University of BRISTOL

Mathematical and Data Modelling 3

EMAT30005

Efficient Imaging In High-Speed Atomic Force Microscopy

November 18, 2020

Authors:
Alfred Brown
Hayden Isaac
Will Leeney
Kit Simmonds

Supervisors:
Dr. Oscar Benjamin
Dr. Naoki Masuda

Abstract

High-Speed Atomic Force Microscopes (HS-AFM) are high-resolution scanning probe microscopes used for the study of samples on the nanoscale, enabling surface topographies to be observed in detail quickly. Due to external factors and mechanical imprecision, the HS-AFM produces images that are often noisy. The current best solution to reduce noise is to take the mean of a collection of frames. Here five algorithms[1] are proposed with the goal of reducing the number of frames, while producing less noisy images that are of similar or better quality than the current method. The quality of an image is evaluated qualitatively (visually) and quantitatively using various evaluation metrics. Our results show that conditional fusion methods performed qualitatively better for less frames on very noisy data than the current method. The findings show a clear dependency between the algorithms performance on the input imagery, and that the algorithms developed make progress in improving the fused image quality.

[1]The code used for this study is readily available for personal use under the link https://drive.google.com/drive/folders/1EOsZiLB0pnrfS26WW_azIeyUscfR3yBE?usp=sharing.

Contents

1 Introduction

The HS-AFM (High-Speed Atomic Force Microscope) has become an essential tool for nanoscientists to understand and investigate topographies and other surface properties that exist on the nanoscale. This microscope is used to map biological processes, chemical reactions and physical processes that occur on the atomic level [1]. The AFM has four main components: a microscopic cantilever with a very sharp tip that is restricted to vertical motion (z-direction); a stage that moves the sample along a fast-scan and slow-scan path (x, y-directions) in the horizontal plane; a laser and a laser-photodetector sensor [2]. The microscope works by reflecting the laser off the tip of the cantilever, such that when a sample moves underneath it and the cantilever is either raised or lowered, the laser-photodetector that receives the reflected beam is able to sense any changes in the height of the cantilever. It thereby produces an image of the surface. See Figure 2 for a visualisation.

One of the many major benefits of the HS-AFM compared to its AFM predecessors is the high speed functionality. The microscope is able to gather a million pixels, one frame, in approximately half a second. However, the frame is subject to many different types of noise, this distorts the image quality and can in some cases also perturb the surface, thereby producing readings that are not true to the sample. The types of noise that can occur vary throughout the scan. Background noise from the environment that the microscope is operating in, can affect the results of the scan. Some noises, such as that caused by talking, can be prevented while the scan is happening. However, there are other types of noise that are random in nature and cannot be prevented without losing the high speed functionality.

The method[2] used to create a less noisy image is called 'image fusion' or 'image stacking'. For clarity, in this report the process is referred to as 'image fusion'. This technique takes multiple frames of the same sample location and uses a method to calculate a new value for each pixel using all of the frames to produce a single image with reduced noise. The current image fusion method, that HS-AFM users at Bristol University implement, calculates the mean value at each pixel across all of the frames.

This paper proposes five algorithms with the intention of reducing the number of frames needed while still attaining less noisy fused images with better or similar quality to those attained through the current method. Specifically, the goal is to reduce the number of frames to less than 20 frames, as this is the maximum number of frames that the Bristol HS-AFM team wishes to use for their image fusion. Qualitative and quantitative analysis is used to determine whether an image is of 'better' or 'similar' quality. Qualitative analysis (visual analysis) is when a team of observers grade the fused image by employing various optical parameters such as spatial details, clarity of features, and lack of recognisable noise [3]. Noise in images generated by the HS-AFM is often easily recognisable to an experienced observer, because they distort the images in particular ways. The reader is encouraged to see Appendix B, where these types of noises are listed with examples. Quantitative analysis is an objective means of evaluating the quality of an image [3]. There are two approaches to quantitative analysis, one that considers a reference image and one that does not. The former case is considered using the *root mean square difference* (RMSD) between a fused and reference image. The latter, by *fusion mutual information* (FMI) and the (RMSD) between fused images. Since the surface topography of a sample is unknown, the reference images used for quantitative analysis is a fused image using the mean of 20 frames of a particular data set, as this is assumed to be the best quality image the current method can generate. Quantitative analysis is therefore only used to evaluate how *similar* in quality the images by the proposed algorithms are to the current method - one cannot assume them to be *better* - we can only do this qualitatively. Furthermore, a consistency measure and a distance measure are also used to assess how much an algorithm varies while fusing frames and how different its resulting image is to the current method of using the mean.

The data used in this study consists of two different sets all taken at the same location. Both of these sets have general background noise but one set has intentionally added noise. Throughout the report these sets are referred to as: 1) the set with increased noise and 2) the noisy set, i.e. noise resulting from uncontrollable sources. For the increased noise set, the noise was generated by tapping the table supporting the HS-AFM, turning on loud machinery and talking in the background. The complete data set consists of two sets of increased noise and just noise for before and after flattening. Flattening is the process in which the surface gradient is estimated as a linear function and then it is subtracted from the the gradient of the surface leaving only the finer surface artefacts. The reason for looking at flattened and unflattened data is to see how the algorithms perform differently on pre-processed frames and on unprocessed frames.

[2]The terms 'method' and 'algorithm' are used in this report to form a hierarchy between overarching ideas that the implemented algorithms use. Some of the algorithms coded for this study share similar characteristics, so we consider them to be implementations of the same method to ease the reader's understanding.

2 Conditional Image Fusion Methods

This section looks at image fusion given certain conditions. The first algorithm, the Conditional Welford's Online Algorithm, places an upper and lower bound condition on each pixel of a frame based on the mean and variance of pixel values of previously observed frames. If a pixel at a certain x, y position of new frame is within the specific boundary determined by previous pixel values at the same position, it is added to the image fusion. The second algorithm, the Conditional Pearson Correlation Algorithm, places a lower bound condition on the value of the Pearson correlation coefficient computed between two frames. This results in only fusing highly Pearson correlated frames.

2.1 Conditional Welford's Online Algorithm

The Conditional Welford's Online algorithm is derived using the approach of Welford [4] – specifically, Welford's Online Algorithm. Utilising this algorithm enables the mean and variance of incoming data to be calculated iteratively, inspecting each value only once; updating both the mean and sample variance alongside each new frame observation. Welford's approach relies on the following recurrence relationships between samples, for the mean and sample variance respectively, to generate the required statistics:

$$\bar{x}_n = \frac{(n-1)\bar{x}_{n-1} + x_n}{n} = \bar{x}_{n-1} + \frac{x_n - \bar{x}_{n-1}}{n}, \tag{1}$$

$$s_n^2 = \frac{(n-2)}{(n-1)} s_{n-1}^2 + \frac{(x_n - \bar{x}_{n-1})^2}{n}. \tag{2}$$

In the HS AFM's case these relationships can be applied to each pixel in a frame, eliminating the need to store the complete frame history that is needed to calculate the mean and median. Through successive iterations, it is reasonable to assume the running mean of pixel z-values converges - since the number of images fused increases and the effects of random noise are negated. The running variance highlights the magnitude of the noise in comparison to the signal, since noise directly influences uncertainty in the data.

The method is conditional, as it precludes pixels outside a specified deviation from the mean from contributing to the running mean or variance. This leads to a different number of pixels being used to calculate the mean after each iteration and therefore the number of pixels used needs to be stored. For this study, the condition is set that the pixel must be within the boundary of the current mean plus or minus one standard deviation, however this condition can be changed by the user.

2.2 Conditional Pearson Correlation Algorithm

The Pearson correlation coefficient is a measure of the linear dependence between two random variables. The two random variables in this case are representative of two frames. The Pearson correlation is computed as,

$$\rho(I_1, I_2) = \frac{1}{N-1} \sum_{i=1}^{N} \left(\frac{(I_1)_i - \mu_1}{\sigma_1} \right) \left(\frac{(I_2)_i - \mu_2}{\sigma_2} \right), \tag{3}$$

where $I_{1,2}$ are the two images being fused and $\mu_{1,2}$ and $\sigma_{1,2}$ are the mean and standard deviation of all pixel values of the images respectively. N here is equal to a million, as the images, which are two dimensional matrices of a 1000 by 1000 pixels, are converted into an array. The Pearson correlation coefficient is bounded between 1 and -1, where a coefficient close to 1 or -1 indicates a high linear correlation. A linear correlation is beneficial given that the frames being fused are of the same sample. The more similar two frames are, the closer the coefficient is to 1, as we expect a positive linear relationship between two frames.

A particular type of error that can occur when the HS-AFM scans a sample, referred to as a drift in Appendix B, is when the scan path shifts ever so slightly causing the generated frames to be out of alignment. The current method of averaging these frames results in a blurry final fused image. The Conditional Pearson Correlation Algorithm avoids this by disregarding frames that are not highly linearly correlated to each other. The algorithm then stores those frames that are highly linearly correlated and fuses them together by computing the mean of each pixel. The condition on the correlation coefficient can be set by the user depending on the number of frames collected. For this study, with the aim of reducing the number of frames to less than 20, the conditional coefficient threshold was set to 0.99, as this was found to reduce the number of images while still attaining images of good qualitative and quantitative quality.

A disadvantage of the Conditional Pearson Correlation Algorithm is that it uses the first frame as a basis on which it conditions all other frames. If the first frame is very noisy and misaligned to the other frames that are to be fused, it will reject them, producing an image of less quality after fusion.

3 Multiresolution Image Fusion Methods

An image fusion technique creates a composite image that retains all the useful information from the source images, and does not introduce features that could interfere with interpretation. The direct approach of summing and averaging can produce unsatisfactory results. If features appear in one source image but not in others, the feature is rendered in the composite at reduced contrast or are superimposed on features from other frames. This section focuses on pixel-level image fusion - this is when the composite is built of spatially registered input images. In this case some generic requirements can be imposed on the fusion schemes:

1. The fusion process should preserve all relevant information of the input imagery in the composite image (pattern conservation).
2. The fusion scheme should not introduce any features or inconsistencies which would distract the human observer or following processing stages.
3. The fusion process should be shift and rotational invariant, i.e. the fusion result should not depend on the location or orientation of an object the input imagery.

The basic idea of the generic multiresolution fusion is that the human visual system is primarily sensitive to local contrast changes, i.e. edges. The fusion methods developed (image pyramid or any wavelet transform) are used to apply a decomposition to the input images, resulting in a multiscale edge representation of the input imagery. The number of levels in the multiscale representation are the decomposition levels, for simplicity this has been set as constant at four levels. It has been shown that the higher number of decomposition levels, the better the quality the fused image is, however at the expense of more complexity [5]. A composite multiscale edge representation is built by a selection of the most salient wavelet coefficients of the input imagery. The fused image is computed by an application of the inverse transform on the composite wavelet representation.

The selection scheme implemented for the coefficients is an area based selection scheme with a consistency verification [6]. This defines the selection of the transform coefficients that carry the 'salient' information for the inclusion in the composite image. Pattern selection is then performed at each sample position of the pyramid for the composite image: The sample value at the position is simply assigned the value of the corresponding sample in the source pyramids that is judged to have the highest salience.

3.1 Laplacian Transform

Image pyramids have been described for multiresolution image analysis. A generic image pyramid is a sequence of images where each image is constructed by lowpass filtering (simple averaging) and subsampling from its predecessor. Due to sampling, the image size is halved in both spatial directions at each level of the decomposition process leading to a multiresolution signal representation. The pyramid transform decomposes each source image into a set of component patterns, the basis functions of the transform. The pyramid transform used to create the pyramids for each image is the Laplacian transform (Laplacian), and is detailed in Liu and Yang [7]. A pyramid is formed for the composite image by selecting coefficients from the source image pyramids. Finally, the composite image is recovered through an inverse pyramid transform. The method is detailed in Zhang et al. [8].

3.2 Discrete Wavelet Transformation

A signal analysis method similar to image pyramids is the discrete wavelet transform (DWT). The main difference is that while image pyramids lead to an over complete set of transform coefficients, the wavelet transform results in a non-redundant image representation.

The discrete 2-dimensional wavelet transform (DWT) is computed by the recursive application of low-pass and high-pass filters in each direction of the input image followed by subsampling. First, the DWT decomposition is applied to all input images, resulting in a multiscale edge representation of the input images. Then a composite multiscale edge representation is built by selecting the most salient coefficients. In the final step, the fused image is computed by the application of the inverse DWT transform on the composite wavelet representation. The method can be found in Li et al. [6]. The wavelet transform has some advantages over the Laplacian pyramid technique:

- The size of the wavelet transform is the same as the image. However, the size of the Laplacian pyramid is 4/3 the size of the image, which means the wavelet transform is more compact.
- The pyramid representation fails to introduce any spatial orientation selectivity into the decomposition process.
- The wavelet kernels can be chosen to be orthonormal, therefore the information contained at different resolution levels is unique. However, the pyramid decomposition contains redundancy between different levels.

3.3 Shift Invariant Discrete Wavelet Transformation

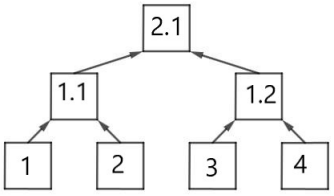
One major drawback of the wavelet transform when applied to image fusion is its well known shift dependency, i.e. a simple shift of the input signal may lead to complete different transform coefficients. This results in inconsistent fused images when invoked in image sequence fusion. To overcome the shift dependency of the wavelet fusion scheme, the input images must be decomposed into a shift invariant wavelet representation. This is done by computing the wavelet transformation for all possible (circular) shifts of the input image.

The actual fusion process in the Shift Invariant Discrete Wavelet Transform (SIDWT) case is identical to the generic wavelet fusion case, the images are first decomposed into their shift invariant representation. A composite shift invariant wavelet representation is then built by the incorporation of the same selection scheme implemented previously [6]. The fused image is constructed by performing the inverse SIDWT transform using the appropriate filters. The SIDWT method is given in Rockinger [9].

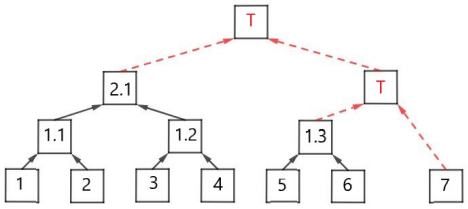
3.4 Triangular Fusion Algorithm

The fusion processes (Laplacian, DWT, SIDWT) create a composite image based on two input images, these methods need to be extended so that any number of images can be fused together. The obvious solution to save the composite image and fuse it with the next input image, repeating until there are no more input images. However, this means the algorithm inherently allocates a weighting to the starting images as each image becomes a part of the fused image and therefore is involved in all future fusions. This is not ideal as it it assumed that every image has an equal probability of being of a 'good' quality, whereas this solution favours the first images to be fused.

The Triangular Fusion Algorithm (TFA) is a framework for fusing more than two images together, where there is minimum dependency on the order of images. This is useful when the fusion algorithm does not allow more than two images at a time. The TFA (see Appendix C, algorithm 1) process is illustrated by Figure 1, and works by combining two images with a fusion method if there are two images that have not been fused on any level. If there are an odd number of frames on a level, the odd image out is passed onto the next level temporarily, this creates an output that will not be stored for future calculations.



(a) TFA implemented on four images, showing three levels of fusing where Image 2.1 is the output.



(b) TFA implemented on seven images, this illustrates how images are fused when there is an odd number of frames on any level, with image T as the output. T is a temporary placeholder for images that need to be fused but not stored in memory.

Figure 1: TFA visualised to show how images are fused with four and seven frames

4 Evaluation Metrics

Quantitative analysis determines the performance of a fused image using either a reference image or no reference image [3]. With a reference image, metrics that can be used for objective analysis are, among others, root mean square difference (RMSD) and relative dimensionless global error (ERGAS). Without a reference image, metrics such as fusion mutual information (FMI) and fusion quality index (FQI) can be used. For this study we used a root mean square error as a comparison measure to the mean of 20 fused images and a fusion mutual information measure in order to quantify the quality of the fusion algorithms and images generated by these algorithms. Further measures are implemented to compare how consistent the algorithms arrive at generating a fused image and how similar these images are to the fused images produced by taking the mean.

4.1 Root Mean Square Difference to Reference Image

The RMSD is an indicator of the spectral quality of a fused image [3]. Since we are only dealing with grey-scale images, where the 'spectrum' refers to the z-values of the sample, the RMSD is useful as it shows to how similar a

fused image is to a reference image. The lower the value of the RMSD, the closer the fused image is to the reference image. The formula used for the RMSD is,

$$RMSD = \sqrt{\frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}\left(I_R(i,j) - I_F(i,j)\right)^2}, \tag{4}$$

where I_R and I_F are the reference and fused image respectively.

4.2 Fusion Mutual Information

Fusion mutual information is used to measure the degree of dependency between an input imagey and a fused image [3]. Let A and B be two images and F be the fused image of both of them. The FMI computes,

$$FMI = MI(A,F) + MI(B,F), \tag{5}$$

the sum of the mutual information of fused image F with images A and B. This is extendable to any number of inputs. The higher FMI value, the likelier it is that the fused image shares high mutual information with both images. A high FMI value is indicative of a good fusion algorithm and if the images A and B are considered of good quality, then the fused image is considered of good quality too. We cannot assume that a high FMI value indicates a better image quality, only that it is of *similar* quality to the two images being fused. The mutual information between an image and the fused image is computed as,

$$MI(A,F) = \sum_{k=1}^{N}\sum_{i=1}^{N} P_{FA}(k,i)\log_2\frac{P_{FA}(k,i)}{P_F(k)P_A(j)}, \tag{6}$$

where P_A and P_F are the histograms of A and F, and P_{FA} is the normalised joint histogram of the two sets of images [10]. The same formula is naturally applicable for $MI(B,F)$.

4.3 Consistency Metric

Since the true surface topography of a sample is not known, we consider the convergence of the RMSD between two frame sets, using each method, collected at a single location. Any two different frame sets at the same location have unique noise associated to them. A comparison of the consistency of the algorithms is drawn by calculating the RMSD at each iteration and analysing the convergence, where present, between these two sets fused images. These two sets will be generated by splitting the data sets that exist in halve.

4.4 Qualitative Analysis

Qualitative analysis of images is performed visually by a team of observers using certain criteria [3]. An image that conforms well to the criteria is considered an image of higher quality. The criteria used is shown in the table below.

Table 1: The criteria used for visual analysis of images generated by fused HS-AFM images.

Criteria	Description	Grade
Sharpness	Image is not blurry and surface topography features are clearly visable	1-4
No return to zero	Caused by a sudden 400-500 nm drop or rise in the data, this can result in parts of the image having lines of the same grey scale value.	1-4
No smeared lines	Caused by debris on the surface of the sample, this can result in parts of the image looking smeared (see Figure 3a)	1-4
No repeating parallel lines	This is usually due to laser misalignment, causing repeated parallel lines to off-vertical surface features (see Figure 3b).	1-4

Grading	
1	Excellent
2	Good
3	Poor
4	Very Poor

5 Results

It should be noted that for comparative purposes taking the median of fused images was also considered.

Qualitative

Visually (see figures 4, 5, 6 and 7), it is observed that overall the Conditional Pearson Correlation Algorithm and Welford's Online Algorithm generated the images of highest qualitative quality for the fusion of 5, 10, 15 and 20 frames. The current method of using the mean performed well for less noisy data, but badly for the induced noisy data whereby surface features became more blurry and eventually almost invisible to the naked eye as more frames were added to the fusion. The SIDWT, Laplacian and DWT algorithms perform very poorly for the induced noisy data, but well for the less noisy data. Instead of losing surface features, however, these algorithms shift the features around the image space, progressively distorting the images as more frames are added. The median generally performs well, except for the very noisy data set, where it loses one of the key features of the sample's surface. These results are summarised in the table below.

Table 2: Summarised table of qualitative results for each of the algorithms. The grades were subjectively decided upon, using the criteria shown above. The values presented for the data sets are rounded average grades from looking at how the algorithms performed after $5, 10, 15$ *and* 20 *frames (see Appendix D). The general performance is then presented by an average in the last column. I. Noise refers to increased noise.*

	Average Grade given Criteria for Algorithms				
	I. Noise flattened	Noise flattened	I. Noise unflattened	Noise unflattened	Average
Mean	4	2	2	2	2.5
Median	4	2	2	2	2.5
Cond. Welford	2	2	2	2	2
Cond. Pearson	1	2	2	2	1.75
DTW	4	2	4	3	3.25
Laplacian	4	2	4	3	3.25
SIDWT	4	2	4	3	3.25

Fusion Mutual Information

The FMI metric shows how much information each algorithm uses from the input imagery to create the fused image, the results of this can be seen in Figure 9. The results show that the FMI values for all algorithms change more in the first six frames than the after this. The Conditional Welford method does not always follow this as for the flattened data it does not experience large changes to FMI throughout. All multiresolution methods converge to a lower FMI value than the conditional methods, this is might be because the methods both calculate a mean. If we compare this with the visual analysis (figures 4, 5, 6 and 7)we see the conditional algorithms produce images of similar quality and the multiresolution results are also similar. It is interesting to note the FMI values fluctuate within a smaller range of values, as this implies that less information is gained per frame. The unflattened data for increased noise fluctuates the most out of the data sets, this means that data sets that have more noise need to stack more frames, this is as expected. Additionally, the conditional models fluctuate less than the multiresolution methods, this is because the conditional methods depend more on the first image, so additional frames fused transfer less information to the fused image.

Consistency Metric

These results analyse Figure 14 where the Summed Root-Mean-Square Difference (RMSD) between two subsets of frames fused image is plotted against the number of frames considered for each method.
It can be seen in Figure 14a that the RMS difference between the z-values of two sets of the same sample (RMSD) using the SIDWT increases for the first six frames and subsequently fluctuates by approximately 2×10^5 z-values for the remaining frames. These fluctuations occur approximately every five to six frames. The first three graphs show the RMSD does not show signs of convergence within 20 frames. In Figure 14d the RMSD shows a negative correlation with the number of frames, indicating convergence. The SIDWT follows a similar pattern to the DWT and the Laplacian, however the RMSD can be seen to be smaller than these throughout frame iterations in Figure 14a. The similarities between the RMSD between these three algorithms continues in the other three data sets, however in some instances the DWT RMSD overlaps or declines below the SIDWT RMSD.

The RMSD of the DWT generally lies between that of the SIDWT and the Laplacian, with the exception of frame 16 in Figure 14c and Frames 16-20 in Figure 14d. The Laplacian has the highest RMSD in the majority of cases and begins to converge with the DWT at higher numbers of frames. The pattern the DWT and the Laplacian follow are reflective of the SIDWT, rising and falling alongside it at every frame.

The RMSD of the Conditional Pearson drops significantly lower than all other algorithms when considering the second frame on all graphs - the lowest RMSD of any algorithm at any number of frames. In figures 14a and 14b, the unflattened data sets, it follows the mean exactly from frame three onward - steadily decreasing. When considering the flattened data sets in figures 14c and 14d the RMSD of the Conditional Pearson remains constant from frame four and two respectively, having the highest final RMSD of any algorithm at 20 frames in Figure 14d.

In all cases the mean shows a negative correlation between the RMSD and number of frames, indicating convergence and consistency. Joint with the Conditional Pearson in Figure 14a the mean displays the highest consistency indicated by the lowest final RMSD at 20 frames, and, exclusively, the lowest final RMSD in figures 14c and 14d.

The RMSD of the median always matches the mean at two frames and generally follows closest to the mean with successive frame fusions. In every case the RMSD of the median declines and in 14b takes the smallest final value at 20 frames.

Table 3: Final summed RMSD in z-values after 20 frames using the consistency metric for each algorithm.

	Final summed RMSD in z-values after 20 frames ($\times 10^5$) (3 S.F.)			
Algorithm	Noise Flattened	Noise Unflattened	I. Noise Flattened	I. Noise Unflattened
SIDWT	22.7	7.30	25.0	7.25
Laplacian	24.5	7.88	29.5	7.84
DWT	24.2	7.69	29.2	7.56
Cond. Pearson	23.5	5.09	38.4	5.51
Cond. Welford	21.9	5.40	30.3	5.51
Mean	17.6	5.09	18.7	5.51
Median	18.6	5.58	19.5	5.24

Difference to Mean of 20 Fused Images

To find out if the difference between the algorithms and the image produced after taking the mean of 20 images is due to a high concentration of large differences or small differences across all of the pixels, histograms of the difference in pixel values between each algorithm and the mean were plotted. They show the distribution of the difference in pixel values with different numbers of frames added to the fusion. The flattened increased noise data and the unflattened increased noise data, shown in figures 11 and 12 respectively, have a larger distribution than their corresponding noise data. After the fusion of 20 frames, the spread of the distribution is greater for the increased noise data than it is for the noise data. The Conditional Welford Algorithm is the most similar to the mean in three out of the four sets. The set it did not do as well in contained the unflattened frames under increased noise, in which the Conditional Pearson Algorithm is the most similar to the mean of 20 frames. All of the algorithms have little change in the distribution of the difference in pixel values after 10 frames have been fused.

6 Discussion

6.1 Analysis of Results and Metrics

The results of the histograms show that the algorithm most similar to the current method of choice, the mean, is the Conditional Welford Algorithm. This is because the Conditional Welford Algorithm calculates the running mean but where it differs is that it will reject pixels if it does not fit the condition set. Since the mean and Conditional Welford Algorithm are not the same, it can be assumed that the Conditional Welford Algorithm is rejecting some pixels and that because of this the final image will be of a better quality. The issue with the Conditional Welford Algorithm is that it will be biased to the initial frame, which can lead to errors. If the first frame has a large amount of noise in it then it will then reject some pixel values, that lie outside the condition, which are not affected by noise.

The histograms also show that the after 10 frames are fused, any future frames have little impact on the difference between the algorithm and the mean of 20 frames. This leads us to believe that after 10 frames the fused image will not change significantly with more frames added. However, this does not mean that those images are of better quality. The figures also show five fused frames for each method. There is a larger difference between the distribution of the five frames to the 10 frames and therefore the five fused frames will change with more frames added.

Generally, the mean operates with greatest consistency when generating a fused image, showing a negative correlation between the RMSD and number of frames for each data set. The median also shares this negative correlation, however in three of four instances has a higher final RMSD for z-values after 20 frames - reflective of a lower consistency. Some algorithms function similarly, the SIDWT being a derivative form of the DWT for instance, which is reflected in the

results. This is most apparent where the SIDWT, DWT and the Laplacian Transformation's RMSD fluctuate against frame number in synchronisation.

Since the Conditional Welford's Online and Pearson Correlation Coefficient ultimately rely on the mean, the three RMSD's for each algorithm also manifest in similar ways, converging to the exact same RMSD in one case. The median shows the lowest final RMSD for the unflattened increased noise data set, a result of its resilience to outliers which suggests it could be a good method for particularly noisy data. When applied to the flattened data sets the Pearson algorithm's RMSD remains fixed following the first four frames; this is the result of an over stringent condition. The correlation coefficient condition causes most frames to be omitted from image fusion in both sets, resulting in constant z-values with each subsequent frame and thereby a constant RMSD between the images.

It is important to evaluate the metrics by which the algorithms are evaluated themselves, this is because it shows how much our results and the conclusions we draw from them mean. Qualitative analysis results provide a visual assessment, this is useful to see the initial clarity of an image, however these results are subjective to the observer. The observers are not experts in the field of nanoscience, so it would be improper to say these results have been tested rigorously. This means that they will not rank feature quality, therefore image quality, the same as someone who works with the HS-AFM on a daily basis. Other metrics used also have disadvantages, for example RMSD uses the mean of 20 images as a reference image. Although this is the current golden standard, it is obvious from the visual assessment of the increased noise flattened data set (see Figure 4) that both conditional algorithms result in images of higher quality than the mean. This means that the reference image used is redundant as the quality metric will not convey that an image can be of better quality than the reference.

The FMI metric is a measure of how much information is gained from the the input imagery, that means FMI is less useful as a quality metric as it does not explicitly judge image quality. However, this metric may be useful in conditioning the addition of new frames to be fused, as conditioning has been shown to improve image quality greatly. This may be done by implementing a condition to stop adding frames if the change in FMI falls to zero. Doing this would result in less frames being fused when the information gain change is low, implying only small improvements to quality. There are other quality metrics that could replace FMI for judging quality [3], future investigations could use more metrics without a reference image to judge quality. In addition, it would be beneficial to compare image quality for a small number of images fused, as it may not be realistic to assume having 20 plus images to stack from. This may be done with a consistency check comparing the average difference for 10 sets of five images.

There are aspects in each algorithm that have not undergone enough testing, and could therefore improve the algorithms' performance. Various selection schemes have been proposed, as shown in the multiresolution fusion methods, such as a simple 'choose max scheme' which would change the performance of those algorithms. In addition, four decomposition levels are used in the method, however Haghighat et al. [5] shows that increasing this will lead to an increase in image quality. There are variables in various algorithms that should be evaluated to define the optimal value for image quality. An example is the conditional coefficient threshold for Pearson's Algorithm, this is currently held constant at 0.99, but could be changed to be a function depending on the number of frames added. Another possibility would be multiplying the variance by a constant to clip pixel values in the conditional Welford algorithm.

In the introduction, the goal was described as reducing the number of frames to less than 20 images while still maintaining fused images of similar or better quality to the current method of taking the mean. The following summarises all results:

- The Conditional Pearson Algorithm generated images of high qualitative quality and did so with less frames compared to the mean.
- The Conditional Welford Algorithm behaved similar to the mean, but better for very noisy data and did so with less frames.
- SIDWT, DWT and the Laplacian algorithms performed well for good quality frames, but suffered for noisy images as they fuse all images. These algorithms did not generally perform similar to the mean.

6.2 Future Work

The methods tested rank differently when fusing different input imagery, which means that no method in particular is categorically better than the rest. Therefore, it would be useful to combine parts of each method to create images of higher quality for any input imagery. The Conditional Pearson algorithm uses a condition for whether an image should be fused at all, which allows frames of poor quality to not be fused. This would help the multiresolution representation algorithms perform better, as they do not judge the quality of an image before fusing. There was not enough time during the project to fully test this idea, however some initial visual analysis has been done (see Figure 8). The Salmagundi Algorithm (SA) (see Appendix C, Algorithm 2) produces sharp images when tested with the same input imagery as the other algorithms. In particular, the feature at the bottom of the flattened data with

increased noise appears to be clearer than the images produced by other algorithms. On the other hand, the SA has not been tested quantitatively or with rigour, so it lacks validation for use. The TFA was constructed so that the multiresolution fusion algorithms would not depend on the first images, however these algorithms could be extended so that more than two images can be combined. This would mean better scale-ability in computation, however the Salmagundi Algorithm relies only on two images being fused together as it has a condition for fusing images.

There are many areas for improvement on the research done on top of this, one such example is using techniques from the emerging field of machine learning. This might be using inference techniques to try and approximate a final fused image without any noise. Another idea could be to use Bayesian learning to encode a likelihood function for how much information could be gained from fusing a new image. These techniques would mean less frames need to be fused overall as the final image would be of higher quality.

Another area that should be duly noted as it affects both scanning speed and image quality is the scan path. All data collected and used was done so with raster scan paths, which results in more data being collected round the edges of a frame, as the input data is time constant. This may be improved upon in further work by implementing non-raster scans, specifically self-intersecting sinusoidal scans. Sinusoidal paths mean that data is collected more evenly throughout the image, however flattening cannot be used on this scan path. Although sensor inpainting can generate an image from any arbitrary path, so any path pattern can be used. In addition, the self intersecting paths have been shown to effectively remove drift so they might be more appropriate for collecting data.

All methods use specific approaches which do not depend on the input imagery, to reduce noise in the image. This may be the most time efficient method to reduce noise, however different noise affects the image quality in certain ways and it may be that different approaches are better at decoding this noise. It would be beneficial for further research to generate frames with a particular noise source and try to classify it to different approaches to image fusion. It would be interesting to compare the use of combining a mean based fusion on data with low variance and a median approach with data that has a high variance.

7 Conclusion

This paper explores extensively methods for fusing multiple images produced by a HS-AFM of a sample location. The methods aimed to increase the image quality and reduce the number of frames needed to fuse. Both qualitative and quantitative techniques have been used to validate the quality of the fused images, including RMSD and FMI. The performance of each algorithm varies on input imagery that has been subjected to different scanning conditions. This means that the suitability of each algorithm is subject to the amount of noise present in the imagery. The Salmagundi Algorithm makes a good start at implementing a blanket solution for any input imagery, however not enough testing has been carried out to draw a reliable conclusion. Future researchers can base their studies on rigorous testing of the SA, to further development of a general fusion algorithm. Alternative work may be based on machine learning techniques to improve the quality of individual images.

References

[1] A. Chen, A. Bertozzi, P. Ashby, P. Getreuer, and Y. Lou. Enhancement and recovery in atomic force microscopy images. *Journal of Material Science and Engineering*, 01(03), 2012. doi: 10.4172/2169-0022.s1.006.

[2] I. A. Mahmood and S. O. Reza Moheimani. Fast spiral-scan atomic force microscopy. *Nanotechnology*, 20(36), 2009. doi: 10.1088/0957-4484/20/36/365503.

[3] P. Jagalingam and A. V. Hegde. A review of quality metrics for fused image. *Aquatic Procedia*, 4:133–142, 2015.

[4] B. P. Welford. Note on a method for calculating corrected sums of squares and products. *Technometrics*, 4: 419–420, 1962.

[5] M.B.A. Haghighat, A. Aghagolzadeh, and H. Seyedarabi. A non-reference image fusion metric based on mutual information of image features. *Computers & Electrical Engineering*, 37(5):744–756, 2011.

[6] H. Li, B.S. Manjunath, and S.K. Mitra. Multisensor image fusion using the wavelet transform. *Graphical models and image processing*, 57(3):235–245, 1995.

[7] G. Liu and W. Yang. Multisensor image fusion based on wavelet transform. In *Process Control and Inspection for Industry*, volume 4222, pages 219–224. International Society for Optics and Photonics, 2000.

[8] Z. Zhang, R.S. Blum, et al. A categorization of multiscale-decomposition-based image fusion schemes with a performance study for a digital camera application. *Proceedings of the IEEE*, 87(8):1315–1326, 1999.

[9] O. Rockinger. Image sequence fusion using a shift-invariant wavelet transform. In *Image Processing, 1997. Proceedings., International Conference on*, volume 3, pages 288–291. IEEE, 1997.

[10] X. Zhang, Z. Liu, Y. Kou, J. Dai, and Z. Cheng. Quality assessment of image fusion based on image content and structural similarity. In *Information Engineering and Computer Science (ICIECS), 2010 2nd International Conference on*, pages 1–4. IEEE, 2010.

[11] O. D. Payton, L. Picco, and T. B. Scott. High-speed atomic force microscopy for materials science. *High-speed atomic force microscopy for materials science, International Materials Reviews*, pages 365–503, 2009.

[12] P. Eaton, Sep 2007. URL http://www.fc.up.pt/pessoas/peter.eaton/artifacts/sampledrift.html. Visited on 2018-02-11.

[13] B. W. Erickson, S. Coquoz, J. D. Adams, D. J. Burns, and G. E. Fantner. Large-scale analysis of high-speed atomic force microscopy data sets using adaptive image processing. *Beilstein journal of nanotechnology*, pages 747–758, 2012.

8 Appendices

8.1 Appendix A - Schematic Diagram of an HS-AFM

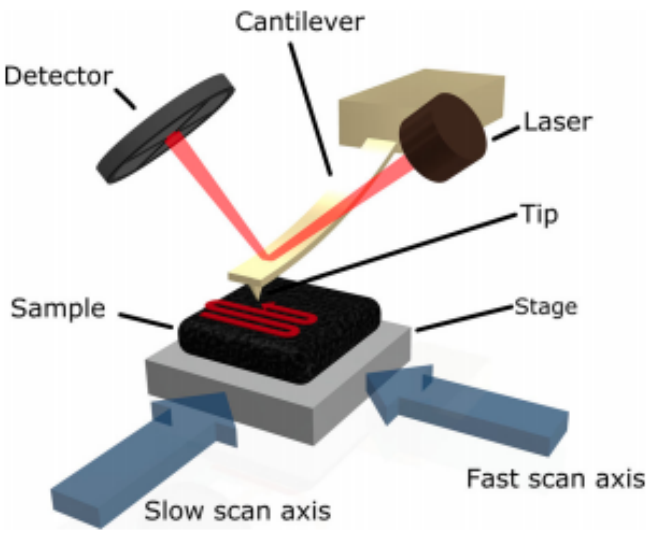


Figure 2: A schematic diagram of a simple AFM. The cantilever's vertical response to the sample is captured by the detector as the sample is moved along the slow and fast scan axes in a raster pattern. For further reading on how the AFM works and for detailed differences between the HS-AFM and the AFM [11]. Note, this diagram is not representative of the HS-AFM that Bristol University uses, however, it is a useful aid to the reader to understand the fundamental idea behind how the HS-AFM works. ©Payton

8.2 Appendix B - Causes of Noise

The HS-AFM is a delicate mechanical microscope and is thereby susceptible to a variety of noise. In order to fully understand the approach taken in this paper, it is necessary to be familiar with the types and causes of noise that can distort the data collected by the microscope.

Parachuting

Parachuting occurs when the cantilever tip loses contact with the surface of the sample, in doing so the z-values obtained for a small number of subsequent pixels are skewed, as the deflection beam over estimates the height of the sample at these points.

Laser Misalignment

The HS-AFM requires extreme precision in its initial setup; the alignment of the laser over the probe must be exact, and any deviation can have consequences. In the case of a misaligned laser, the noise this causes typically manifests itself in the form of a ripple like effect over the captured image, with repeating and fading patterns, often reflective of prominent surface features. (see Figure 3b)

Drift - Thermal and DC

Drift occurs when the sample moves with respect to the microscopes stage, this can be the result of sliding or thermal expansion and contraction; the sample can be seen to drift across the frame [12].

Signal out of Range

When a signal falls out of range it is, perhaps, the most easily perceived form of noise - that is to say, it is clearly visible when present in image data forming large abrupt bands. This noise is caused when a z-value reading breaches an upper height threshold; resulting in a 'reset' of subsequent z-values.

Surface Debris

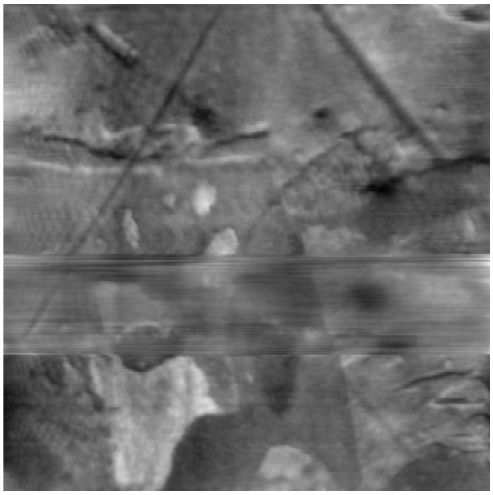
When loose debris is present on the surface of a sample it can easily collide and stick to the cantilever tip, causing a temporary loss of resolution within the frame (see Figure 3a).

Trace and Retrace Misalignment

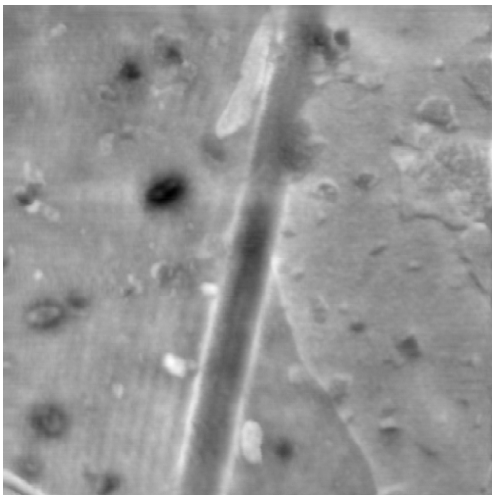
Trace and Retrace refer to the direction of the scan. Trace being the initial direction and retrace being the reverse. The misalignment between frames is typically caused by hysteresis in the motion of the stage in the trace and retrace directions caused by the piezoelectric motors which drive the motion.

Flattening

Flattening maps the surface of a sample onto a flat plane. Any material being scanned has some kind of slope to it and if the gradient is too great then all of the smaller surface texture will become negligible compared to the overall height difference between the ends of the slope. This is a problem. A frame that has not been flattened will produce z-values for each pixel that will primarily be from the sloped surface making the frame more difficult to see. To counter this problem a polynomial relative to the slope is subtracted from the frame to remove the large differences in height and produce pixel values that display the surface texture [13].



(a) Noise in a frame due to surface debris, notice the loss in resolution in a band of the frame.



(b) Noise in a frame due to a misaligned laser in the initial microscope setup, notice the repeating lines parallel to the off-vertical surface feature.

Figure 3: Two examples of noise in HS-AFM imaging.

8.3 Appendix C - Algorithms

It should be noted that the procedure $Fuse(a, b)$ is the appropriate fusion procedure for the method and takes two images as the input and returns a single fused image. $corrcoef(c, d)$ is a function that computes the Pearson's correlation coefficient between two images and $size(e, f)$ is a function that returns the size of e in dimension f.

Algorithm 1 Triangular Fusion Algorithm

procedure ALGORITHM(*FramesToStack*, *NumFrames*)
  $lengths$, $LastOdd$, $Fusion \leftarrow initialisation$
  $Fusion(:, :, 1, 1) \leftarrow FramesToStack(:, :, 1)$
  if $NumFrames = 1$ **then return** $Fusion(:, :, 1, 1)$
  $Fusion(:, :, 2, 1) \leftarrow FramesToStack(:, :, 2)$
  $Fusion(:, :, 1, 2) \leftarrow \text{fuse}(Fusion(:, :, 2, 1), Fusion(:, :, 1, 1))$
  if $NumFrames = 2$ **then return** $Fusion(:, :, 1, 2)$
  $z \leftarrow 1$
  while $z \neq NumFrames$ **do**
    $s \leftarrow size(Fusion, 4)$
    $w \leftarrow 1$
    $LastOdd \leftarrow 0$
    while $w \neq s$ **do**
      if $(lengths(w) \bmod 2 \neq 0)$ and $(LastOdd = 0)$ **then**
        $Fusion(:, :, \frac{z}{2^w}, w + 1) \leftarrow \text{fuse}(Fusion(:, :, \frac{z}{2^{w-1}}, w), Fusion(:, :, \frac{z}{2^w} - 1, w))$
        $Y = Fusion(:, :, \frac{z}{2^w}, w + 1)$
      else if $(lengths(w) \bmod 2 = 0)$ and $(LastOdd = 1)$ **then**
        $temp \leftarrow \text{fuse}(Fusion(:, :, \frac{z-1}{2^{w-1}}, w), temp)$
        $Y = temp$
      else if $lengths(w) \bmod 2 \neq 0$ **then**
        $temp \leftarrow \text{fuse}(Fusion(:, :, lengths(w), w), temp)$
        $Y = temp$
        $LastOdd \leftarrow 2$
      else
        $temp \leftarrow Fusion(:, :, z, 1)$
        $Y = temp$
        $LastOdd \leftarrow 2$
      $w \leftarrow w + 1$
    $lengths(1) \leftarrow lengths(1) + 1$
    while $i \neq s$ **do**
      if $(lengths(i) \bmod 2 = 0)$ **then**
        $lengths(i + 1) = \frac{lengths(i)}{2}$
      $i \leftarrow i + 1$
    $z \leftarrow z + 1$
  return Y

Algorithm 2 Salmagundi Algorithm

procedure ALGORITHM(*FramesToStack*, *NumFrames*)
  $lengths$, *LastOdd*, *Fusion* ← *initialisation*
  $Fusion(:,:,1,1) \leftarrow FramesToStack(:,:,1)$
  if $NumFrames = 1$ **then**
    $Y = Fusion(:,:,1,1)$ **return** Y
  $Fusion(:,:,2,1) \leftarrow FramesToStack(:,:,2)$
  $Fusion(:,:,1,2) \leftarrow \text{fuse}(Fusion(:,:,2,1), Fusion(:,:,1,1))$
  if $NumFrames = 2$ **then**
    $Y = Fusion(:,:,1,2)$ **return** Y
  $z \leftarrow 1$
  while $z \neq NumFrames$ **do**
    $s \leftarrow size(Fusion, 4)$
    $w \leftarrow 1$
    $LastOdd \leftarrow 0$
    $X = lengths(1) + 1$
    if $\text{corrcoef}(Y, FramesToStack(:,:,z)) > 0.99$ **then**
      while $w \neq s$ **do**
        if $(lengths(w) \bmod 2 \neq 0)$ and $(LastOdd = 0)$ **then**
          $Fusion(:,:,\frac{X}{2^w}, w+1) \leftarrow \text{fuse}(Fusion(:,:,\frac{X}{2^{w-1}}, w), Fusion(:,:,\frac{X}{2^w} - 1, w))$
          $Y = Fusion(:,:,\frac{X}{2^w}, w+1)$
        else if $(lengths(w) \bmod 2 = 0)$ and $(LastOdd = 1)$ **then**
          $temp \leftarrow \text{fuse}(Fusion(:,:,\frac{X-1}{2^{w-1}}, w), temp)$
          $Y = temp$
        else if $lengths(w) \bmod 2 \neq 0$ **then**
          $temp \leftarrow \text{fuse}(Fusion(:,:,lengths(w), w), temp)$
          $Y = temp$
          $LastOdd \leftarrow 2$
        else
          $temp \leftarrow Fusion(:,:,X,1)$
          $Y = temp$
          $LastOdd \leftarrow 2$
        $w \leftarrow w + 1$
      $lengths(1) \leftarrow lengths(1) + 1$
      while $i \neq s$ **do**
        if $(lengths(i) \bmod 2 = 0)$ **then**
          $lengths(i+1) = \frac{lengths(i)}{2}$
        $i \leftarrow i + 1$
    $z \leftarrow z + 1$
  return Y

8.4 Appendix D - Qualitative Results

Below are the generated visual results. Please rotate the next following pages to landscape to compare the images better for each of the algorithms. The results appear in the following order: FMI, RMSD, Consistency.

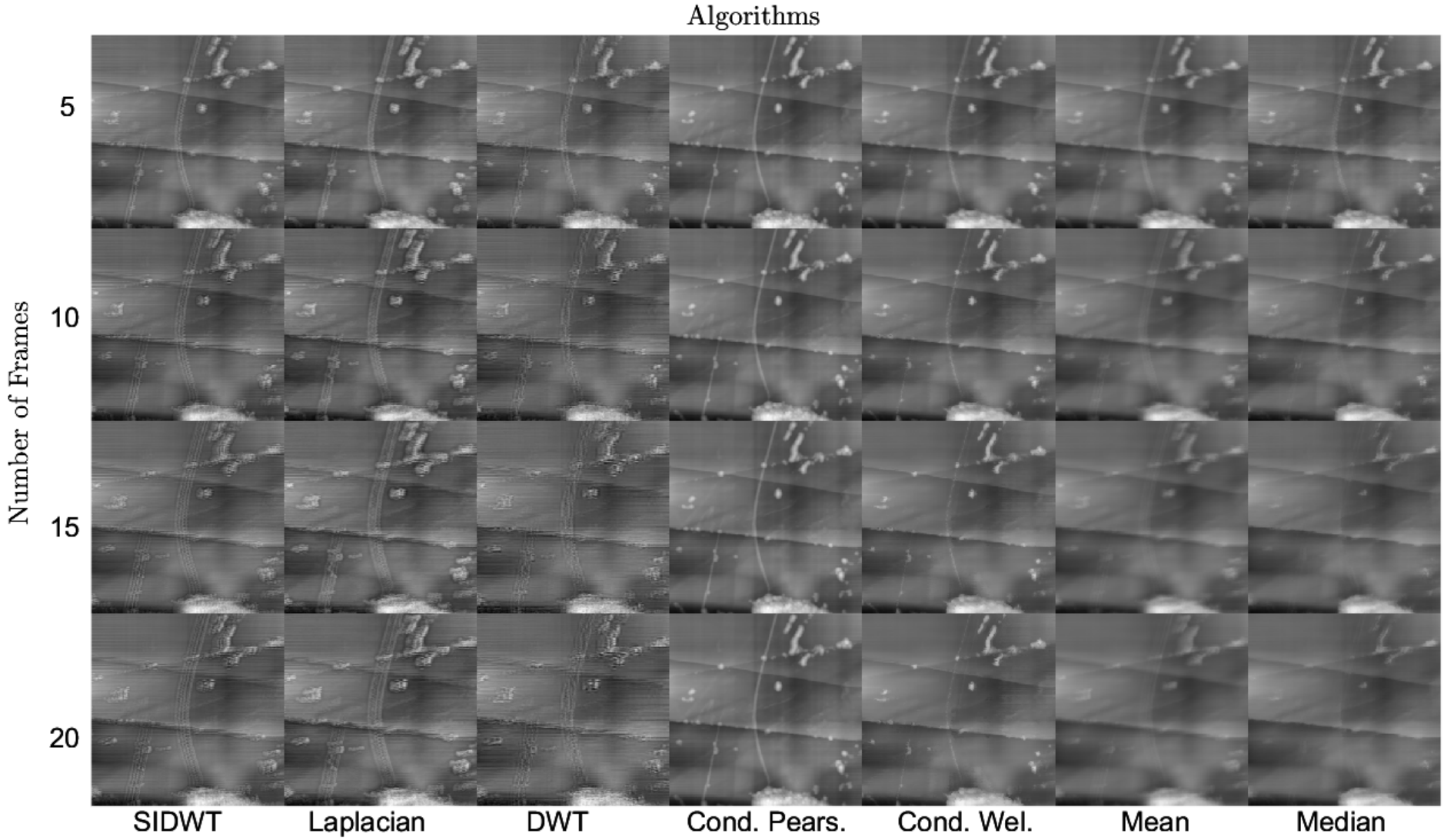


Figure 4: A montage of images generated by each of the algorithms for 5, 10, 15 *and* 20 *frames for the* ***Increased Noise flattened data set****. View this figure from top to bottom for each column to see how the generated images progress with added frames for each algorithm. For this data set, it is noticeable that the Conditional Welford Online Algorithm and Conditional Pearson Algorithm generate less blurry, sharper images than the mean after just considering five frames. The mean, median, SIDWT, Laplacian and DWT algorithms become more distorted and blurry as the number of frames increases.*

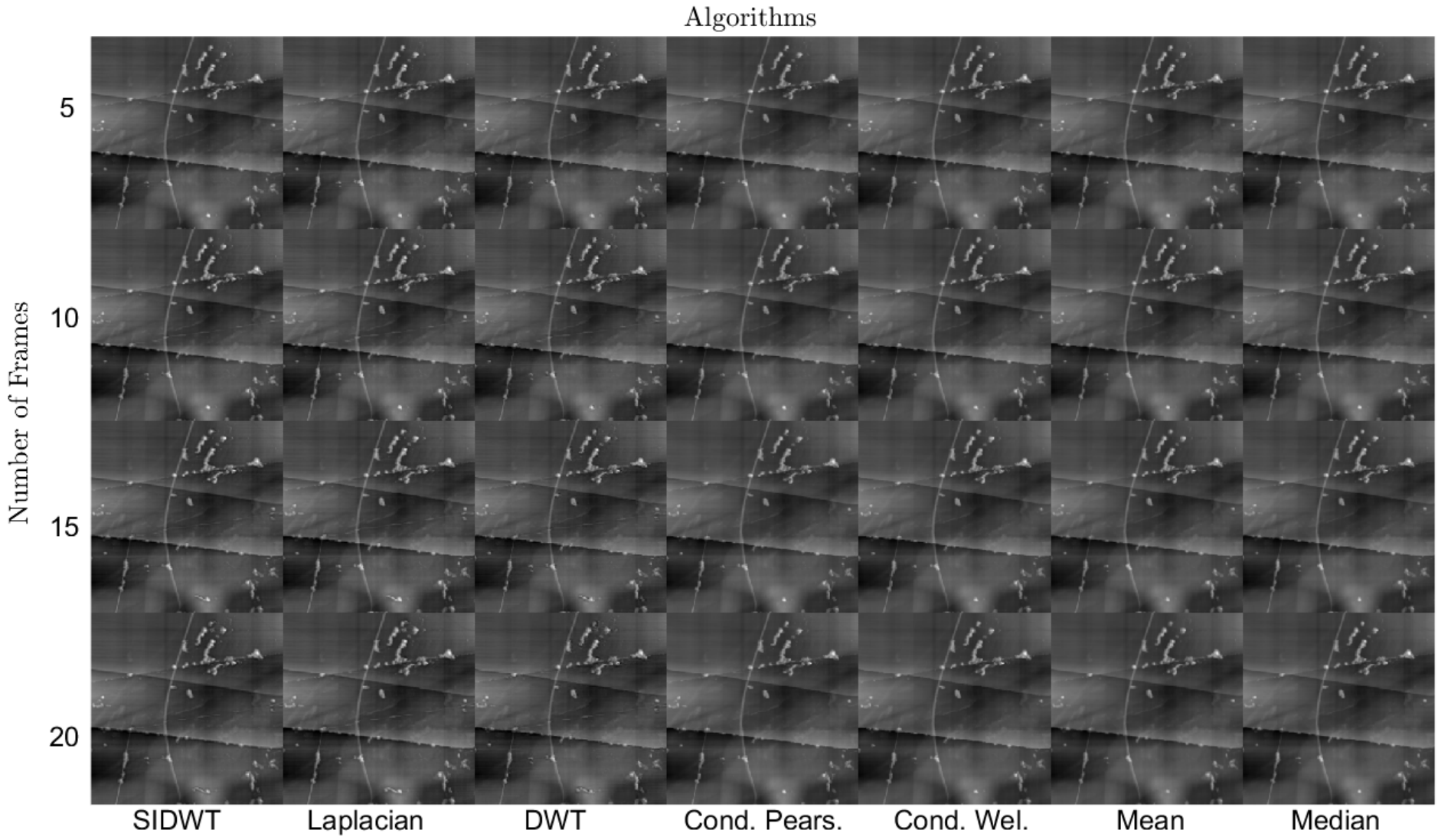


Figure 5: A montage of images generated by each of the algorithms for 5, 10, 15 *and* 20 *frames for the* ***Noisy flattened data set****. For this data set all algorithms are able to maintain the surface features. The SIDWT, Laplacian, DWT algorithms show some of small features clearly.*

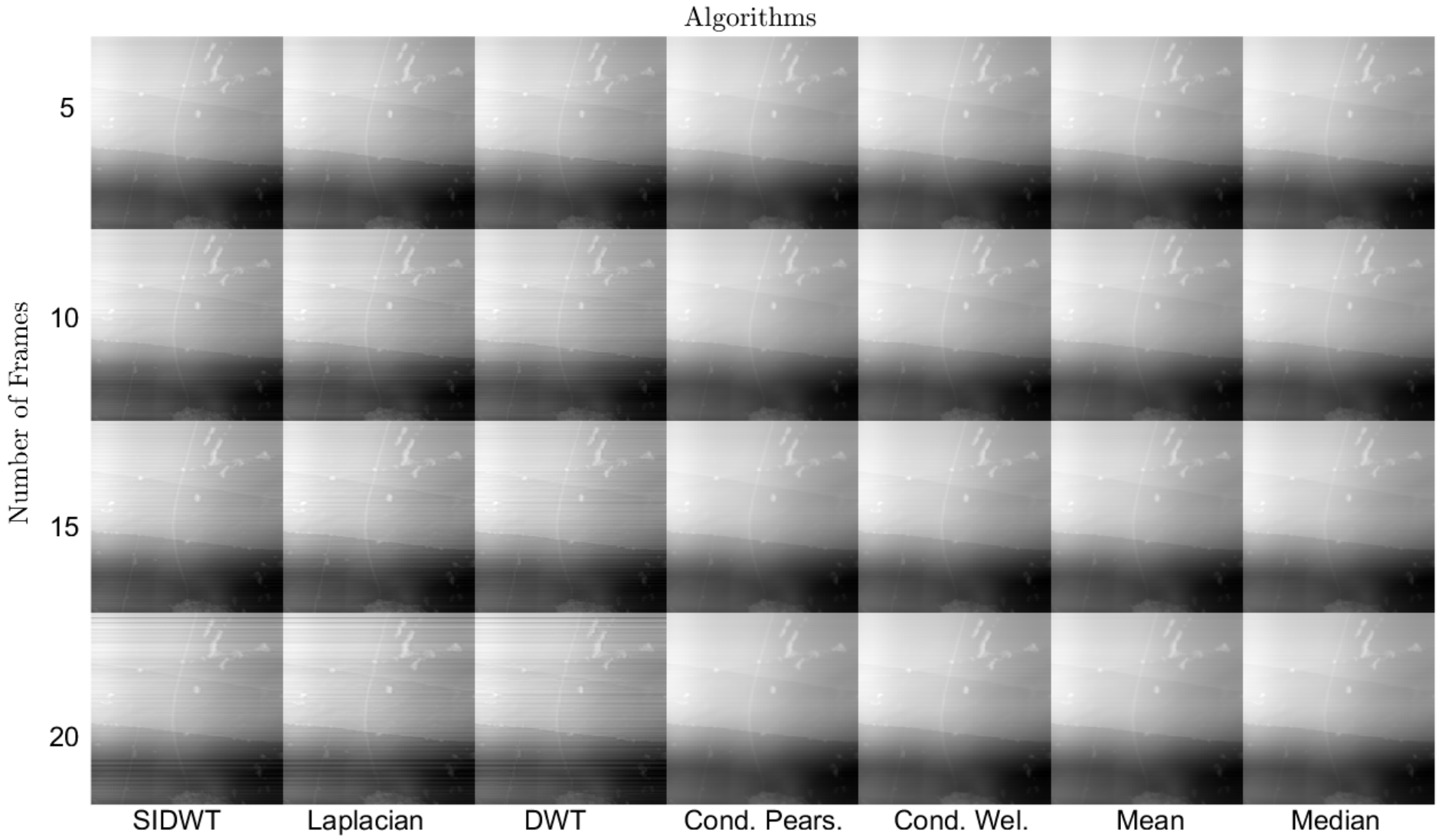


Figure 6: A montage of images generated by each of the algorithms for 5, 10, 15 *and* 20 *frames for the* ***Increased Noise unflattened data set****. The SIDWT, Laplacian, DTW become progressively worse, while the Conditional Pearson, Conditional Welford, Mean and Median all remain similar as more frames are added.*

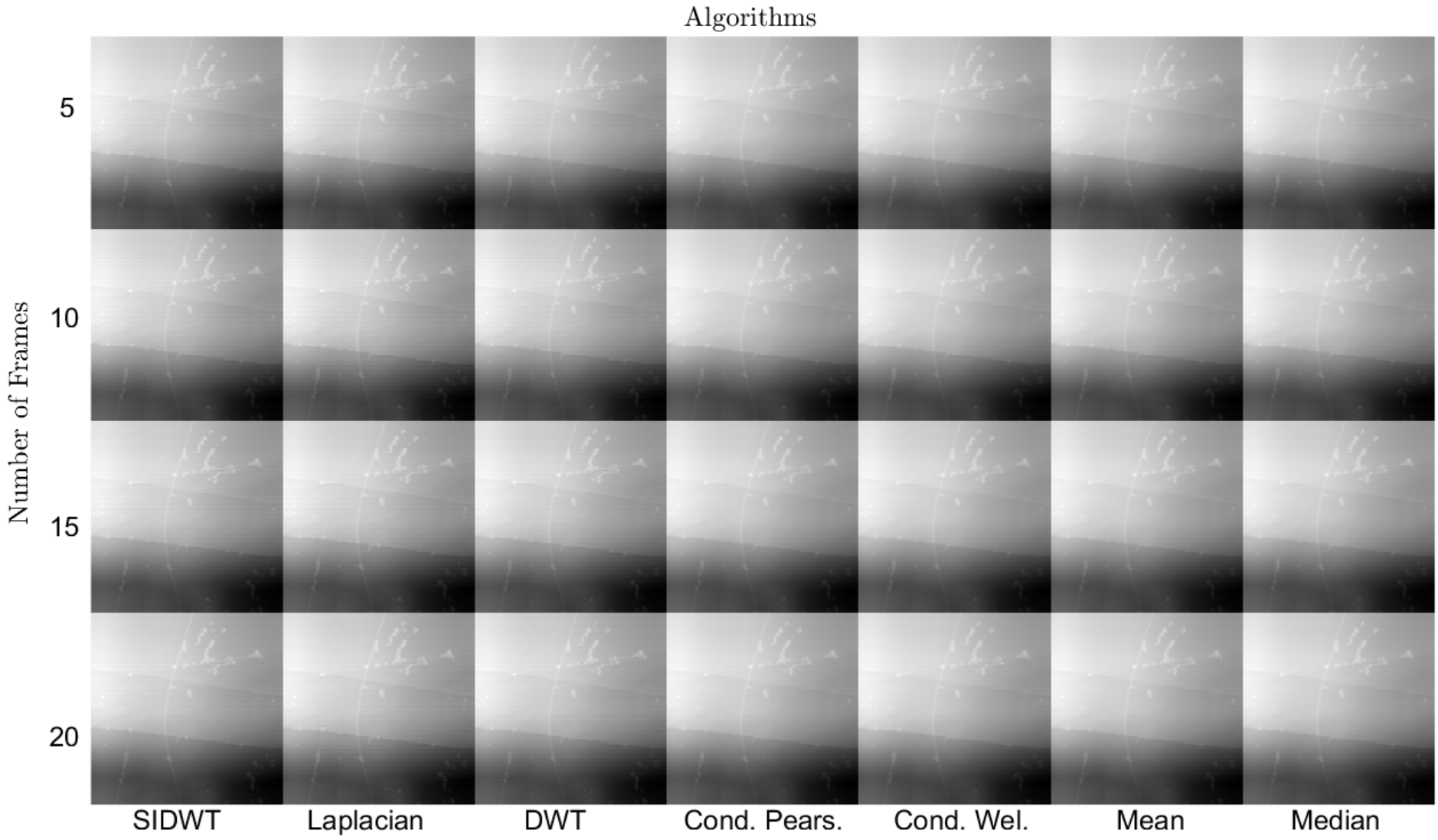


Figure 7: A montage of images generated by each of the algorithms for 5, 10, 15 *and* 20 *frames for the* ***Noisy unflattened data set****. The algorithms seem to generate similar results, however it is visible that the SIDWT, Laplacian and DTW algorithms have more horizontal lines.*

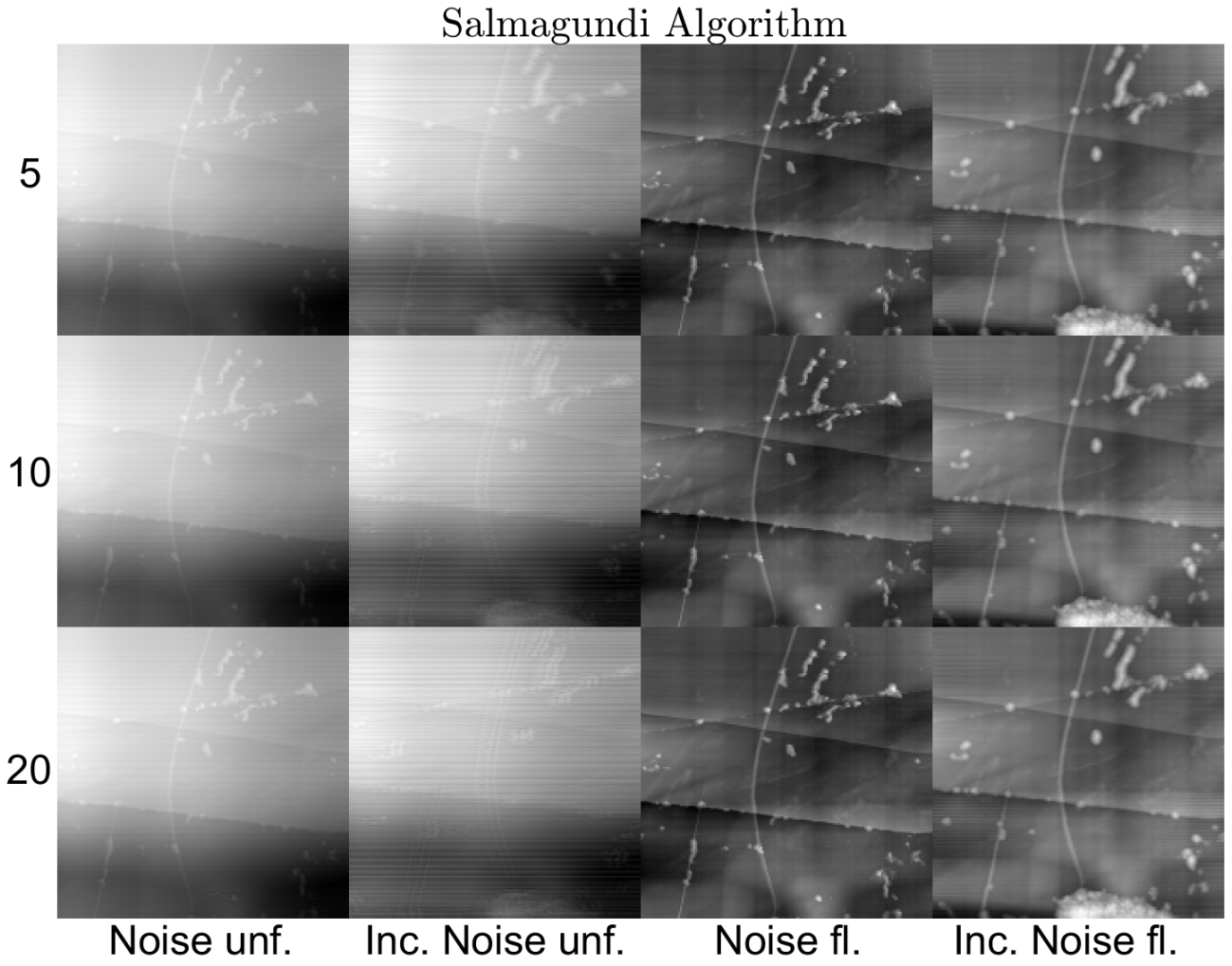


Figure 8: A montage of images generated by the Salmagundi Algorithm for 5, 10, *and* 20 *frames for a (un)/flattened data set with/(out) increased noise. From initial visual assessments the images look to be of better or equal quality than the previous algorithms in for the flattened data set but not than the conditional algorithms on unflattened images.*

8.5 Appendix E - Quantitative Results

FMI Measure

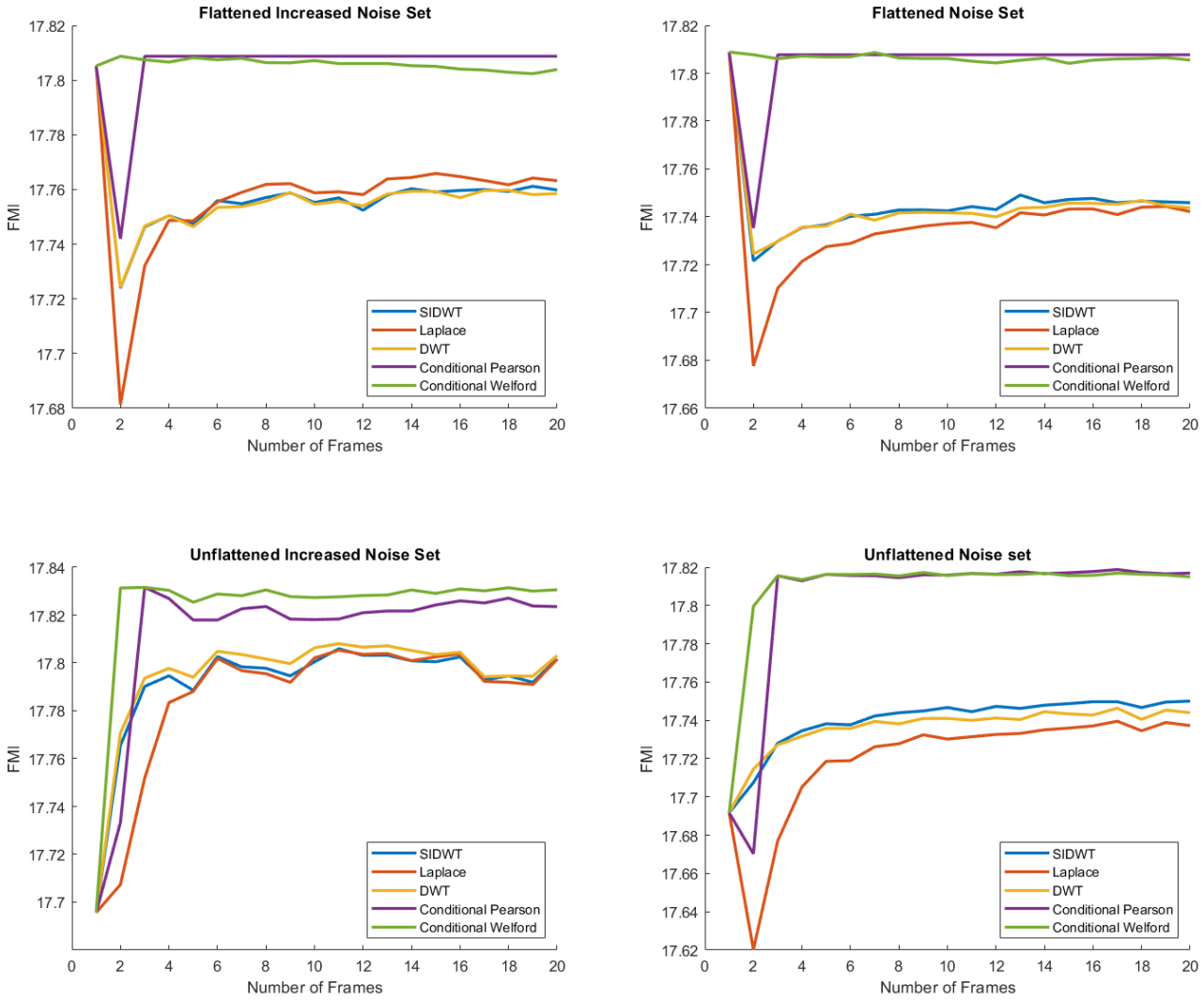


Figure 9: A collection of figures showing the FMI results for every data set over 20 frames. A higher FMI means that the fused image is more similar to mean of 20 frames. When fusing less than six image, the FMI changes by a larger amount when a new frame is added compared to past this point.

RMSD Measure

The following histograms show the spread of the difference in pixel values between each method and 20 frames of the mean. It also shows how the difference varies with a different amount of frames used in the fusion methods. The amount of frames used are 1, 5, 10, 15 and 20. With the conditions set in the algorithms, the first frame will always be the same as they all take the first frame of the set to begin the fusion method.

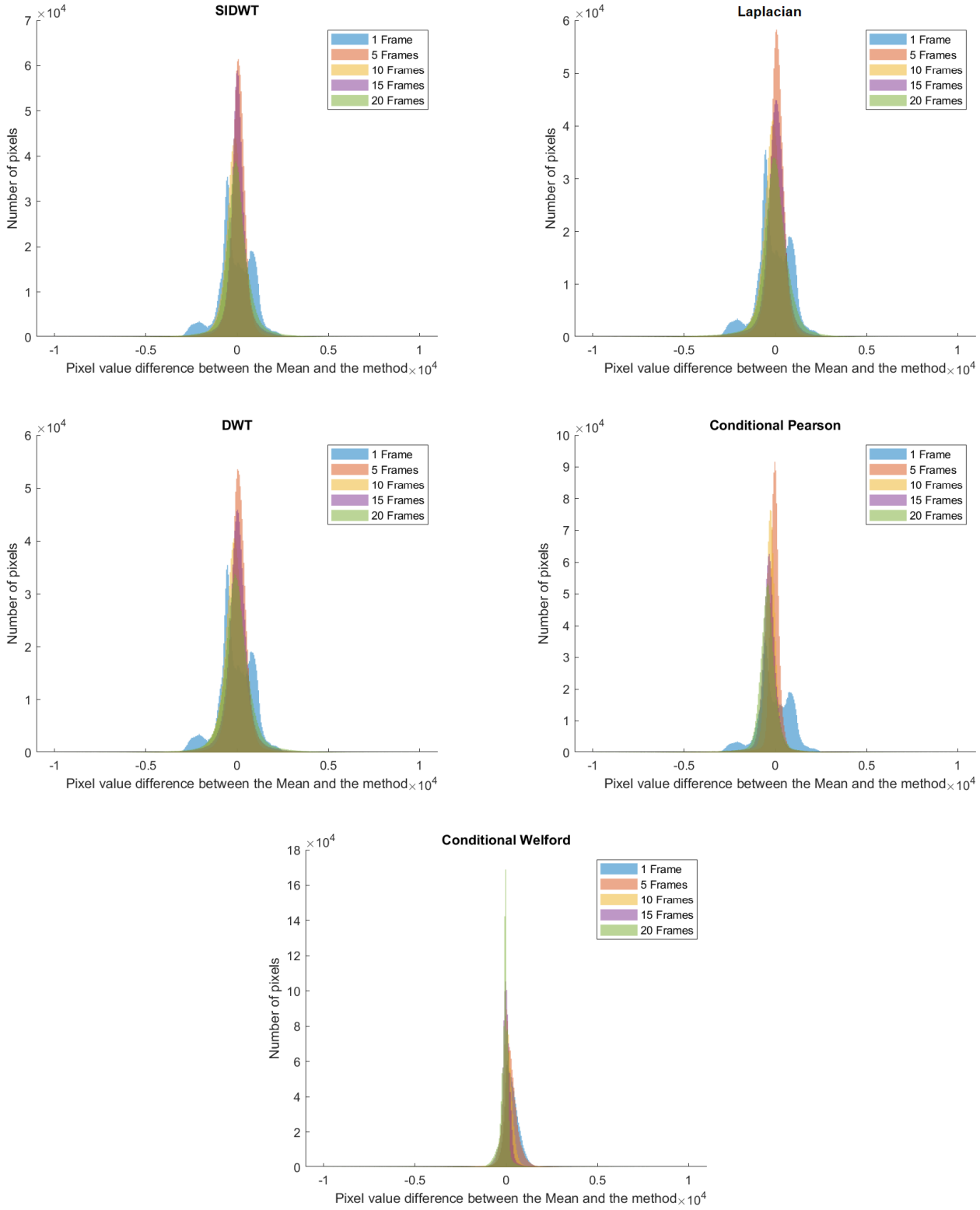


Figure 10: The frames used to produce these graphs are from the ***flattened set without increased noise****. The Conditonal Welford Algorithm is the most similar to the mean using flattened frames as the distribution of the difference is smaller at 20 frames than the other algorithms. As more frames are taken, the distribution of the differences start to converge.*

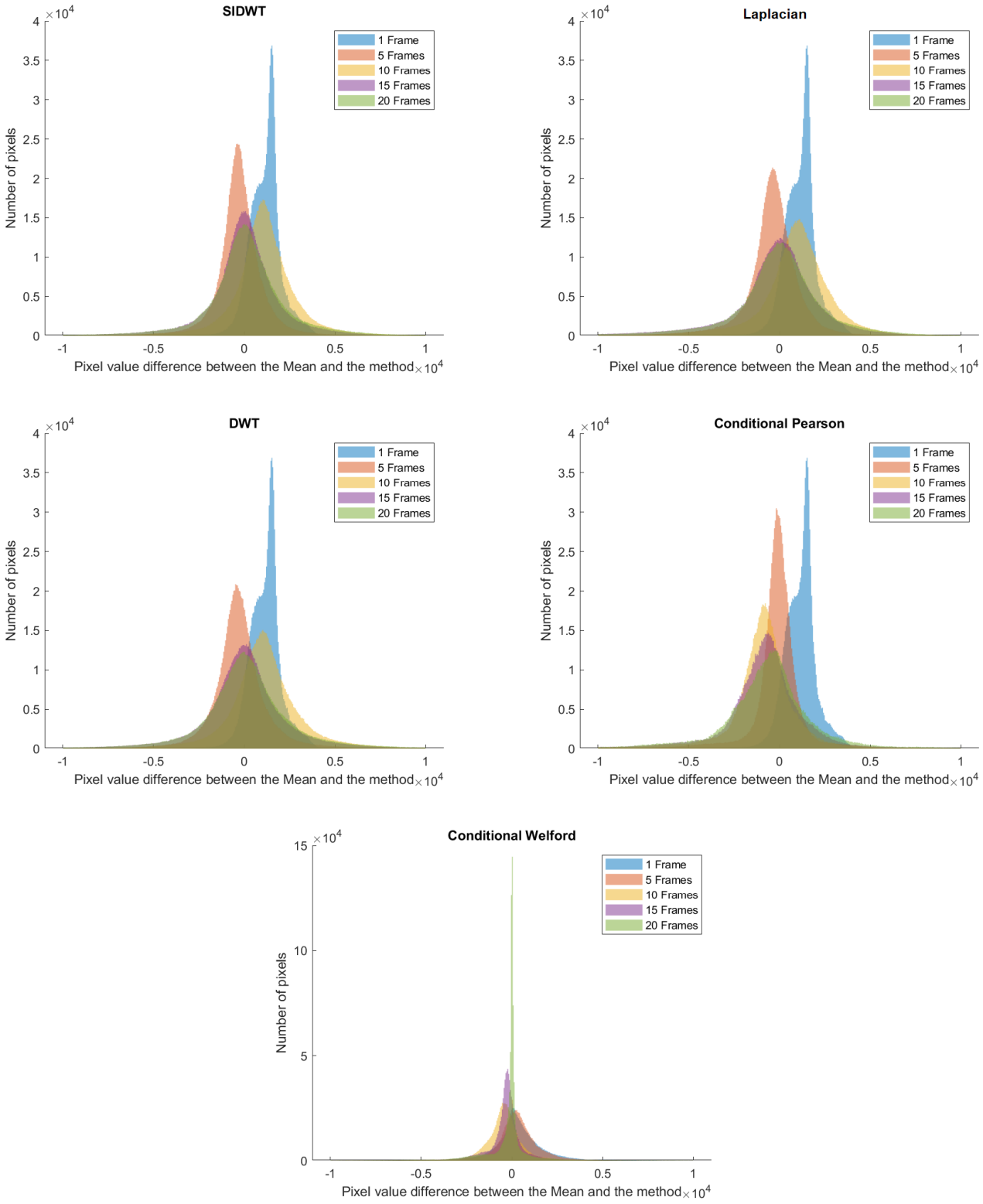


Figure 11: These graphs show the ***flattened set with increased noise****. The Conditional Welford Algorithm is the most similar to the mean which makes sense as it uses a running mean but will reject pixels outside of the condition. The other four methods are similar for 1 frame and 20 frames but the distribution varies for the other amount frames. The different algorithms have a similar distribution to the mean but that does not mean that they have the same quality.*

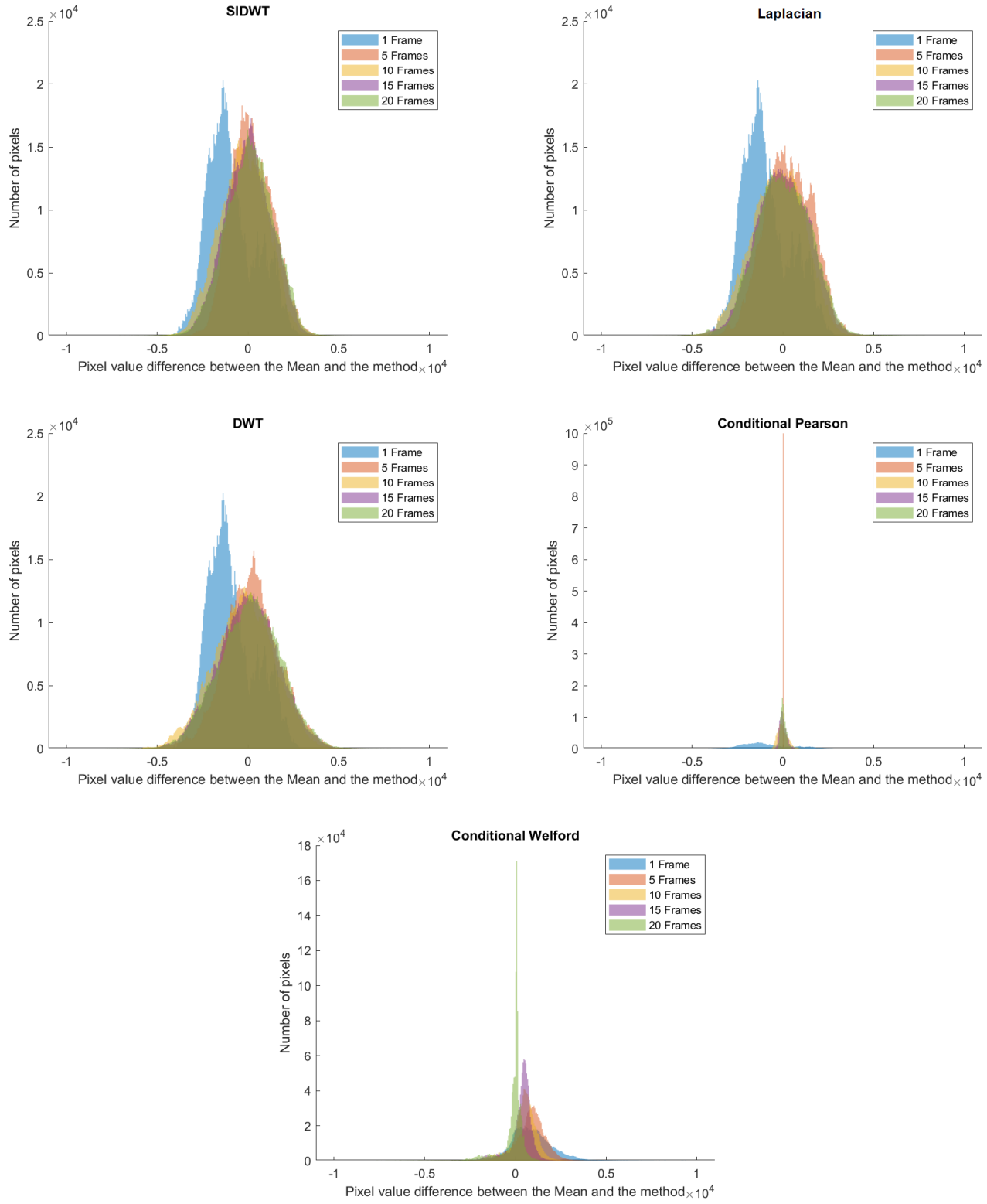


Figure 12: These histograms show the difference between the mean and the methods in the ***unflattened set with increased noise****. Here the conditional Pearson algorithm is the most similar to the mean in 5 frames but differs when more frames are added. It shows 5 frames is not enough to give an accurate image as it is likely to change with more frames added. For 20 frames the two conditional algorithms are the most similar to the mean. The spread of difference is less for the Conditional Pearson Algorithm than the other algorithms*

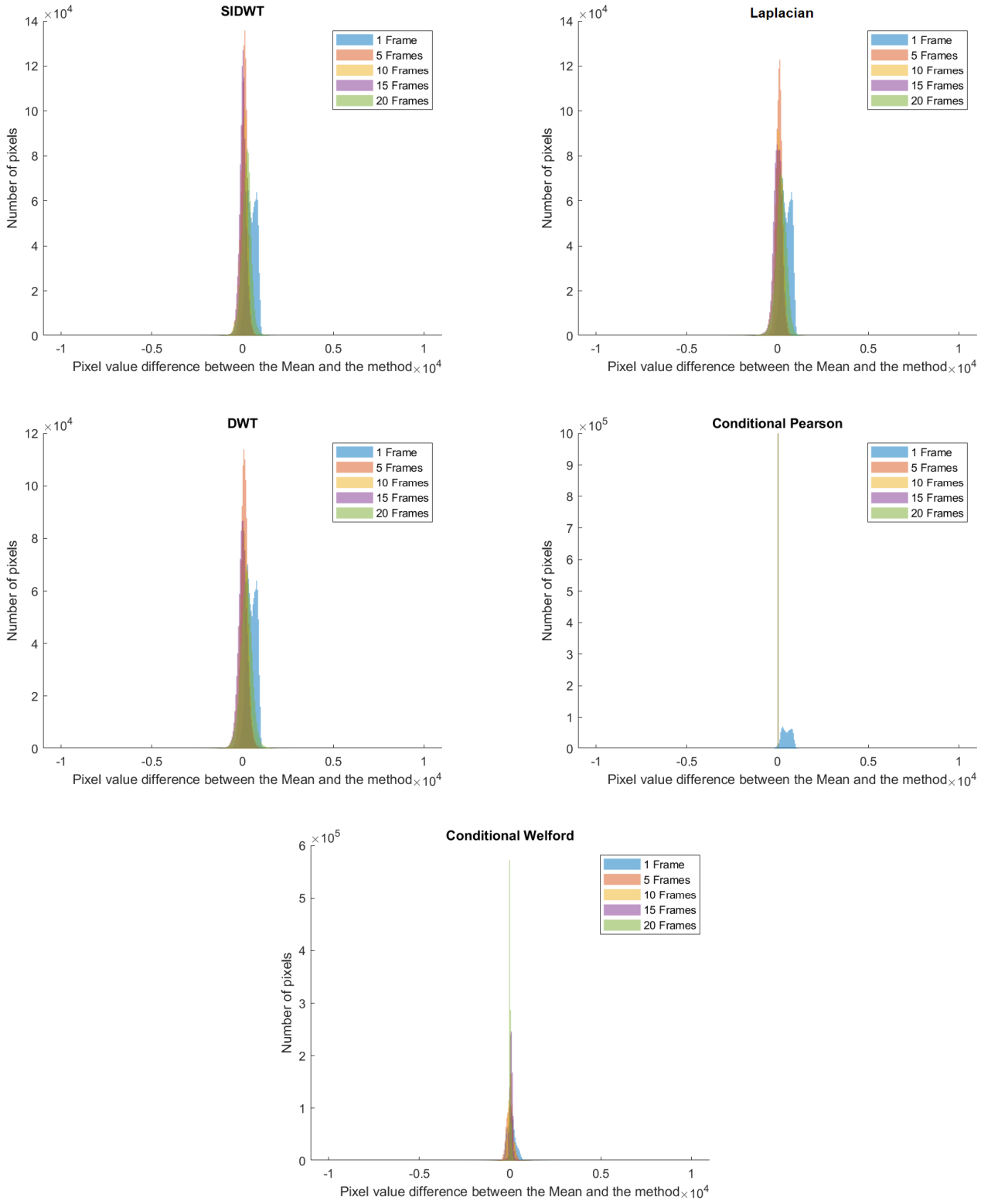
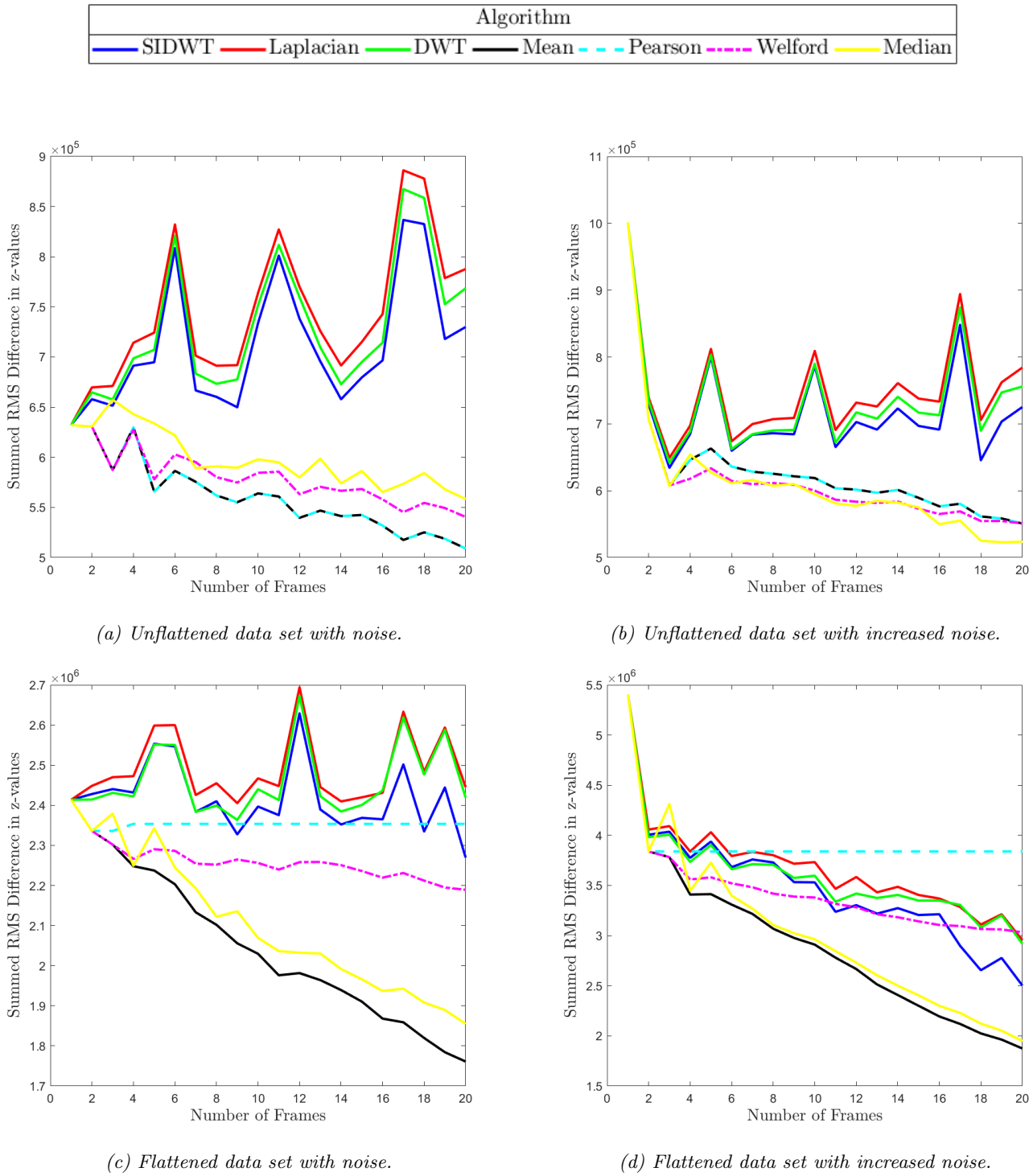


Figure 13: For these histograms the data set used is the ***unflattened set without increased noise****. Here, after 5 frames have been fused, the Conditional Pearson Algorithm is exactly the same as the mean as all one million pixels have a difference of zero. The Conditional Welford Algorithm is also only very marginally different to the mean. The other multiresolution fusion algorithms are also similar to the mean.*

Consistency Measure



(a) Unflattened data set with noise.

(b) Unflattened data set with increased noise.

(c) Flattened data set with noise.

(d) Flattened data set with increased noise.

Figure 14: Big