

# AUTOMATED LINE FLATTENING OF ATOMIC FORCE MICROSCOPY IMAGES

S. A. Tsiftaris, J. Zujovic, and A. K. Katsaggelos

Department of Electrical Engineering and Computer Science  
Email: {stsift, jzu850, aggk}@eecs.northwestern.edu  
Northwestern University, 2145 Sheridan Rd., Evanston, IL 60208, USA

## ABSTRACT

In this paper, an automated algorithm to flatten lines from Atomic Force Microscopy (AFM) images is presented. Due to the mechanics of the AFM, there is a curvature distortion (bowing effect) present in the acquired images. At present, flattening such images requires human intervention to manually segment object data from the background, which is time consuming and highly inaccurate. The proposed method classifies the data into objects and background, and fits convex lines in an iterative fashion. Results on real images from DNA wrapped carbon nanotubes (DNA-CNTs) and synthetic experiments are presented, demonstrating the effectiveness of the proposed algorithm in increasing the resolution of the surface topography.

**Index Terms**— curve fitting, nanotechnology, polynomial approximation, object detection.

## 1. INTRODUCTION

AFM functions by bringing a cantilever tip in physical contact with (or close proximity to) the sample, revealing nanometer scale topographical information [1]. Fig. 1. shows a block diagram of an AFM. The repulsive force from the surface applied to the tip bends the cantilever. The amount of bending is measured and fed back to control the vertical movement of the sample in order to keep the contact force constant. The vertical movement follows the surface profile and is recorded as the surface topography. AFM is used to capture images of cells, materials, biomolecules etc. An example of a typical AFM image of DNA-CNTs is shown in Fig. 2.

Due to the mechanics of the AFM, there is a curvature distortion (bowing effect) present in the acquired images. This can be observed in Fig. 2, where intensities are low in the middle of the image while they are high at the sides. The tip follows arc-like lines in the image, creating a spherical or parabolic shape, depending on the scanner [2]. To compensate for this (known as line flattening or plane fitting), objects in an AFM image are manually labeled to generate an exclusion mask, which usually is rigid (either parallelogram or ellipse) and may not adequately represent the shape of the sought after object. The data of each row in the image outside the mask are fitted by a polynomial, which is subsequently subtracted from all the data values of the line.

With respect to the existing procedures, the processing of AFM images of small objects (such as DNA-CNTs) can be substantially improved for the following reasons:

- (i) it is labor intensive and its automation is highly desirable;
- (ii) the manual labeling of the objects in the image is highly inaccurate, since the objects are not easily distinguishable in the

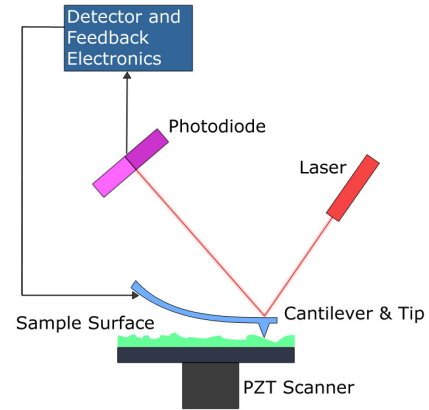


Fig. 1. AFM block diagram (Wikipedia Foundation).

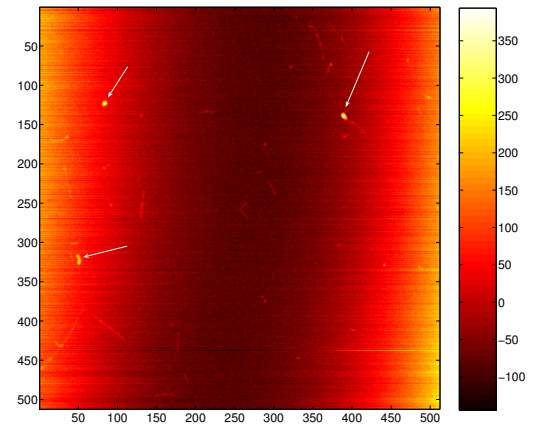
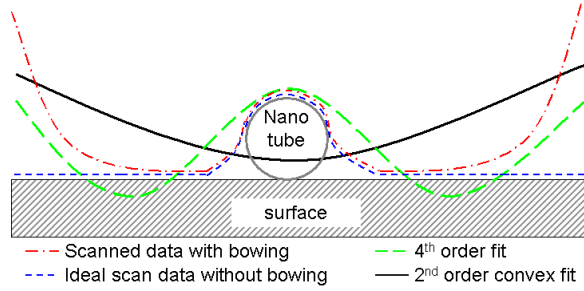


Fig. 2. AFM image of DNA-CNTs (indicated with 3 arrows).

- (iii) with the existing line flattening techniques, the fitted polynomials are non-convex, in disagreement with the intrinsic physics, thus reducing the accuracy of the recovered surface topography.

In this paper, a method to automatically and accurately fit convex polynomials on background only points of a line is presented. In tandem to select the background points the algorithm iterates between fitting and automated classification of background and object points on a line. It is superior in comparison to the techniques used so far because it speeds up the process, uses less human resources, and produces more accurate results.



**Fig. 3.** Bowing effect illustration.

This paper is organized as follows: Section 2 provides the problem formulation and discusses possible solutions; Section 3 discusses the proposed algorithm in detail. Section 4 offers the results and discussion while Section 5 concludes with summary comments and possible future extensions.

## 2. BACKGROUND AND PROBLEM FORMULATION

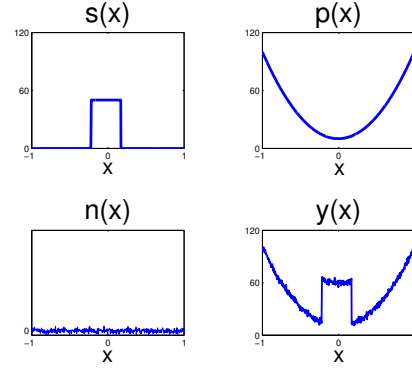
To demonstrate the significance of convexity and object point exclusion, an illustrative synthetic example of a nanotube imaged by an AFM is shown in Fig. 3. The dash-dot line represents the actual data recorded by the AFM, containing the bowing artifact, and the short-dashed line represents the recording if there were no curvature distortion. Due to tip convolution [1], there is blurring at the object boundaries. In an ideal scenario, the recorded height of the tube should be equal to the recorded width but due to the tip convolution problem the width is highly inaccurate. Therefore, the goal is to recover the height of the object while removing the bowing effect. The dashed line represents a non-convex polynomial least squares (LSQ) fitted to the recorded curve. Subtracting this polynomial from the recorded line would completely remove the object in the flattened image, and the information of its height would be lost. It will also create dark patches in the two minima. Enforcing convexity will ensure a better approximation (solid line) of the curvature distortion, and also will avoid possible exclusion of the object in the flattened image.

To the best of the authors' knowledge there is no direct literature on the topic of simultaneous and automated object exclusion and line flattening in AFM images. An apparent global 2D solution would be to fit parabolic surfaces on the whole image [3]. However, some artifacts in AFM images are intrinsic to the particular moment of operation, like vertical scanner drift, its internal non-linearities etc, and hence cannot be adequately modeled with a parabolic 2D surface. In [4] a Gaussian Mixture Model is employed to model a smooth, and distinguishable from the signal, background. In AFM, neither the background is smooth (the necessity of processing each line is stressed previously), nor it can be assumed that the background and foreground pixels are distinguishable only by their values (as can be seen in the original AFM image in Fig. 2).

There is however a relation to the MRI shading removal problem. However, in [5], [6] either the noise is ignored, or a Gaussian signal distribution is assumed, and hence they are not applicable here. The algorithm presented in [7], despite being formulated for images, can be applied to the 1-D case (to accommodate the AFM problem), but histogram-based entropy estimation is not robust in the case of large object coverage, and the optimization routine (Powell's method) does not enforce any convexity constraints.

Each observed scan line can be modeled as:

$$y_i(x) = s_i(x) + p_i(x) + n_i(x), i=1, \dots, N, \quad (1)$$



**Fig. 4.** Signal  $s(x)$ , polynomial  $p(x)$ , noise  $n(x)$  and  $y(x)$ .

where  $y_i(x)$ ,  $s_i(x)$ ,  $p_i(x)$ ,  $n_i(x)$  represent the raw data, signal, (convex) polynomial, and noise, respectively;  $x$  is the horizontal coordinate,  $i$  corresponds to the line number in the image, and  $N$  is the total number of lines. An example is shown in Fig. 4.

The overall objective is to

$$\text{estimate } s_i(x) \text{ given } y_i(x). \quad (2)$$

To find signal points in the line one could resort to image based object detection or segmentation. However, the presence of noise and the limited *a priori* knowledge of the object's shape and size render both methods not applicable as general solutions.

Since  $s_i(x)$  cannot be explicitly estimated, the problem can be best formulated as

$$\text{estimating } p_i(x) \text{ given } y_i(x), \quad (3)$$

where  $p_i(x)$  is a convex polynomial.

## 3. PROPOSED APPROACH

Each line in an AFM image is flattened iteratively in a two step process: (i) K-means based classification into object and background points, followed by (ii) convex polynomial fitting on the background points. In more detail the steps are:

1. For  $x$  in *Background*, fit a convex polynomial  $p_i^{(k)}(x)$  ( $k$  is the iteration number) and subtract it from the data
$$z_i^{(k)}(x) = y_i^{(k)}(x) - p_i^{(k)}(x). \quad (4)$$

Initially  $y_i^{(1)}(x) \equiv y_i(x)$ , and all points are considered as background.

2. Assuming a known noise variance  $\sigma_i^2$ , decide if the line has objects based on:

$$\text{var}(z_i^{(k)}(x)) > \alpha \cdot \sigma_i^2, \quad (5)$$

where  $\alpha$  is a user given parameter. If Eq. 5 is satisfied continue, otherwise go to step 6.

3. For all  $x$ , using the K-means algorithm, cluster the signal
$$c_i^{(k)}(x) = y_i(x) - p_i^{(k)}(x), \quad (6)$$
into *Object* and *Background* classes.

4. Increase the iteration number  $k$ , and for  $x$  in *Background*

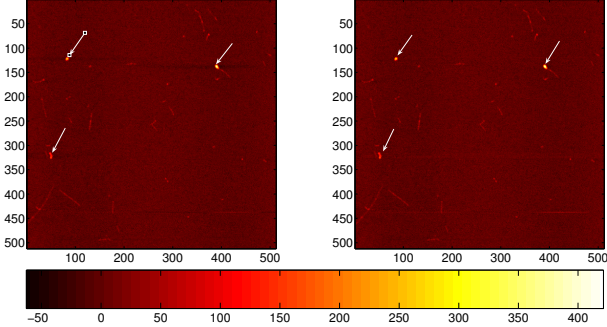
$$y_i^{(k)}(x) = y_i(x). \quad (7)$$

5. Repeat steps 1 to 4.

6. No object data further detected, output the polynomial from the last iteration
$$p_i^*(x) = p_i^{(k)}(x), \quad (8)$$

and the  $x$  belonging to the final *Object* class.

In the following steps 1 and 3 are detailed, and noise variance estimation methods for step 2 are discussed.



**Fig. 5.** Image in Fig. 2 (left) flattened with 3<sup>rd</sup> degree polynomials without any object detection; (right) flattened with 3<sup>rd</sup> degree polynomials using the proposed algorithm.

### 3.1. Fitting Convex Polynomials

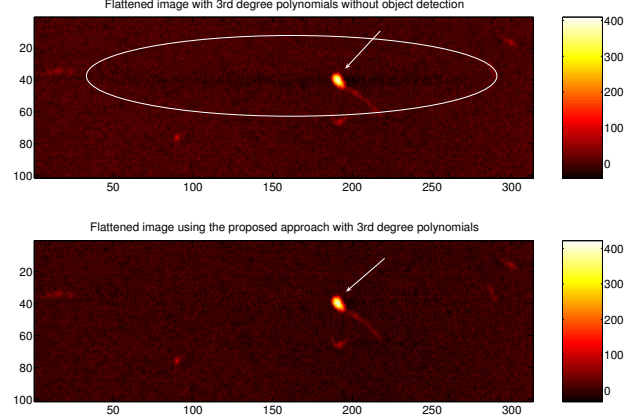
For fitting convex polynomials the following methods were tested: (a) *Constrained optimization* with Sequential Quadratic Programming (SQP) [8], or sum-of-square (SoS) polynomials [9]; however, SQP might not converge if the initial guess is not close to the solution, and SoS is computationally very intensive. (b) *Fitting ellipses* [10] which are easy and fast to implement; however, they are restricted to only second degree curves and hence may not fit adequately the boundary points. (c) *Direct least square fitting of convex polynomials* as in [11]; this method, however, did not yield accurate results. The SQP method was finally chosen for its accuracy and computational efficiency.

### 3.2. K-means clustering

The variance test in step 2 is chosen since it can detect even small objects. The larger the object, the worse the first fit, the greater the variance, thus detection confidence grows with the object size, which is expected. The flattened line in Eq. 6 is clustered in *Background* and *Object* classes. This is an 1D rather than a 2D clustering procedure, given that the points are clustered only by their  $c_i^{(k)}(x)$  value, not their  $x$  position. The background points should have lower values than the object points, and therefore the cluster with the smallest centroid is marked as “background”, while the one with the largest is labeled as “objects”. Since K-means clustering is sensitive to initialization, the min and max of  $c_i^{(k)}(x)$  are chosen to be the initial centroids. The points marked as objects are excluded from the signal and a new polynomial is fitted.

### 3.3. Noise estimation

Estimation of the noise variance is of critical importance. In the current incarnation of the algorithm, the noise variance is estimated on a per line basis. After the first fit, the polynomial is subtracted from the original data, and local variances are found using a sliding window. The location of the peak in the local variances histogram is used as the estimated noise variance. This method may not perform well if there is large object coverage in a line. In the opposite case a global approach is utilized, where each line in an image is first fitted with a polynomial, then the variance of each line is measured, and the line with the minimum variance is used with the above procedure to find an estimate of the noise variance. Alternatively, an approach similar to [12] can be used; however in this case difference operators are used to identify edges in images which are very susceptible to the noise in AFM images. As a final option, the user can supply a variance estimate.



**Fig. 6.** (top), (bottom) detail of left and right image of Fig. 5 respectively, to highlight the ability of the proposed algorithm to segment the data and eliminate negative contrast regions (ellipse).

## 4. RESULTS AND DISCUSSION

To obtain the results the algorithm of section 3 was implemented in MATLAB (The Mathworks, Inc). The optimization toolbox and the function *fmincon* (which utilizes an SQP solver) were used to fit convex polynomials of degree two to five. Initially a polynomial was LSQ fitted to the data. If this polynomial was not convex the *fmincon* routine was used to find a convex polynomial with initial starting coefficients of the LSQ fitted polynomial. The latter is used to ensure a good starting point, and improve the convergence of the SQP algorithm. In the subsequent sections, first results on flattening Fig. 2 are presented, followed by tests on synthetic raw line data to highlight the robustness of the algorithm.

### 4.1. Flattening of AFM images of DNA-CNTs

The image of Fig. 2 was used as a test image. 3<sup>rd</sup> degree polynomials were fitted to each line without any object exclusion (automatic or manual) and subtracted from each line, as shown in Fig. 5 left, and with the proposed algorithm, shown in Fig. 5 right. The height of the object at pixel location (139,390) (rightmost arrow) was 398 in the raw image while in the flattened images, it was 413.8 in Fig.5 left and 425 in Fig. 5 right, respectively. (Please notice in Figs. 2 and 5 that the minimum values are negative.). For clarity, Fig. 6 zooms in around this object. In addition to the accurate height recovery, we see that the dark patches around the object in the top image, Fig. 6, which are an artifact of bad fitting and introduce a non-existent dent-like deviation of the background, have disappeared on the bottom image, Fig. 6. This is a clear indication that the fitting is correct, since Fig. 3 showcases how bad fitting creates dark patches. Further experiments with more DNA-CNT images exhibited similar performance.

### 4.2. Flattening of synthetic AFM line data

Raw AFM lines were synthesized following a similar construct as in Fig. 4. Results obtained by the proposed algorithm on a synthetic  $y(x)$  are shown in Fig. 7. In this figure it can be seen that: (i)  $p^*(x)$  of 2<sup>nd</sup> degree (dashed curve) provides a very accurate estimate of the original  $p(x)$  (solid); (ii) the convexity constraint is necessary, since the 5<sup>th</sup> degree unconstrained curve (dash-dot) completely obscures the object; and (iii) the K-means based classification (+ markers) adequately identifies the object.

For more exhaustive experiments, various lines were generated with  $p(x)=ax^2+bx+c$ , with  $(a, b, c)$  taking values

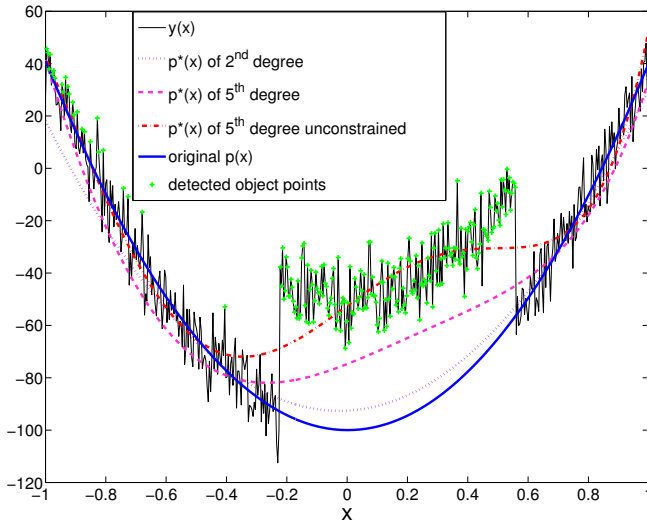


Fig. 7. An example line fitted with various polynomials.

from the set  $\{(140.5, -0.5, -100), (60.5, -0.5, -50), (80.5, -70.5, -50)\}$ , and a pulse  $s(x)$  of height  $A$   $\{20, 75, 200, 400\}$ , width  $w$   $\{10, 20, 100, 300\}$  and delay  $L$   $\{50, 100, 200\}$ . Gaussian random noise of zero mean and variance  $\sigma^2$   $\{4, 16, 100, 400\}$  was also added. For a given noise variance, while the other parameters remained the same, the algorithm was run 20 times by fitting 2<sup>nd</sup> and 5<sup>th</sup> degree convex and 5<sup>th</sup> degree unconstrained polynomials. At each trial the height of the recovered signal defined as  $s^*(x) = y(x) - p^*(x)$  for  $x$  in *Object*, was estimated by its average  $A^* = E\{s^*(x)\}$ . The final estimate  $A^*$  was the average of all 20 trials  $A^* = E\{A^*\}$ .

In Fig. 8 the ratio of estimated height and pulse height ( $A^*/A$ ) is shown as function of the ratio of pulse height and noise standard deviation ( $A/\sigma$ ) fitted with the three polynomial types. It is evident that enforcing convexity is important since the unconstrained polynomials fail to recover the signal and hence register large errors. On the other hand, despite the fact that the lines were generated with a binomial, fitting 5<sup>th</sup> degree convex polynomials recovers the signal in low signal to noise ratios ( $A/\sigma < 10$ ), while 2<sup>nd</sup> degree polynomials recover the signal more accurately in larger signal to noise ratios. It should be noted that in 3% of all trials for which  $\sigma^2$  was large,  $A$  was small and  $w$  was large, we failed to recover the signal with accuracy greater than 70%.

## 5. CONCLUSION

In this paper a novel algorithm for the automatic line flattening of AFM images was presented. The algorithm offers distinct advantages when compared to previous solutions, since it accommodates convex polynomials thus increasing the accuracy of the fit, and can automatically detect the presence of objects and segment them using a K-means algorithm. Results on real as well as synthetic data were presented that showed the robustness of the proposed method and its ability to resolve the relevant features of the objects. Line flattening is an integral part of AFM image analysis and precedes any other operation such as tip deconvolution, feature extraction, etc. Tests with more images will be performed in the future, as well as, comparisons to other methods, such as the one of [7].

## 5. ACKNOWLEDGMENTS

The authors would like to thank Prof. Hersam and Dr. Ostojic from Materials Science Dept. at Northwestern for the images.

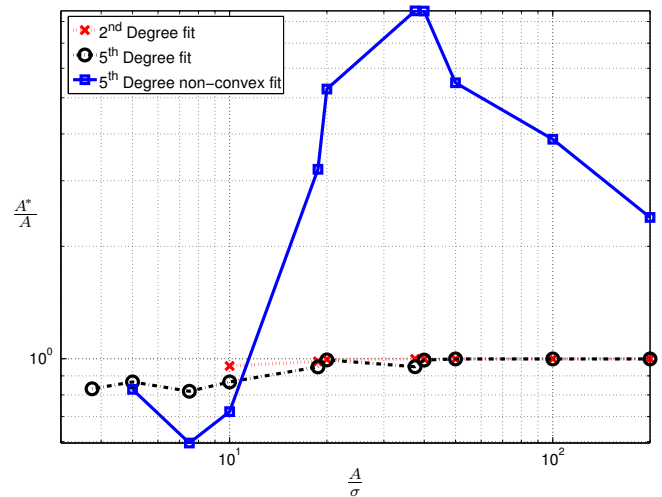


Fig. 8. The ratio  $A^*/A$ , for various polynomials, as a function of  $A/\sigma$ , fitted with 2<sup>nd</sup> and 5<sup>th</sup> degree convex, and 5<sup>th</sup> degree unconstrained polynomials.

## 6. REFERENCES

- [1] D. Sarid, *Scanning Force Microscopy*, New York: Oxford University Press, 1991.
- [2] D. Ricci, P.C. Braga, "Recognizing and avoiding artifacts in AFM imaging," *Methods in Molecular Biology*, vol. 242, pp. 25-37, 2004.
- [3] M. Dai, T.S. Newman, C. Cao, "Least-squares-based fitting of paraboloids," *Patt. Rec.*, vol. 40, no. 2, pp. 504-515, Feb 2007.
- [4] A. Saranlı, "A Gaussian-mixture based approach to spatial image background modeling and compensation," in *Proc. of EUSIPCO*, 2007, pp. 1457-1461.
- [5] J. Sled, A. Zijdenbos, A. Evans, "A Nonparametric Method for Automatic Correction of Intensity Nonuniformity in MRI Data," *IEEE Trans. on Medical Imaging*, vol. 17, no. 1, pp. 87-97, Feb 1998.
- [6] A. Noe, S. Kovacic, J. Gee, "Segmentation of Cerebral MRI Scans Using a Partial Volume Model, Shading Correction and an Anatomical Prior," in *Proc. of SPIE*, 2001, pp. 1466-1477.
- [7] Z. F. Knops, J. B. A. Maintz, M. A. Viergever, J. P. W. Pluim, "Normalized Mutual Information Based Registration Using K-means Clustering and Shading Correction," *Medical Image Analysis*, vol. 10, issue 3, pp. 432-439, June 2006.
- [8] R. Fletcher, *Practical Methods of Optimization*, Chichester: John Wiley and Sons, 1987.
- [9] A. Magnani, S. Lall, S. Boyd, "Tractable fitting with convex polynomials via sum-of-squares," in *Proc. of the 44th IEEE CDC-ECC*, 2005, pp. 1672-1677.
- [10] A. Fitzgibbon, M. Pilu, R. M. Fisher, "Direct least square fitting of ellipses," *IEEE Tran. on Pattern Analysis and Machine Intelligence*, vol. 21, no. 5, pp. 476-480, May 1999.
- [11] D. J. Hudson, "Least-squares fitting of a polynomial constrained to be either non-negative, non-decreasing, or convex," *J Roy. Stat. Soc. B*, vol. 31, no. 1, pp. 113-118, 1969.
- [12] K. Rank, M. Lendl, R. Unbehauen, "Estimation of image noise variance," in *IEEE Proceedings - Vision, Image, and Signal Processing*, vol. 146, no. 2, April 1999.