Methodology for Increasing the Measurement Accuracy of Image Features

Michael Majurski, Joe Chalfoun, Steven P. Lund, Peter Bajcsy, and Mary Brady
National Institute of Standards & Technology
100 Bureau Drive, Gaithersburg, MD 20899
{michael.majurski, joe.chalfoun, steven.lund, peter.bajcsy, mary.brady}@nist.gov

Abstract

We present an optimization methodology for improving the measurement accuracy of image features for low signal to noise ratio (SNR) images. By superimposing known background noise with high quality images in various proportions, we produce a degraded image set spanning a range of SNRs with reference feature values established from the unmodified high quality images. We then experiment with a variety of image processing spatial filters applied to the degraded images and identify which filter produces an image whose feature values most closely correspond to the reference values. When using the best combination of three filters and six kernel sizes for each feature, the average correlation of feature values between the degraded and high quality images increased from 0.6 *(without filtering) to* 0.92 *(with feature-specific filters), a* 53% *improvement. Selecting a single filter is more practical than having a separate filter per feature. However, this results in a* 1.95% *reduction in correlation and a* 10% *increase in feature residual root mean square error compared to selecting the optimal filter and kernel size per feature. We quantified the tradeoff between a practical solution for all features and feature-specific solution to support decision making.*

1. Introduction

Image features are computed in cell biology to extract quantitative information regarding cell state, differentiation, biological activity, and cell dynamics. The motivation for our work is the improvement of measurement accuracy for image features extracted from time-lapse fluorescent images of stem cell colonies. Due to cell sensitivity to light, only brief low intensity light could be used to excite fluorophores, producing images with low signal to noise ratios (SNRs) and hence resulting in questionable accuracy of image features.

Our objective is to mitigate the effects of image noise on extracted features via image de-noising (filtering) with respect to quantitative metrics. Quantitative imaging can play an important role in scientific experiments as a means to monitor and communicate behavior of complex systems (e.g. cell colonies) by recording features extracted from objects of interest. An image feature is a function whose input is an image. Ideally, the image itself is representative of the current state of the biological system being imaged such that changes in the images are representative of changes in the system. In such cases, image features can provide useful summaries to help monitor and communicate the systems behavior. However, when images have a low SNR, poor focus, or other distortions, the information extracted via feature evaluation may not reflect the behavior of the underlying system. That is the ability to extract meaningful image feature values is linked to the quality of the acquired images. Unfortunately, due to experimental constraints ideal high quality images can be time consuming to acquire, impractical, expensive, or damaging to the specimen. This forces the acquisition of lower quality images. Image processing algorithms can help mitigate measurement inaccuracies caused by low quality images. The difficulty lies in selecting which image processing algorithms to apply for a given feature measurement. We are interested in determining the ordered set of image processing operations that result in images whose feature values convey similar meaning to those of the same image if it was of higher quality, in other words, having clear signal and negligible noise.

In this paper, we focus on the effects of image noise, as opposed to other factors that may degrade image quality. Image features measured from images with low SNRs can be very poorly correlated with ground truth feature values, defined as those extracted from the same signal images with minimal noise. Any conclusions and insights based upon those feature measurements can be unreliable and biased. As SNR decreases the meaningful signal variations among the different images becomes lost to the noise. As a result, the extracted features begin to characterize the behavior of the noise instead of the signal. Ground truth features should be measured from very high SNR images so feature values are predominantly a function of signal only, and minimally influenced by noise. Differ-

ent types of features might display various robustness to noise. These effects should be investigated before interpreting the features. Figure 1 shows the effect of image noise, displaying a feature value scatterplot for Texture.Average.ENTROPY at an SNR of 2, plotting the measured values against the ground truth values. The red $y = x$ line with a slope of 45 $^\circ$ denotes where the measured feature value (y-axis) is equal to the ground truth feature value (x-axis). With a correlation value of $\rho = 0.584$ there is only a vague linear relationship between the measured and ground truth values. All features are drawn from Bajcsy *et al*. [1] and their mathematical definitions are available at https://isg.nist.gov/deepzoomweb/stemcellfeatures.

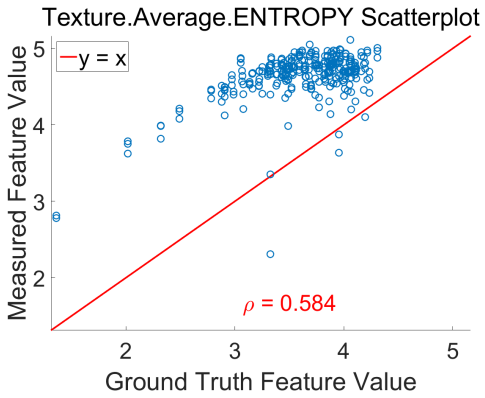

Figure 1. Example low quality (SNR of 2) image feature measurements which do not correlate well with ground truth. The correlation (ρ) between measured and ground truth feature values is 0.584. The red $y = x$ line with a slope of 45 $^\circ$ denotes where the measured feature value (y-axis) is equal to the ground truth feature value (x-axis).

Previous research has addressed either image quality assessment or the accuracy of feature based application specific results. The intermediate step of analyzing feature quality has been neglected. Image quality assessment utilizes two types of metrics, those based on quantitative deviation from ground truth (SNR, PSNR, MSE [19]) and those designed to mimic human visual perception (FSIM [19], RFSIM [18], MS-SSIM [15], SSIM [14]). In order to evaluate the quality of de-noised images researchers have created synthetic noise models and optimized the selection of de-noising algorithms over using synthetic images [2, 6, 7, 11, 12, 13, 17]. In contrast, we leverage measured images, both high quality ground truth signal and noise. We then create evaluation data based only on measured images. Our methodology of operating on image features is less dependent upon the final application problem.

This paper presents an experimental optimization methodology for improving the measurement accuracy of image features. This methodology consists of four steps: (1) sequester a small subset of the specimen and acquire a set of reference signal images with the highest quality (SNR) possible, (2) acquire a set of background noise images typical of the target experiment, (3) combine the high quality images and the background noise images to create a set of pseudo-real images with an SNR typical of the target experiment, and (4) optimize the application of an ordered set of image processing algorithms in order to maximize the correspondence in feature values measured from the reference signal images and the pseudo-real images.

The low SNR (pseudo-real) images mimic those of the target experiment but contain a known signal component. Therefore features computed from the pseudo-real images can be compared with features computed from the reference signal images. This enables the design, optimization, and refinement of the target experiment image processing pipeline with respect to accurately measuring the image features of interest. Figure 2 shows an overview of the proposed methodology. The reference signal and measured noise images are combined into a pseudo-real image typical of the target experiment. Multiple image processing pipelines are applied to the pseudo-real image generating processed images. Features are then extracted from the processed images and compared to feature values extracted from the reference signal images.

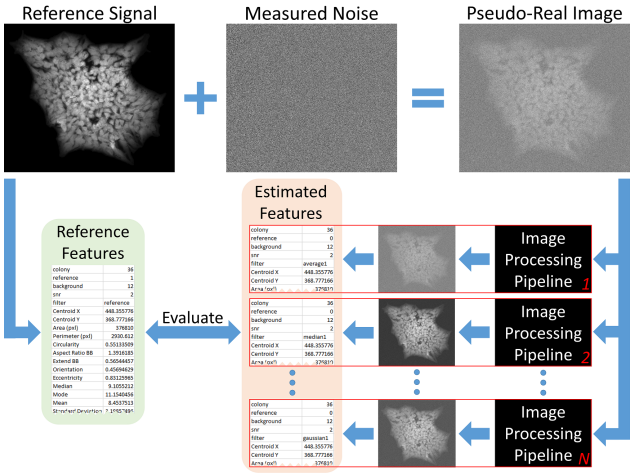

Figure 2. Overview of the methodology for improving the accuracy of measured image features. The reference signal and measured noise images are combined into a pseudo-real image typical of the target experiment. The pseudo-real image is passed through an image processing pipeline to generate a processed image. Features are extracted from the processed image and compared to features extracted from the reference signal image, enabling the evaluation of several processing pipelines to determine which improves the accuracy of the extracted features best.

Pseudo-real images (created by combining measured images) are preferable to synthetic images (created entirely from computer simulation of signal and noise models) because they are more relevant to the imaging experiment in question. Both pseudo-real images and synthetic images

contain a known signal component enabling comparison to ground truth. However, the synthetic image noise model might not match the target imaging experiment. By relying on experimentally observed reference signal and noise the pseudo-real images avoid this problem. The downside of the pseudo-real images is that any noise present in the reference signal images cannot be overcome, setting an upper limit on the improvement in feature measurement accuracy. By allowing the experimentalist to define the reference signal and the expected noise the image quality model can be reduced to just SNR, assuming the selection of reference signal images is such that the underlying population of signal profiles is sufficiently narrow that the optimal filter is principally a function of SNR and not the signal itself. In cases where that is not true, the reference image set can be refined until the assumption is met. If required, this methodology can be performed multiple times to optimize filter selection for each sub-type of signal. Other factors affecting filter selection, which might be experiment dependent, will be defined by the experimentalist as part of acquiring the reference signal and noise images.

2. Methods

This experimental methodology for increasing image feature measurement accuracy consists of two major stages. The first stage is the creation of the pseudo-real images and the second stage is the optimization and evaluation of the image processing required to increase the accuracy of the measured image features in generated pseudo-real images.

2.1. Pseudo-Real Image Creation

In order to evaluate the accuracy of extracted image features, a known reference signal similar to ones own experiment is needed. Such a signal is acquired by imaging a representative subset of the experimental specimen with the highest image quality possible. Since this sample is not being relied upon for the actual experiment feature data, it is not subject to the same experimental constraints which limit image quality. For example, a longer exposure time than is practical for the real experiment could be used. These high quality images are called the reference signal because they contain the reference foreground signal from which the ground truth image features are extracted.

The next step is to acquire a sample of the background noise expected in the target experiment. These background images should mimic all noise sources in the target experiment while containing no foreground data. This is done by imaging a background area with the same acquisition setup as used in the actual experiment.

The pseudo-real images are created by combining the reference signal and measured background noise images. Thus the pseudo-real images mimic the expected target experiment images while containing a known reference signal component. They are created by multiplying the reference signal images by a scalar and then adding the background noise. The following procedure is used to create the pseudo-real images with a desired SNR.

1. Compute the mean of the foreground pixels (as identified by segmentation), μ_F, from the reference signal image, I_F.
2. Compute the standard deviation of the background pixels, σ_B, from the background noise image, I_B.
3. Given a desired SNR value of k for the pseudo-real image, compute the rescale factor $(k * \sigma_B)/\mu_F$.
4. Multiply the reference signal image by the rescale factor then add the background images, $I_k = \frac{k*\sigma_B}{\mu_F} \times I_F + I_B$.

This pseudo-real image creation procedure is shown in Figure 3. The formula used to identify the rescale factor for a specified SNR is derived from the Rose criterion as described in [16].

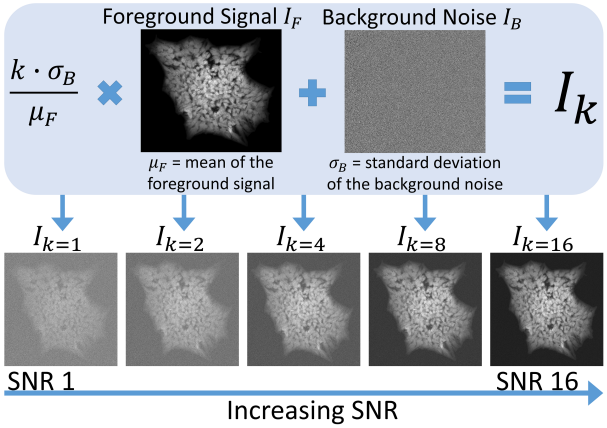

Figure 3. Diagram outlining the pseudo-real image creation. The reference signal (foreground) image is scaled by a factor $(k * \sigma_B)/\mu_F$ and added to the measured background noise image to create a pseudo-real image with a specified SNR value. Note: all displayed images were auto-contrasted using the same algorithm for visual clarity.

2.2. Process Optimization

We then determine an ordered set of image processing operations required to improve the feature measurement accuracy. Ground truth feature values are measured from the reference signal images, denoted *Reference*. These are then compared with features measured from the pseudo-real images after processing, denoted *Processed*. Correlation (ρ), described by Eq. (1), is used to compare sets of feature values.

$$\rho = \frac{cov(Processed, Reference)}{sd(Processed) * sd(Reference)} \quad (1)$$

The correlation coefficient measures the linear relationship strength between the processed feature values and their corresponding reference values. A correlation coefficient of 1 indicates that there are scalar values $a > 0$ and b such that the equation $Reference = a*Processed+b$ holds exactly. When primary interest lies in the relative sizes of pairwise distances among any set of points, as opposed to the exact values of the features themselves, correlations near 1 can be interpreted as meaning the two quantities are nearly equivalent.

3. Experimental Results

This section presents experimental results obtained by applying the proposed optimization methodology to cellular microscopy images. There are many factors involved in keeping cells alive, but with respect to imaging, prevention of photo-toxicity is the important one. It is desirable to have the highest SNR images possible to discern biological information with the highest fidelity. However, acquiring these high SNR images requires exciting the fluorophores within the cells with high-intensity light. This exposure damages or kills the cells (due to photo-toxicity) and causes accelerated photo-bleaching of the specimen. Note that we are mainly concerned with fluorescent imaging modalities where the sample must be probed with excitation light. For transmitted light modalities photo-toxicity is less of a concern. Thus, our primary interest lies in strategies to balance the competing interests of using minimally invasive imaging techniques to avoid affecting cell behavior or survival and acquiring high quality images which contain the required information.

3.1. Measured Microscopy Images of Cells

The target imaging experiment consists of a time-lapse acquisition of the H9 human embryonic stem cell (hESC) line over the course of 5 days on a microscope equipped with a controlled environment incubation chamber (Kairos Instruments LLC, Pittsburgh, PA). This cell line was engineered to produce green fluorescent protein (GFP) under the influence of the native OCT-4 promoter using a published homologous recombination plasmid construct developed by the James Thomson lab [20] and obtained from Addgene (Addgene, Cambridge, MA). Experimental imaging is performed using a Zeiss 200M microscope (Carl Zeiss Microscopy, LLC, Thronwood, NY) every 45 minutes via a Coolsnap HQ camera (Photometrics, Tucson, AZ) in a grid of 16×22 field of views (FOVs) with 10% overlap covering approximately 180 mm^2. Each individual FOV (image) is 1040×1392 pixels.

The individual target experiment images are stitched into a single mosaic per time point using MIST (Microscopy Image Stitching Tool) [3]. Foreground and background masks are generated by segmenting the phase-contrast stitched images using the Empirical Gradient Threshold technique [5]. The stitched mosaic images are flat-field corrected and background subtracted [4]. Using the foreground masks a set of 61 intensity and texture image features, taken from [1], are extracted from each colony. The intensity features are statistical moments: mode, mean, mode, standard deviation, skewness, kurtosis, etc. The texture features are based on Haralick texture features [10] which generate four values per feature type, the average amplitude, principle component angle, orthogonal component angle, and principle component value.

Since this is a time-lapse experiment, the cells need to be kept alive and minimally disturbed by the high intensity light used in imaging. Therefore, experimental conditions constrain imaging to phase contrast (less-damaging transmitted light) and low SNR fluorescent imaging. Many regions of interest exhibit SNRs of roughly 2. The goal of the cell imaging is to classify stem cell colonies based on homogeneity and to analyze the homogeneity distribution of these colonies through time. The classification of cell colonies is based on the intensity and texture features extracted from the fluorescent images. Therefore, it is important to compute the features with the highest accuracy possible under these circumstances. We apply the proposed methodology on this problem to find the optimal image processing steps that increase the accuracy of the measured features.

3.2. Pseudo-Real Image Creation

For this application the reference signal image dataset consists of 100 stem cell colonies imaged in the fluorescent channel with a long exposure time and high power excitation light to create very high SNR images. All of these colonies fit within a single FOV and are larger than 1000 pixels in area. It is important to note that in acquiring these images with the aforementioned acquisition parameters, the colonies were both damaged and photo-bleached, making this acquisition method unsuitable for the target time-lapse experiment.

Typical background noise for the target experiment is acquired by imaging the specimen background consisting of cell culture media, culture dish, and any extracellular matrix protein coatings under the same acquisition parameters as the real experiment. In addition to any background autofluorescence, the CCD camera noise is captured in these background images. We acquired 30 background images with different spatial locations on the plate typical of the conditions expected in the target time-lapse imaging experiment.

Conditional random sampling is applied to the set of

100 reference signal colony images and 30 measured background noise images to produce the set of pseudo-real images. Each colony image is combined with 3 background images. Each background image is selected 10 times for a total of 300 combinations. The subsampling, as opposed to a complete factorial design, is used to restrict computational requirements to a reasonable level. Next, each colony image containing ideally pure signal is combined with its selected backgrounds to create 5 target SNR levels (1, 2, 4, 8, and 16). The colony images were segmented using a manually selected threshold (foreground is greater than 500 intensity units) to set the background of the image to 0. Before this adjustment the reference signal image background (non-colony pixels) contained just dark current noise from the CCD camera with intensity values of approximately 200 units. The colony foreground contains pixels of approximately 4000-8000 grayscale intensity units coming from a 14bit CCD camera with an output range of 0-16284 intensity units.

3.3. Optimization of Image Processing Filters

For this application we are interested in selecting the spatial image processing filter and kernel size for each feature which produces the most accurate measurement of that feature. While this methodology enables the design and optimization of arbitrary image processing pipelines with respect to feature measurement accuracy, we have limited the complexity of the processing pipeline to a depth of one operation and a small set of manually selected spatial image filters (Average, Median, or Gaussian) [8, 9]. These filters were chosen because they are commonly used methods of reducing image noise. Each filter is parameterized by a kernel size of which six were tested (3x3, 5x5, 7x7, 9x9, 13x13, 17x17). In order to evaluate the image processing, each feature was computed for each combination of filter type and kernel size.

3.4. Numerical Results

Each filter and kernel size combination is applied to the pseudo-real images and all 61 features are extracted from the processed images. This enables the analysis of how the feature values change as a function of the image filter, kernel size, and image SNR. The target experiment of this study has an expected image SNR of 2. The optimal filter and kernel size can be selected for each feature by selecting the filter and kernel which maximizes the correlation in Eq. (1) between the processed and reference feature values.

Applying the filter selected for each feature increases the average correlation from 0.601 to 0.919. Of the 61 features evaluated, 77% are optimized with the Gaussian filter, 18% with the Average filter, 3.3% with the Median filter, and 1.6% with No Filter. Kernel sizes 5x5 and 7x7 are the most common at 20% and 61% respectively. The majority of the features (57%) have the same optimal filter and kernel size, 7x7 Gaussian. Figure 4 shows a histogram of the feature correlation values for no filter and the optimal filter per feature, highlighting the increase in correlation.

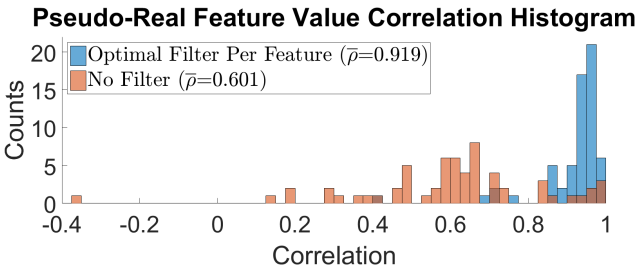

Figure 4. Histogram of the extracted feature correlation with ground truth values for pseudo-real images with an SNR of 2. Average correlation ($\bar{\rho}$) is listed in the legend.

The feature correlation (Figure 4) without filtering has an average correlation of 0.601 and only a few features with correlations above 0.8. Once filtering has been performed the majority of the features have correlations with ground truth above 0.8. There are two groups of features that do not respond well to filtering. The first group contains just the statistical moment feature Mode ($\rho \approx 0.4$). The second group contains all of the Haralick principle component angle texture features. All optimal filter per feature correlation values below 0.9 are principle component angle texture features with the exception of Mode. Without these two groups the optimal filter per feature average correlation is 0.942.

By averaging correlation across all features a single optimal image processing filter, the 7x7 Gaussian, can be found for this experiment. Doing this results in a slight loss in average accuracy compared to selecting the optimal filter for each feature. Among the features whose per feature optimal filter differs from 7x7 Gaussian there is a 1.95% loss in average correlation and a 10.04% increase in average feature residual RMSE Eq. (2).

$$residualRMSE = \sqrt{\frac{\sum_{i=1}^{N}(Proc_i - Ref_i)^2}{N}} \qquad (2)$$

The correlation metric selects the filter which results in the strongest linear relationship between the ground truth feature values and the processed feature values. If exact feature values are required a linear transformation can be applied to the processed feature values. This is demonstrated in Figure 5 where the best filter for the feature Texture.Average.ENTROPY at an SNR of 2, 5x5 Gaussian, results in a bias in the processed feature values. This bias is corrected with a linear transformation (slope $a = 1.143$ and intercept $b = -1.115$) reducing the residual RMSE from 0.545 to 0.135.

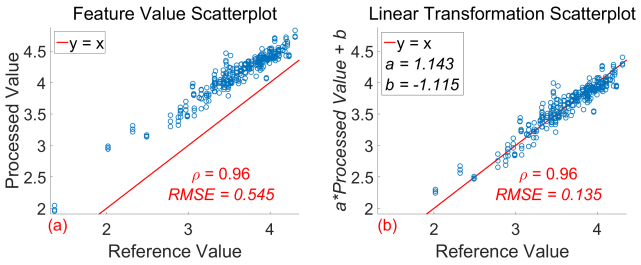

Figure 5. Feature Texture.Average.ENTROPY (SNR of 2) processed with the 5x5 Gaussian filter. The original processed feature values are shown in (a) with a bias residual RMSE of 0.545. The linear transformation of the processed feature values is shown in (b) with a lower residual RMSE value of 0.135.

To examine the relationships between the processed image feature values and the ground truth feature values a series of exploratory plots were generated. The first, shown in Figure 6, contains scatterplots of the processed feature values plotted against the ground truth feature values. This figure is organized into a two dimensional grid of scatterplots. Within each plot the feature value for an individual image is shown as a single point and the line marks $y = x$, where the processed value equals the reference value. The text superimposed on each scatterplot is the corresponding correlation coefficient, see Eq. (1).

With no filter applied (indicated by the 1x1 kernel size) there is a clear bias in the computed features that decreases with processing. As the kernel size increases the correlation values increase and the feature values show a reduction in bias. The effects of different image filter types is most evident in the 3x3 kernel size plots. Of the 3x3 filters, the Average filter has th least bias and highest correlation. Moving across the row of 3x3 kernel size scatterplots, the correlation decreases and the distribution gets further from the $y = x$ line. Increasing the kernel size reduces the disparity in results between filter types. The optimal filter for this feature is Gaussian with kernel size 5x5.

Due to the high dimensionality of the feature accuracy data it is hard to conceptualize the full picture. Therefore, a summary plot was created where the correlation values previously printed on the scatterplot are plotted as a function of image feature, filter type, kernel size, and image SNR. This plot is shown in Figure 7 and can be found in supplementary document 1. Each image SNR block contains 4 sub-blocks, the Gaussian filter block 'Gau', the Median filter block 'Med', the Average filter block 'Ave', and the No filter block 'None'. Within each filter block, the kernel size increases from bottom to top, 3 to 17. Per column within each SNR block the maximum correlation value is shown by printing the relevant kernel size. Figure 7 shows that there is considerable variability in the optimal filter and kernel size between different features. Overall, as the image SNR increases the optimal kernel sizes shrink.

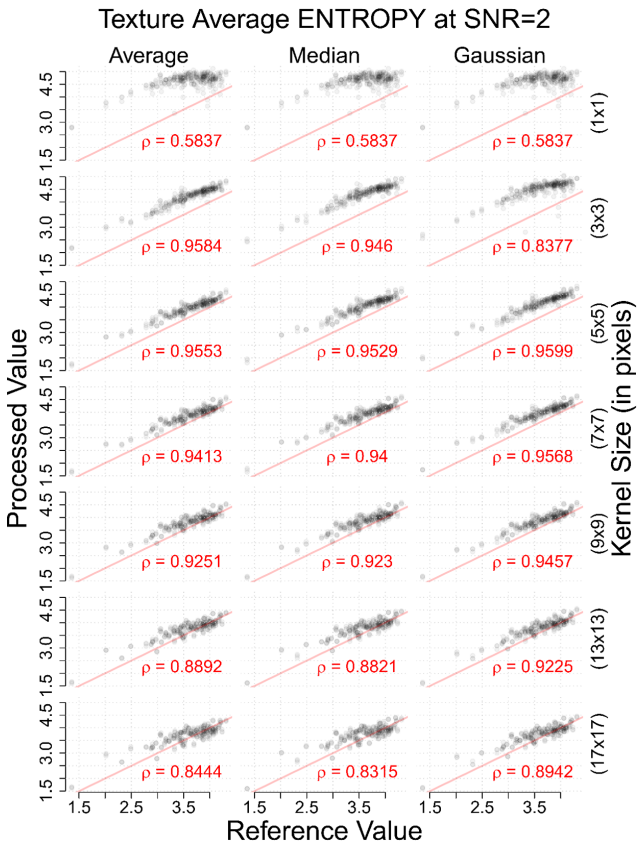

Figure 6. Feature value scatterplots for the feature Texture.Average.Entropy at an SNR of 2 given the different filter types and kernel sizes. This figure is organized into a two dimensional grid of plots. Within each plot the feature value for an individual image is shown as a single point. The line marks $y = x$, where the processed feature value equals the reference feature value. The superimposed text on each scatterplot is the correlation coefficient (ρ) for that scatterplot.

Reducing the dimensionality of the data once more is done by averaging correlation values across all features to produce a single value per filter type, image SNR, and kernel size. This creates plots where feature correlation is shown as a function of kernel size for each image SNR and filter type. Figure 8 depicts plots of these feature correlations processed with Average, Median, and Gaussian filters, where each point shows correlation averaged across all 61 features.

4. Discussion

There are several general observations that can be gleaned from Figure 8. First, a dominant factor in determining the processed feature measurement accuracy is the image SNR. The higher the acquired image SNR the more accurate the feature measurement which is logical since higher SNRs have less noise to distort the feature measure-

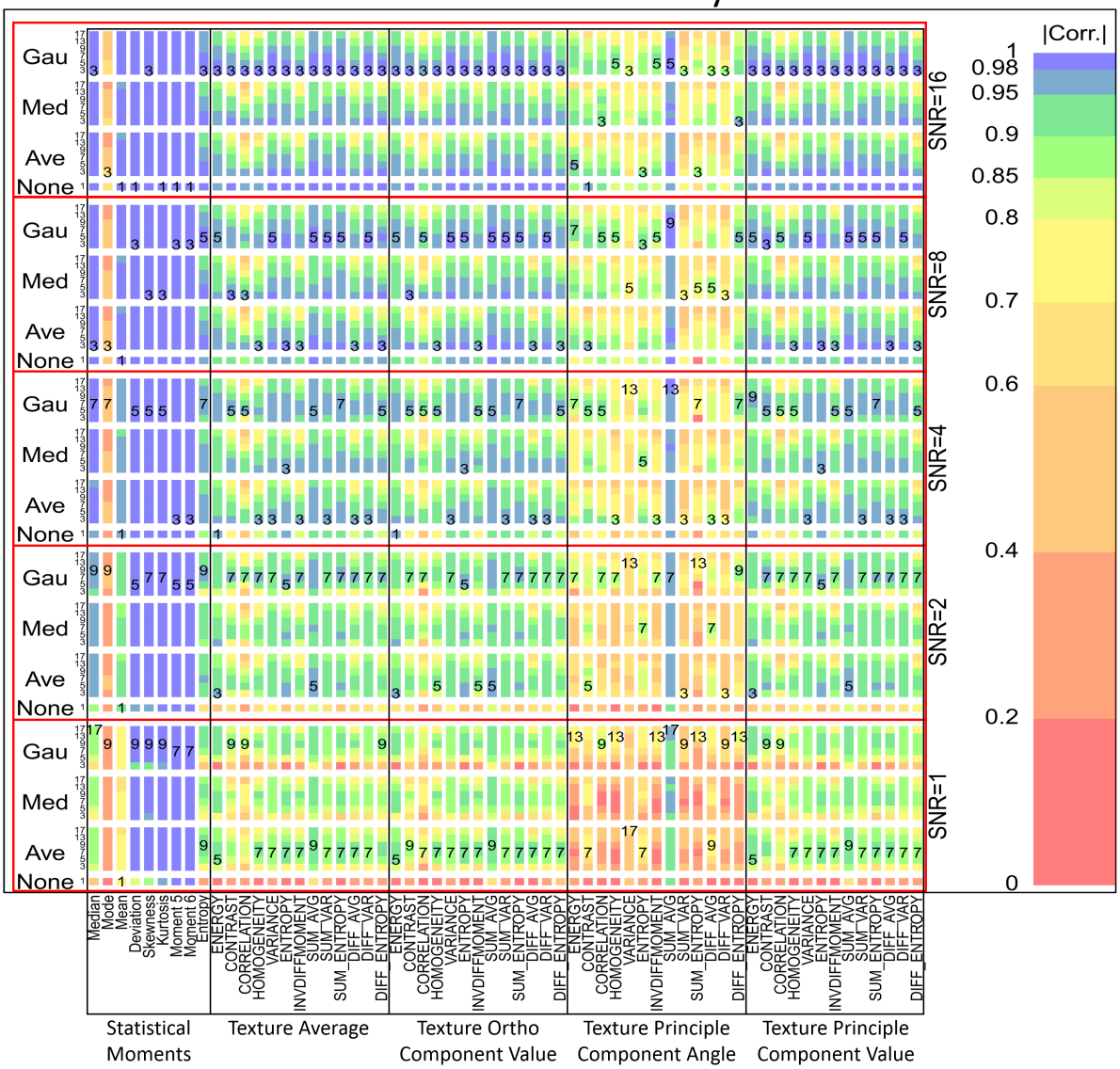

Figure 7. Correlation summary plot. Each image SNR block contains 4 sub-blocks, the Gaussian filter block 'Gau', the Median filter block 'Med', the Average filter block 'Ave', and the No filter block 'None'. Within each filter block, the kernel size increases from bottom to top, 3 to 17. Per column within each SNR block the maximum correlation value is shown by printing the relevant kernel size.

ments. If the image SNR is high enough there is little to no accuracy gained by filtering the images. For example, at an SNR of 16 a 3x3 kernel provides a minor increase in feature measurement accuracy, but a 5x5 kernel provides equal or worse accuracy than no filter. Second, as the image SNR decreases, larger filter kernels are required to obtain a given level of feature measurement accuracy. For example, at an SNR of 4 a 3x3 Gaussian kernel produces roughly the same accuracy as a 5x5 Gaussian kernel at an SNR of 2. Third, Gaussian filters require a larger kernel size to accomplish the same effect as the Median or Average filters.

The time-lapse stem cell colony imaging experiment presented here has an SNR of approximately 2. Given that constraint, the optimal image filter and kernel size for each feature should be selected such that the correlation between the processed feature values and ground truth feature values is maximized. The per feature filter selection accuracy data is available in supplementary document 2 for each image SNR level. Looking at just the filter type selection for an SNR of 2, 77% are optimized with the Gaussian filter, 18% with the Average filter, 3.3% with the Median filter, and 1.6% with No Filter. The optimal kernel size distribution is

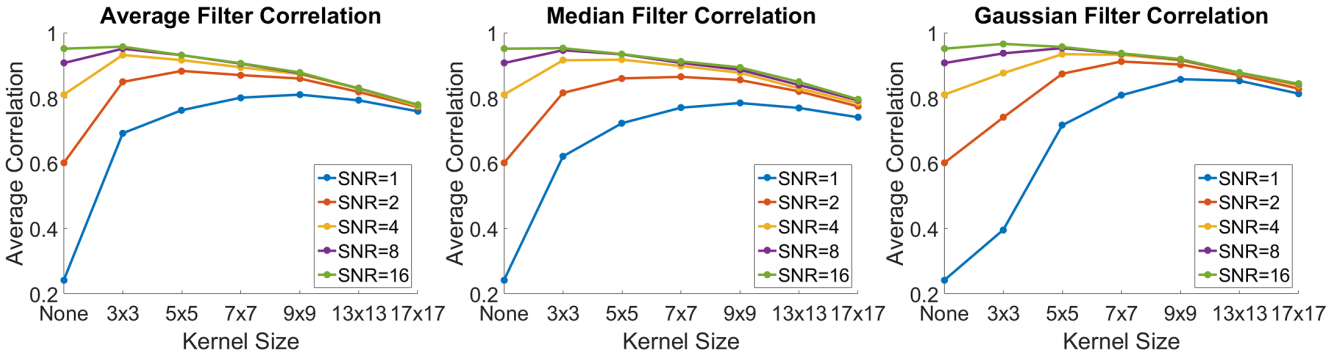

Figure 8. Average feature correlation values for each Filter, SNR, and Kernel size combination. The first plot was processed with an Average filter, the second a Median filter, and the third a Gaussian filter. Within each plot correlation is shown as a function of kernel size for multiple SNR values.

more spread out with 60.65% being 7x7, 19.67% 5x5, 8.2% 3x3, 6.56% 9x9, 3.23% 13x13, 1.64% No Filter, and 0.0% 17x17. The most common filter and kernel size combination is 7x7 Gaussian which is optimal for the majority of the features (57%). This effect shows up in Figure 7 as a fairly consistent row of '7's written within the 'Gau' block of 'SNR=2'.

Texture features which compute a principle component directionality angle did not improve nearly as much as the other evaluated features. These features accounted for all but 1 of the features that did not obtain a correlation of 0.9 or greater under any considered filter. This shows up in Figure 7 as a vertical block of lower correlation values across all SNRs.

Whether one selects a single image processing filter for the entire experiment or a filter per feature, these results are only relevant for the target experiment under consideration. The numerical results cannot be generalized but the methodology can. Changes in the target experiment (different cell line, different features, etc.) would require this pre-experiment to be redone in order to find the optimal filter(s) and kernel size(s) for the new target experiment. The power of this approach is its flexibility and extensibility. This optimization methodology can be applied to different experiments, image conditions, image modalities, and image features. For small scale experiments it might not be reasonable to perform such a pre-experiment to help design the target experiment and its data processing. However, as long as the pre-experiment does not constitute an unreasonable effort, it can help inform the accuracy of the target experiment.

5. Conclusions

This work was motivated by a desire to understand the impact on stem cell colony classification when using image features derived from low SNR images. We devised a methodology to quantify the improvement of feature measurement for a given image pre-processing method. As a proof of concept, we chose three basic filtering techniques as pre-processing steps. We found that selecting the best filter per feature produces a 53% improvement in feature correlation with ground truth, from 0.6 to 0.92. Selecting a single filter for all features results in a 1.95% reduction in correlation and a 10% increase in residual RMSE.

6. Future Work

We intend to measure the impact of using image features derived from pre-processed images on colony classification accuracy. The pool of image processing operations is going to be expanded to include more advanced image enhancement and noise reduction algorithms.

7. Acknowledgments

This work has been supported by NIST. We would like to acknowledge the team members of the computational science in biological metrology project at NIST for providing invaluable inputs to our work. We would also like to thank specifically Kiran Bhadriraju, Greg Cooksey, Michael Halter, John Elliot, and Anne Plant from Biosystems and Biomaterials Division at NIST for acquiring the image datasets.

8. Disclaimer

Commercial products are identified in this document in order to specify the experimental procedure adequately. Such identification is not intended to imply recommendation or endorsement by the National Institute of Standards and Technology, nor is it intended to imply that the products identified are necessarily the best available for the purpose.

References

[1] P. Bajcsy, A. Vandecreme, J. Amelot, P. Nguyen, J. Chalfoun, and M. Brady. Terabyte-Sized Image Computations on

Hadoop Cluster Platforms. *2013 IEEE International Conference on Big Data*, pages 729–737, oct 2013.

[2] S. Bharadwaj, H. Bhatt, M. Vatsa, R. Singh, and A. Noore. Quality Assessment Based Denoising to Improve Face Recognition Performance. In *Computer Vision and Pattern Recongnition Workshops (CVPRW)*, pages 169–174, 2011.

[3] T. Blattner, J. Chalfoun, B. Stivalet, and M. Brady. A Hybrid CPU-GPU System for Stitching of Large Scale Optical Microscopy Images. *International Conference on Parallel Processing (ICPP)*, 2014.

[4] J. Chalfoun, M. Majurski, K. Bhadriraju, S. Lund, P. Bajcsy, and M. Brady. Background Intensity Correction for Terabyte-Sized Time-Lapse Images. *Journal of Microscopy*, 257(3):226–238, 2015.

[5] J. Chalfoun, M. Majurski, A. Peskin, C. Breen, P. Bajcsy, and M. Brady. Empirical Gradient Threshold Technique for Automated Segmentation Across Image Modalities and Cell Lines. *Journal of Microscopy*, 260(1):86–99, 2015.

[6] M. Elad and M. Aharon. Image Denoising Via Sparse and Redundant Representations Over Learned Dictionaries. *IEEE Transactions on Image Processing*, 15(12), 2006.

[7] R. Eslami and H. Radha. Translation-Invariant Contourlet Transform and its Application to Image Denoising. *IEEE Transactions on Image Processing*, 15(11):3362–3374, 2006.

[8] R. C. Gonzalez and R. E. Woods. *Digital Image Processing*. Prentice-Hall, New Jersey, 2nd edition, 2008.

[9] R. C. Gonzalez, R. E. Woods, and S. L. Eddins. *Digital Image Processing Using Matlab*. Pearson Prentice Hall, New Jersey, 2004.

[10] R. M. Haralick, K. Shanmugam, and I. Dinstein. Textural Features for Image Classification. *IEEE Transactions on Systems, Man, and Cybernetics*, 3(6), 1973.

[11] B. Matalon, M. Elad, and M. Zibulevsky. Improved Denoising of Images Using Modelling of a Redundant Contourlet Transform. *Optics & Photonics 2005*, 2005.

[12] J. Portilla, V. Strela, M. J. Wainwright, and E. P. Simoncelli. Image Denoising Using Scale Mixtures of Gaussians in the Wavelet Domain. *IEEE Trans Image Processing*, 12(11):1338–1351, 2003.

[13] J.-L. Starck, E. J. Candès, and D. L. Donoho. The Curvelet Transform for Image Denoising. *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, 11(6):670–84, 2002.

[14] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image Quality Assessment: From Error Visibility to Structural Similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004.

[15] Z. Wang, E. P. Simoncelli, and A. C. Bovik. Multi-Scale Structural Similarity for Image Quality Assessment. *IEEE Asilomar Conference on Signals, Systems and Computers*, 2, 2003.

[16] R. Watts, Y. Wang, P. A. Winchester, N. Khilnani, and L. Yu. Rose Model in MRI: Noise Limitation on Spatial Resolution and Implications for Contrast Enhanced MR Angiography. In *Intl. Society Mag. Reson. Med.*, volume 4, page (8) 462, 2000.

[17] K. Youssef, N. N. Jarenwattananon, and L. S. Bouchard. Feature-Preserving Noise Removal. *IEEE Transactions on Medical Imaging*, 34(9):1822–1829, 2015.

[18] L. Zhang, L. Zhang, and X. Mou. RFSIM: A Feature Based Image Quality Assessment Metric Using Riesz Transforms. In *International Conference on Image Processing (ICIP)*, pages 321–324, 2010.

[19] L. Zhang, L. Zhang, X. Mou, and D. Zhang. FSIM: A Fast Feature Similarity Index for Image Quality Assessment. *IEEE Transactions on Image Processing*, 20(8), 2011.

[20] J. A. Zwaka, Thomas P and Thomson. Homologous Recombination in Human Embryonic Stem Cells. *Nature biotechnology*, 21(3):319–321, 2003.