A Comparison of De-noising Techniques for FIRST Images *

Imola K. Fodor and Chandrika Kamath
Center for Applied Scientific Computing
Lawrence Livermore National Laboratory
Livermore, CA 94551
E-mail: fodor1@llnl.gov and kamath2@llnl.gov

Abstract

Data obtained through scientific observations are often contaminated by noise and artifacts from various sources. As a result, a first step in mining these data is to isolate the signal of interest by minimizing the effects of the contaminations. Once the data has been cleaned or de-noised, data mining can proceed as usual. In this paper, we describe our work in de-noising astronomical data from the Faint Images of the Radio Sky at Twenty-Centimeters (FIRST) survey. We are mining this survey to detect radio-emitting galaxies with a bent-double morphology. This task is made difficult by the noise in the images caused by the processing of the sensor data. We compare three different approaches to de-noising: thresholding of wavelet coefficients advocated in the statistics community, traditional filtering methods used in the image processing community, and a simple thresholding scheme proposed by FIRST astronomers. While each approach has its merits and pitfalls, we found that for our purpose, the simple thresholding scheme worked relatively well for the FIRST data set.

1 Introduction

In the last few decades, there has been an explosion in the amount of scientific data generated through observations, experiments, and simulations. To effectively and efficiently explore these large, multi-dimensional data sets, scientists are increasingly turning to data mining techniques.

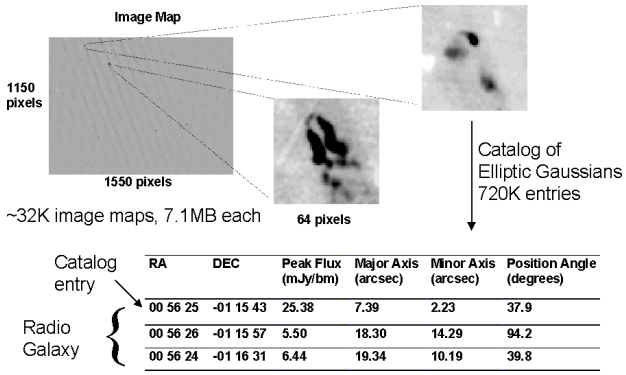
Sapphire [13] is a project in large-scale data mining at the Center for Applied Scientific Computing at Lawrence Livermore National Laboratory. We view data mining as an iterative and interactive process involving data pre-processing, search for patterns, knowledge evaluation, and possible refinement of the process based on input from domain experts or feedback from one of the steps. For scientific data, extensive pre-processing is often needed prior to the recognition of patterns. Depending on the domain and the problem, this may involve several steps. In the case of data obtained through observations or experiments, one of these steps may be the reduction of noise that typically contaminates the observed signal.

Removing noise from an observed signal is an active research area, with extensive work done by both the statistics and image processing communities [2, 11, 3, 14]. In addition to being visually unappealing, noise in the data can also have a detrimental effect as important features may become unrecognizable, resulting in poor quality of any subsequent processing. There are many sources of noise in images, such as dust on the optics, the effects of the camera's sensors and associated electronics, stray photons, electromagnetic interference, and post-processing of the data collected by the sensors.

Noise in images is frequently modeled as Gaussian or Poisson noise if the noise is due to the sensors, or salt and pepper noise if it is due to defective pixels or dust particles. The noise can be either additive or multiplicative and is often considered to be un-correlated to the signal. If the type of noise in an image is known, it becomes much easier to select an appropriate method to reduce the noise. However, as we describe in this paper, modeling the noise may be non-trivial in some cases.

As part of the Sapphire project, we are working with a variety of data sets, which are contaminated with different types of noise. To remove the noise from these data sets, we have implemented a comprehensive

*In the Proceedings of the Third Workshop on Mining Scientific Datasets, held in conjunction with the First SIAM International Conference on Data Mining, Chicago, April 2001



RA	DEC	Peak Flux (mJy/bm)	Major Axis (arcsec)	Minor Axis (arcsec)	Position Angle (degrees)
00 56 25	-01 15 43	25.38	7.39	2.23	37.9
00 56 26	-01 15 57	5.50	18.30	14.29	94.2
00 56 24	-01 16 31	6.44	19.34	10.19	39.8

Figure 1: An example of a FIRST image map with two bent-double galaxies.

set of de-noising techniques based on spatial filtering and wavelet thresholding techniques. In this paper, we compare and contrast the effectiveness of these methods on one of our test-bed applications, namely, the Faint Images of the Radio Sky at Twenty-Centimeters (FIRST) astronomical survey. We are analyzing this data set to help astronomers identify radio-emitting galaxies with a bent-double morphology. Until recently, FIRST astronomers identified such galaxies by visually inspecting the radio images. In addition to being subjective, this manual analysis is becoming infeasible as the survey is growing in size. Our goal is to replace this visual inspection by a semi-automated approach from data mining. In a previous publication [9], we reported on the success of our initial approach. In this paper, we describe the work we have done in de-noising the images to improve the accuracy of the features extracted.

This paper is organized as follows: Section 2 introduces the FIRST survey, describes the noise in the images, and illustrates why the presence of noise makes data mining difficult. In Section 3, we provide a brief description of the three de-noising techniques considered, namely, wavelet thresholding, spatial filtering, and simple thresholding. In Section 4 we discuss our experiences with applying these techniques to the FIRST data. Finally, Section 5 concludes with a summary of our findings and proposed future work.

2 The FIRST Survey

The FIRST survey [1] was started in 1993 with the goal of producing the radio equivalent of the Palomar Observatory Sky Survey. When complete, FIRST will cover over 10,000 square degrees of the sky, to a flux density limit of 1.0 mJy (milli-Jansky). Using NRAO's Very Large Array (VLA), FIRST observations through 1999 have covered about 8,000 square degrees, producing more than 32,000 two-million pixel images. At a threshold of 1mJy, there are approximately 90 radio-emitting galaxies, or radio sources, in a typical square degree. Radio sources exhibit a wide range of morphological types that provide clues to the source class, emission mechanism, and properties of the surrounding medium. Of particular interest are sources with a bent-double morphology as they are tracers of large clusters of galaxies. Identifying bent-doubles is a particularly challenging task, because of the wide variety of morphologies involved. In the early stages of the survey, FIRST scientists visually inspected sub-images of the survey to identify bent-double galaxies. We now use this set, along with several non-bent doubles, as a training set for data mining.

The data from FIRST, both raw and post-processed, are accessible via a user-friendly interface at the FIRST website [6]. There are two forms of data available for use — image maps and a catalog. In Fig. 1, we show an image map containing examples of two bent-doubles. Each map covers an area approximately 0.45 square degrees, with pixels that are 1.8 arc seconds wide. These large image maps, obtained from processing the raw data from the 27 telescopes of the VLA, are mostly background noise that appears as streaks in the images. These streaks, at various angles and thicknesses, can be seen in the image map in Fig. 1 and in its zoomed-in version in Fig. 4, panel (a). It is beyond the scope of this paper to discuss how the extensive, but careful, post-processing of the raw data results in the noise in the images; the interested reader is referred to [12, 1]. Suffice it to say that this background noise can interfere with

correctly identifying which pixels belong to a galaxy and which are part of the background. Since our identification of a galaxy as a bent-double is dependent on the shape of the galaxy, it is important to correctly identify the pixels that form the galaxy.

In addition to the image maps, FIRST also provides a source catalog [17], obtained by processing an image map by fitting two-dimensional elliptic Gaussians to each radio source. For example, the lower bent-double in Fig. 1 is approximated by more than seven Gaussians while the upper one is approximated by three Gaussians. Due to an upper limit to the number of Gaussians that are used to fit each radio source, highly complex sources are not approximated well using just the information in the catalog. Each entry in the catalog corresponds to the information on a single Gaussian, including the Right Ascension (RA) and Declination (Dec) for the center of the Gaussian, the major and minor axes, the peak flux, the position angle of the major axis, etc.

In our first iteration of the data mining process, we derived features from the catalog to characterize the galaxies, built decision trees based on the labeled training set provided by the astronomers, and then classified all the galaxies in the catalog using the decision trees. The method proved to be successful, as we even found a bent-double that was overlooked by the astronomers in a manual search [7, 9]. In the second iteration of data mining, we used the trees generated in the first iteration to enhance our relatively small training set. This resulted in a lower classification accuracy. In the third iteration of data mining, we are using several approaches to improve the accuracy, including deriving additional features directly from the images. This would help us address the concern that some features of the galaxies, which are apparent in the images, are lost while converting the image data into the catalog. However, before identifying the image pixels that form a galaxy, we need to first remove the noise from the images.

3 De-noising Techniques for Image Data

The problem of de-noising image data can be stated as follows: given the observed zero-mean data $Y_{i,j}$ which is a noisy version of the signal $X_{i,j}$:

$$Y_{i,j} = X_{i,j} + \epsilon_{i,j}, \quad i = 1, \cdots, I, \quad j = 1, \cdots, J, \tag{1}$$

find an "optimal" estimate of the signal $X_{i,j}$. Depending on the distribution of the noise, and the quantity optimized, this estimation problem can have different solutions. A common approach is to assume that $\{\epsilon_{i,j}\}$ is an independent and identically distributed (*iid*) Gaussian $N(0, \sigma^2)$, independent of the signal, and use the minimal mean square error (MSE) to evaluate the optimality of the estimate.

We considered three different approaches for de-noising the FIRST image data. The first approach, based on thresholding wavelet coefficients, uses the statistical properties of the wavelet decomposition of an image. The second approach, popular in the image processing community, uses spatial filters to smooth an image. The third method is used by FIRST astronomers to generate the catalog from the images. We next describe these techniques in more detail.

3.1 De-noising by Thresholding of Wavelet Coefficients

De-noising data using thresholding of wavelet coefficients is a subject that has been studied by several researchers in the statistics community. Unfortunately, we did not find any work that compared and contrasted the different approaches or provided practical guidance on the selection of the different options that are available in each method. To rectify this, we first conducted a study of various wavelet de-noising methods on known test images with simulated noise added to them [8]. By comparing the performance of the various de-noisers with respect to the original, known, noiseless images, we were able to narrow the options available. The optimal method found in this experiment was then applied to the FIRST data. In this section, we provide a brief summary of wavelet de-noisers and our experimental results with images containing additive simulated noise.

Wavelet de-noising is one particular way of obtaining the estimates $\hat{X}(i,j)$ in Equation (1). If $\mathbf{Y}$, $\mathbf{X}$, and ϵ denote the observed data, the noiseless data, and the error matrices in Equation (1), respectively, then the three main steps of the process are given as follows:

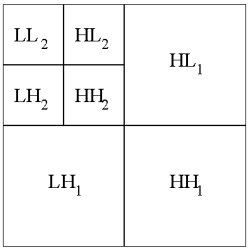


Figure 2: Wavelet decomposition subbands using a decimated transform with two multi-resolution levels.

1. Calculate the empirical wavelet coefficient matrix $\mathbf{w}$ by applying a wavelet transform $\mathbf{W}$ to the data:

$$\mathbf{w} = \mathbf{W}\ \mathbf{Y} = \mathbf{W}\ \mathbf{X} + \mathbf{W}\epsilon, \tag{2}$$

2. Threshold the detail coefficients of $\mathbf{w}$ to obtain the estimate $\hat{\mathbf{w}}$ of the wavelet coefficients of $\mathbf{X}$:

$$\mathbf{w} \rightarrow \hat{\mathbf{w}}, \tag{3}$$

3. Inverse transform the thresholded coefficients to obtain the de-noised estimate:

$$\hat{\mathbf{X}} = \mathbf{W^{-1}}\ \hat{\mathbf{w}}. \tag{4}$$

The number N of the coefficients $\mathbf{w}$ in Equation (2) can vary depending on the type of transform used. For our work [8], we used decimated transforms [10, 2], as they have $N = IJ$ coefficients, regardless of the number of multi-resolution levels K used in the decomposition. Fig. 2 displays the subbands of a two-level decimated decomposition. The N wavelet coefficients are grouped into subbands according to the $K = 2$ multi-resolution levels and four spatial directions. The directions reflect the order of the high-pass (H) and low-pass (L) filtering along the two dimensions of the original image. For example, the level one coefficients are decomposed into subbands LH_1 (vertical detail, first level), HL_1, (horizontal detail, first level), HH_1 (diagonal detail, first level), and LL_1 (smooth, first level). The smooth part is then similarly decomposed into the four subbands corresponding to the second multi-resolution level. The process could be further iterated for additional multi-resolution levels.

The differences in the wavelet-based de-noising methods lie in the details of the three de-noising steps in Equations (2)-(4). First, one chooses the wavelet filters, the number of multi-resolution levels, and the boundary treatment rule, and obtains the wavelet coefficients $\mathbf{w}$ in Equation (2). Next, if w denotes a single generic detail coefficient, its thresholded version $\hat{w}$ in Equation (3), is obtained through

$$\hat{w} = \hat{\sigma}\ \delta_\lambda(w/\hat{\sigma}), \tag{5}$$

where $\delta_\lambda()$ denotes the thresholding (or shrinkage) function, λ the threshold, and $\hat{\sigma}$ an estimate of the noise in Equation (1).

Note that we suppressed in the notation the possible dependence of the noise estimate, the threshold, and the thresholding function on the multi-resolution level or on the subband. We consider three ways in which the threshold values can be calculated and applied:

- global: one value for all coefficients across all subbands and levels
- level-dependent: one value each for the coefficients on a given level; the same value is used across subbands on a level
- subband-dependent: one value each for the coefficients on a given subband on a given level

The various de-noising methods in the literature differ in the details of the implementation of Equation (5), that is, in selecting the thresholding function which determines how the threshold is applied, estimating the noise, and selecting the shrinkage rule to determine the threshold λ. In [8], we compared the following choices:

(a) (b) (c) (d)

Figure 3: De-noising results with the Lena image with $\sigma = 20$, symmlet12 wavelet, three multi-resolution levels, periodic boundary treatment. (a) Original image. (b) Noisy image, $MSE = 399.50$. (c) Subband-dependent "SURE" rule with soft thresholding, $MSE = 61.59$. (d) Global universal rule with hard thresholding, $MSE = 103.95$.

- thresholding functions: hard, soft, garrote, and semisoft
- thresholding rules: "universal" [4], "minimizing the false discovery rate" [15], "top" [2], "hypothesis testing" [11], "SURE" [5], and "Bayes" [3]

We based our noise estimates on wavelet coefficients from different subbands, and combined them with estimators such as the sample standard deviation and the median absolute deviation (MAD).

Based on our experimental results using various test images and various levels of additive Gaussian simulated noise, we concluded that the subband-dependent "SURE" and the "Bayes" methods with soft thresholding were superior to all the other methods. Fig. 3 shows a typical result from the comparison study. Panel (a) shows the original image, (b) the original image contaminated by noise with $\sigma = 20$, (c) the result of the best de-noiser, and (d) the result of an inferior de-noiser. Using the mean square error (MSE) values with respect to the original noiseless image, we see that there is a large difference between the good and the bad de-noiser. The former reduces the MSE of 399.50 of the noisy image to 61.59, while the latter results in a noticeably higher value of 103.95. The results shown in Fig. 3 are characteristic to all images and noise levels we evaluated, and are robust across the wavelet used, the number of multi-resolution levels in the decomposition, and the boundary treatment rule. A main factor influencing the results is the method of noise estimation. We found that using a robust estimator, such as the MAD, with the diagonal wavelet coefficients from the first level decomposition provided accurate noise estimates.

3.2 De-noising by Spatial Filters

A common approach to de-noising images in the image processing community is the use of spatial filters. These approaches have been studied extensively [16, 14]. The specific filters we have implemented include:

1. Mean filters
2. Gaussian filters
3. Unsharp Masking filters. These filters subtract the mean filtered image from the original image.
4. Scaled Unsharp Masking filters. These filters calculate

$$(1.0 + \alpha)original\_image - (\alpha)mean\_filtered\_image \tag{6}$$

   where α is a real number. Our experiments indicated that $\alpha = -0.8$ gave relatively good results for our test images. However, this value may vary with the image.
5. Alpha-trimmed mean filters
6. Median filters

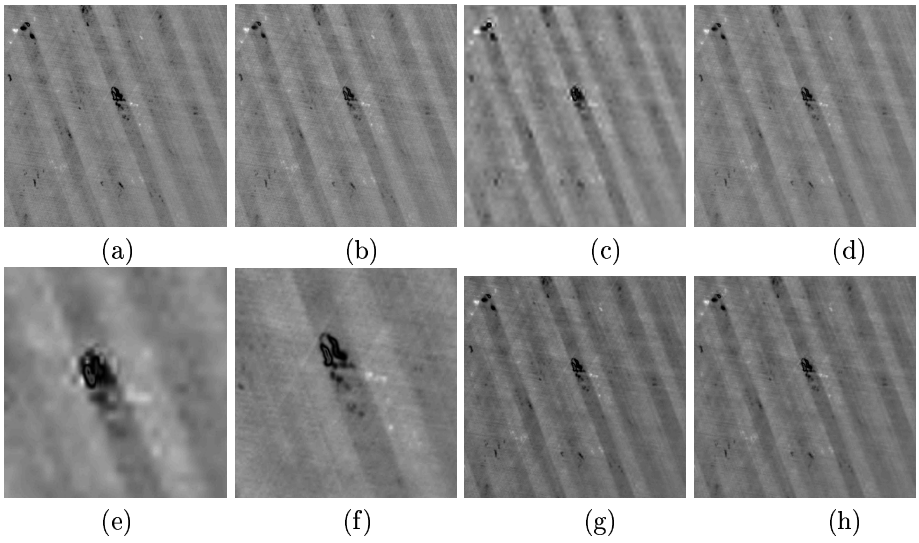


Figure 4: De-noising results with a typical FIRST image with symmlet12 wavelet, three multi-resolution levels, periodic boundary treatment. (a) FIRST image. (b) Subband-dependent "SURE" rule with soft thresholding. (c) Subband-dependent "hypothesis test" rule with soft thresholding. (d) Global "universal" rule with hard thresholding. (e) Zoomed-in version of (c). (f) Zoomed-in version of (d). (g) Level-dependent "top" rule with soft thresholding. (h) Level- and subband-dependent "top" rule with soft thresholding.

7. Mid-point filters. The value calculated is the average of the minimum and maximum within the filter mask.
8. Minimum mean squared-error filters
9. Scaled unsharp masking filter followed by a mean filter
10. Mean filter applied twice
11. Gaussian filter applied twice
12. Minimum mean square error filter followed by a mean filter
13. Minimum mean square error filter followed by a Gaussian filter

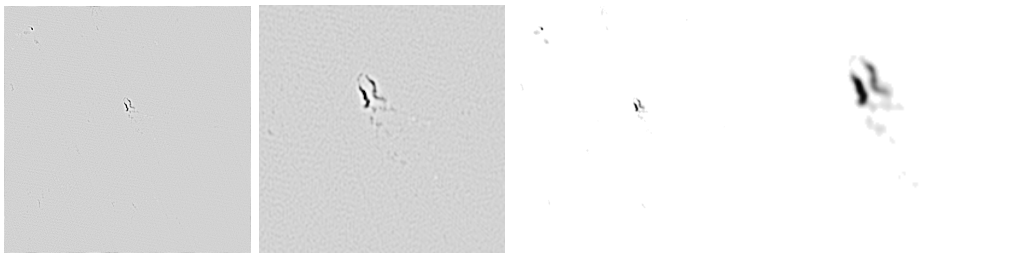
Our results with test images with additive simulated noise indicated that the simple spatial filters were very competitive with the wavelet-based techniques. For values of $\sigma = 10$, a simple Gaussian filter of size 3 worked well, while for higher σ values, the combined minimum mean squared error filter, followed by either the mean or the Gaussian filter was the best.

3.3 De-noising using Simple Thresholding

In this approach, we borrow an idea from FIRST astronomers and use a simple thresholding of the pixels for de-noising. For the FIRST data, an estimate of the root mean square (RMS) of the noise at each pixel is available as coverage maps for the area covered in the survey [6]. In order to generate the catalog from the images, FIRST astronomers start by dropping all pixels below 5*RMS in value, as they consider them to be noise.

4 Experimental Results for De-noising FIRST Images

In this section, we compare and contrast the different de-noising techniques using a sample image from the FIRST survey. Fig. 4 displays a 512×512 pixel FIRST image in panel (a). The pixels in the original image contain floating-point radio-intensity values in Janskys. This is in contrast with standard images which



(a) (b) (c) (d)

Figure 5: (a) (Original - 3×3 mean-filtered) difference image. (b) Zoomed-in version of (a). (c) Image with simple threshold, threshold $= 5 * RMS$, $RMS = 0.373$. (d) Zoomed-in version of (c).

have grayscale values in the $[0, 255]$ range. Panel (b) corresponds to the best de-noiser, as determined in Section 3.1. Panel (c) is the result of the subband-dependent "hypothesis testing" shrinkage rule with soft thresholding. This is an inferior result with significant blurring around the radio galaxies. Panel (d) shows the result of the "universal" approach with hard thresholding. This was an inferior method for our test images, but for the FIRST image, led to a result that was comparable with the best. The previously best "SURE" method still provides a reasonable output, as compared with the other wavelet de-noising methods. Panels (e) and (f) show the zoomed-in detail of panels (c) and (d), respectively. Unlike the test cases reported in Section 3.1, none of the wavelet de-noising methods we considered de-noised the FIRST images sufficiently. The streaks of the original image remain in the de-noised images and the results of the study reported in [8] do not carry over to the FIRST data set. We believe that this is because the *iid* additive Gaussian noise assumption in the wavelet de-noising methods of Section 3.1 is not satisfied by the FIRST observations. Further, as the noise is very structured, the wavelet-based denoiser treats it as features, preserving it in the denoised image.

It is possible to tailor the wavelet de-noising to the FIRST data for example, by aggressively de-noising in the diagonal direction to perhaps remove more of the "streaks". Fig. 4, panels (g) and (h) illustrate the effect of such a subband-dependent de-noising. The image in panel (g): kept 15% of all detail coefficients at level 1, 40% of all detail coefficients at level 2, and 80% of all detail coefficients at level 3. In panel (h), we kept only 1% of the detail coefficients in the HH subband on all three levels, while keeping the values for the other two subbands (HL and LH) unchanged. There is no noticeable visual difference between the two images. Killing more diagonal coefficients does not result in a superior image, as the streaks in the original image do not seem to correspond to more noise in the diagonal direction.

In comparison, the spatial filters performed better at removing the noise from the FIRST images. To reduce the effects of blurring, we used a small filter of size 3×3. Many of the spatial filters were not able to reduce the noise in the images sufficiently. However, the unsharp masking filter performed relatively well as shown in Fig. 5, panel (a). The galaxies are clearly visible as the signal in very low noise. Panel (b) is the zoomed in version, showing the details. A similar result was obtained by subtracting the Gaussian filtered image from the original image.

The simple thresholding approach used by FIRST astronomers also worked very well. Using a threshold of $5 * RMS$ with an $RMS = 0.373$ milliJanskys obtained for the center of the image from the FIRST coverage map, we obtained the image in panel (c) of Fig. 5. Panel (d) shows the zoomed-in version.

5 Conclusions and Future Work

For the FIRST images, we found that de-noising with wavelets did not lead to desired results. While these techniques are very effective in images with Gaussian speckle noise such as the noise occurring in Synthetic Aperture Radar (SAR) images, they did not work well for the FIRST data. In contrast, simpler techniques such as unsharp masking or thresholding based on domain knowledge worked well. However, wavelets can still be exploited in a data mining framework for tasks such as feature extraction, image registration, and image compression.

Our plans for future work in the detection of bent-double galaxies in the FIRST data set include

extracting additional features from the images after we have de-noised them using the simple thresholding approach. We expect that an enhanced feature set will allow us to improve the accuracy of classification. We will report on our experiences at the workshop.

6 Acknowledgement

UCRL-JC-142085 - This work was performed under the auspices of the U.S. Department of Energy by University of California Lawrence Livermore National Laboratory under contract No. W-7405-Eng-48.

References

[1] Becker, R., White, R., and Helfand, D. The FIRST survey: Faint images of the radio sky at twenty-cm. *Astrophysical Journal 450* (1995), 559.

[2] Bruce, A., and Gao, H. *S+WAVELETS User's Manual.* StatSci Division of MathSoft, Inc., 1994.

[3] Chang, G., Yu, B., and Vetterli, M. Adaptive wavelet thresholding for image denoising and compression. *IEEE Trans. Image Processing 9* (2000), 532–1546.

[4] Donoho, D. L., and Johnstone, I. M. Ideal spatial adaptation via wavelet shrinkage. *Biometrika 81* (1994), 425–455.

[5] Donoho, D. L., and Johnstone, I. M. Adapting to unknown smoothness via wavelet shrinkage. *Journal of the American Statistical Association 90*, 432 (December 1995), 1200–1224.

[6] Faint images of the radio sky at twenty-cm. http://sundog.stsci.edu/.

[7] Fodor, I., Cantu-Paz, E., Kamath, C., and Tang, N. Finding bent-double radio galaxies: A case study in data mining. In *Interface : Computer Science and Statistics, Volume 33, New Orleans, LA* (April 2000).

[8] Fodor, I., and Kamath, C. On denoising images using wavelet-based statistical techniques. *Manuscript in preparation* (2001).

[9] Kamath, C., Cantu-Paz, E., Fodor, I., and Tang, N. Searching for bent-double galaxies in the FIRST survey. In *The 2nd Workshop on Mining Scientific Datasets* (2001), Army High Performance Computing Research Center, http://www.ahpcrc.umn.edu/conferences/July2000/index.html.

[10] Mallat, S. A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence 11* (1989), 674–693.

[11] Ogden, R. *Essential Wavelets for Statistical Applications and Data Analysis.* Birkhäuser, 1997.

[12] Perley, R., Schwab, F., and Bridle, A. *Synthesis Imaging.* National Radio Astronomy Observatory, 1985.

[13] Sapphire: Large-scale data mining and pattern recognition. http://www.llnl.gov/casc/sapphire/.

[14] Umbaugh, S. *Computer Vision and Image Processing: A Practical Approach using CVIPtools.* Prentice Hall, 1998.

[15] Vidakovic, B. *Statistical Modeling by Wavelets.* Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., 1999.

[16] Weeks, A. *Fundamentals of Electronic Image Processing.* SPIE/IEEE Series on Imaging Science and Engineering. SPIE Optical Engineering Press and IEEE Press, 1996.

[17] White, R., Becker, R., Helfand, D., and Gregg, M. A catalog of 1.4 GHz radio sources from the FIRST survey. *Astrophysical Journal 475* (1997), 479.