AUTOMATIC RECOGNITION OF DNA NANOSTRUCTURES IN ATOMIC FORCE MICROSCOPY (AFM) IMAGE: FIRST EXPERIENCE ON DNA PLIERS

Yuexing Han
TITECH, Japan
hanyuexing@gmail.com

Akito Hara
TITECH, Japan
hara.a.ac@m.titech.ac.jp

Akinori Kuzuya
Kansai-Univ, Japan; JST
kuzuya@kansai-u.ac.jp

Ryosuke Watanabe
Kansai-Univ, Japan
K947410@kansai-u.ac.jp

Yuichi Ohya
Kansai-Univ, Japan
yohya@kansai-u.ac.jp

Akihiko Konagaya,
TITECH, Japan
kona@dis.titech.ac.jp

ABSTRACT

Recently, bottom-up fabrication techniques are expected to alleviate the limitations of top-down fabrication. Bottom-up fabrication requires self-assembling facilities to construct complex structures including DNA nanostructures and molecular robots. In this paper, we focus on the automatic recognition of flexible DNA origami named "DNA pliers" on AFM image. Recognition of DNA pliers is challenging since DNA pliers can have several forms: parallel, cross and anti-parallel forms, depending on hinge angles. Our experience suggests that the combination of the curvature scale space method and convexity-concavity detection works well for DNA nanostructure recognition if appropriate contour information of DNA nanostructures is available from an AFM image.

Keyword: Atomic Force Microscope (AFM) image analysis, curvature scale space technique, convexity-concavity detection, DNA nanostructures, Molecular robotics

1. INTRODUCTION

Recently, bottom-up fabrication techniques are expected to alleviate the limitations of top-down fabrication, such as photolithography and etching, which have long driven advances in nanotechnology but are now approaching their physical limits. Bottom-up fabrication requires self-assembling facilities to construct complex structures including DNA nanostructures [1] and molecular robots [9].

DNA is one of such ideal building blocks that construct complex nanostructures because a strand of DNA can specifically bind with its complementary counter-strand by Watson-Crick base paring to form a double helix with high affinity. In addition, each double helix consisting of various sequences of DNA has the same diameter and

pitch. High affinity of DNA double helixes enables the concept of "programmed self-assembly" of designed DNA sequences. It should be noted that DNA sequences are useful not only for holding information but also for constructing complex nanostructures. In this sense, DNA sequences are considered as the key devices in molecular robotics.

"DNA origami", a pioneer of DNA nanostructures, was first reported by Rothemund using atomic force microscopy (AFM) images of nanoscale geometric shapes such as smiley faces and world map hemispheres [8]. The key to this technique is the use of single-stranded DNA as the scaffold for the self-assembling of DNA sequence with a length of 7249 nucleotides. The following extensive study of DNA origami enables to construct not only arbitrary 2D nanostructures but also 3D nanostructures with hollow polyhedron [2, 10, 6].

In DNA origami, AFM plays an essential role to observe the shapes of DNA nanostructures by scanning the surface of the target objects on mica with a very tiny but precise cantilever. It enables to obtain a high-resolution image in the order of fractions of a nanometer, roughly 1000 times better than the optical diffraction limit. However, AFM images are difficult to analyze due to the high occurrences of image noise and contaminations as well as resolution limits.

In this paper, we focus on the automatic recognition of flexible DNA origami named "DNA pliers" on AFM image [2, 10]. A pair of DNA pliers has a hinge to connect two DNA nanostructures and branches for holding a small molecule in closed form, as shown in Figure 1. There are two closed forms, parallel form (Figure 1 (a)) and anti-parallel form (Figure 1 (c)) due to the anisotropic shape of the DNA pliers. Cross form (Figure 1 (b)) is an intermediate shape between the parallel form and the anti-parallel form. Since DNA pliers may take various forms on the same AFM image, automatic recognition of DNA pliers are strongly required to reduce human labor and human errors. In order to recognize DNA nanostructures and their types, many researchers have to mark them in the AFM image by hand. The method of artificial mark is difficult when the number of AFM images is large, and it is easy to make marking mistakes for a DNA by hand. This research, i.e., automatic recognition of DNA pliers, is urgently required to reduce human labor and eliminate human errors.

In order to automatically recognize the various forms of DNA pliers on the AFM image, we adopted an image processing pipeline consisting of outer contour information preprocessing, curvature information extraction and convexity-concavity detection. The preprocessing step extracts contour information of DNA pliers

eliminating image noises, contaminations and DNA pliers aggregations which are difficult to analyze in further steps. The curvature information extraction process focusses on the peak points that are useful for the classification of DNA pliers' forms with curvature scale space (CSS) method [3, 4, 5]. The convexity-concavity detection process gives additional information to distinguish DNA pliers' forms with regards to convexity-concavity values of landmarks obtained by controlling landmarks number (CLN) method [12].

The structure of this paper is as follows. Section 2 describes the preprocessing techniques to obtain outer contours of DNA pliers on AFM image. The design of the curvature information extraction process and the convexity-concavity detection processes are described in section 3 and 4, respectively. Furthermore, an algorithm flow is shown in section 5. Section 6 demonstrates the effectiveness of our approach with some AFM images. Finally, a brief discussion and conclusion are given in section 7.

(a) (b)

(c)
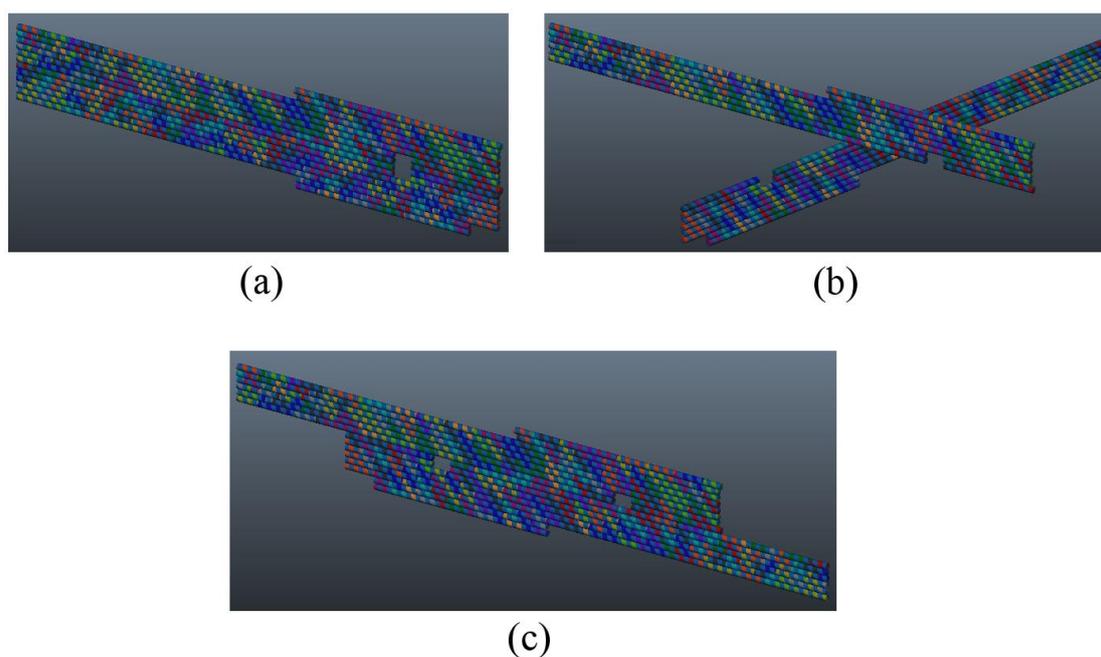
Figure 1. Typical DNA pliers' forms designed by **cadnano** and **MAYA**: (a) parallel DNA pliers, (b) cross DNA pliers, (c) anti-parallel DNA pliers.

2. Preprocessing and Outer Contour Detection

An AFM image contains not only DNA nanostructures, but also various noises, which are one of obstacles to extract DNA nanostructure shapes. These noises include small points and stains, as shown in Figure 2 (a). In order to detect outer contour information of DNA pliers, we applied some preprocessing techniques including noise elimination in AFM images followed by the distinction of single DNA pliers and

aggregated DNA pliers with regards to lengths and areas of target objects.

In order to eliminate noises on AFM images, conventional noise filtrating techniques are used as in Figure 2. First, an AFM image is changed from color to gray, then, an average value of the image pixels is used as a threshold to eliminate small signal noises. The lower threshold and the higher threshold of DNA pliers' area are estimated by the partial average value of non-black points, and some constant value (95 and 180 as the lower and higher threshold in our study, respectively), which was empirically determined in our previous studies. Dilating and eroding techniques are used for the elimination of noise areas mainly caused by contaminates. Finally, the gray image is changed to a binary image with an appropriate threshold for outer contour detection.

It should be noted that the obtained binary images contain not only single DNA pliers in various forms but also the aggregation of DNA pliers. Since our objective is the classification of single DNA pliers, the aggregation areas must be removed. For this purpose, the current implementation uses the lengths and areas of single DNA pliers' forms as the criteria of selection. A sophisticated detection algorithm is now under development.

a   b   c
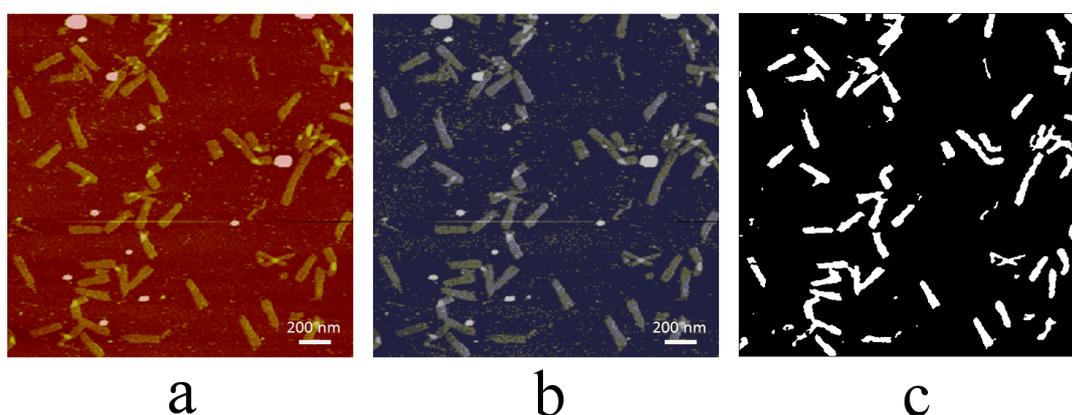
Figure 2. Preprocessing of AFM images: (a) original AFM image, with colored DNA pliers in yellow and noises in pink, (b) gray image, (c) final binary image

3. Curvature Information Extraction

The curvature scale space (CSS) method is a technique to find points whose curvatures are zero-crossings on a curve at varying levels of detail. The technique can reveal multi-scale, curvature-based shape information of planar curves known as CSS representation. The CSS representation is widely known for recognizing arbitrary shapes with noisy curves in any arbitrary scales or orientations [3, 4, 5].

In order to find rules to classify DNA pliers with regards to CSS representation, we adopted four model DNA pliers in Figure 3. The model-based approach enables us to pick up a characteristic CSS representation corresponding to typical DNA pliers' forms. Actual DNA pliers may have different CSS representations, for example, by means of twisted structures of DNA pliers. However, the main characteristics necessary for DNA pliers recognition are mostly preserved in CSS representations of DNA pliers in our study, mainly due to the large shape differences of the model structures. Further research is necessary for the other DNA structures.
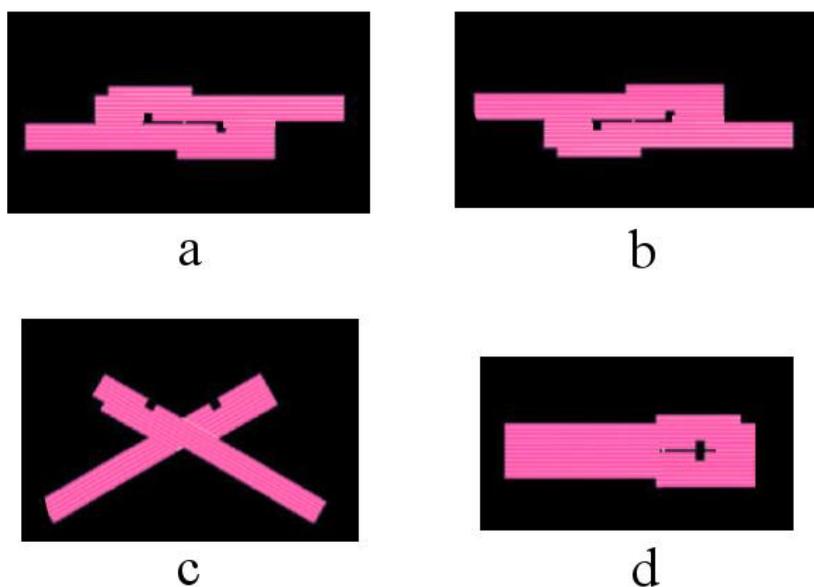
Figure 3. Four ideal DNA pliers shape models: (a) and (b) anti-parallel DNA pliers, (c) cross DNA pliers, and (d) parallel DNA pliers.

Figure 4 shows the ideal CSS representations corresponding to the anti-parallel DNA pliers, cross DNA pliers and parallel DNA pliers. Only outer contours are used in our study since inner contours in DNA pliers are too small to detect in real AFM images. Note that the outer contours of anti-parallel forms are symmetric in rotation; peak positions are slightly different on CSS representations due to the rotation dependency. The CSS representations are characterized by four large peaks for anti-parallel DNA pliers and cross DNA pliers, and by two large peaks for parallel DNA pliers, respectively. Other peaks are ignored in our further analysis since they are too small and difficult to be distinguished from noises.

Figure 5 shows the four examples of CSS representations of DNA pliers extracted from the AFM image in Figure 2. The CSS representations become very complicated due to the noises in the AFM image when compared with the ideal CSS representation in Figure 4. However, most DNA pliers still hold characteristic peak positions in their

CSS representation. Therefore, extraction of the large peaks is important for the classification of parallel DNA pliers and others. Sometimes, more than five large peaks can be found due to the sharp-big noise in the AFM image. In order to eliminate pseudo peaks, the distances between large peaks are used. This distance is between two points on the outer contour corresponding with the two peak points to judge whether the peak points are valid or not. If the distance is too short, it means that only one peak point is valid and the other ought to be removed. The number of large peak points (NPP) plays an important role in our classification of DNA pliers, although further study must be made for other DNA structures.
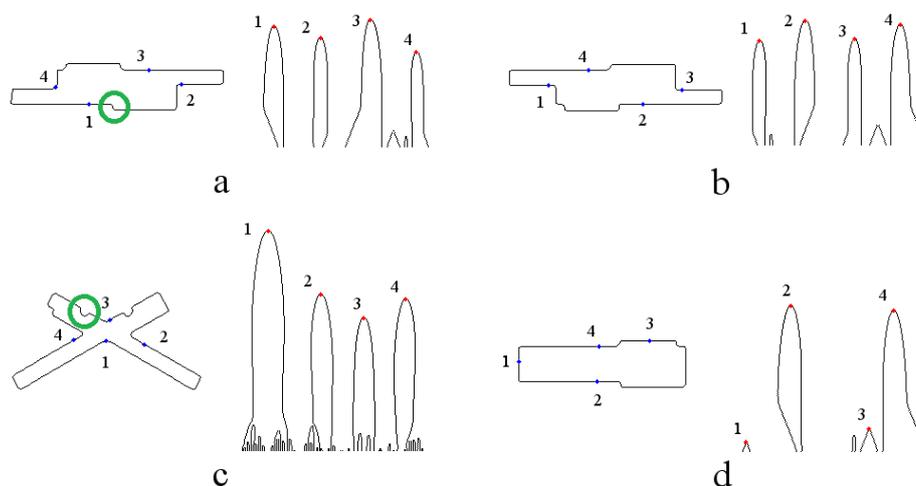
Figure 4. CSS representations of outer contours: Red points indicate peaks on CSS representation. Blue points represent corresponding points on the outer contour with indexes.
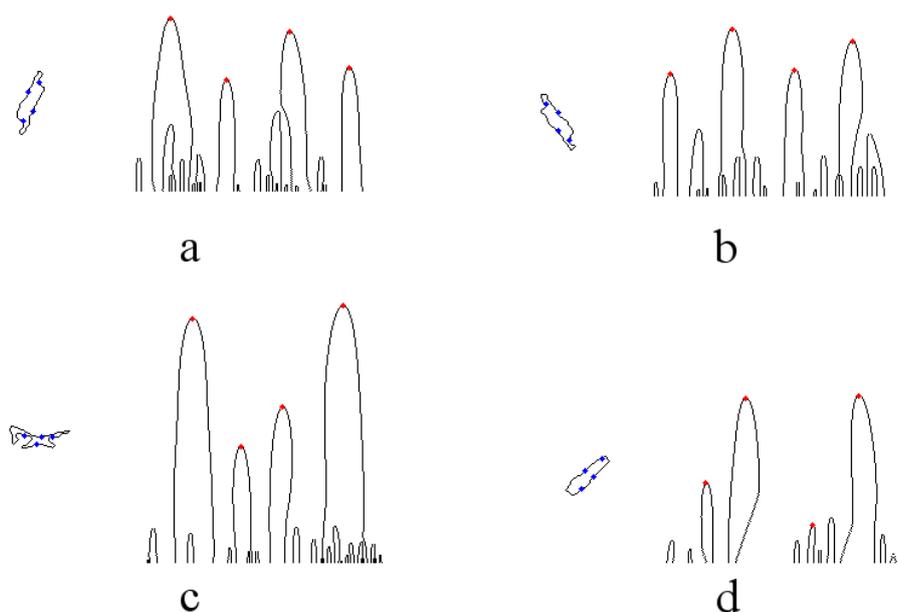
Figure 5. Four examples of outer contours and corresponding CSS representations of DNA pliers found on AFM image: blue points on the outer contours are corresponding to peak points on CSS representation.

4. Convexity-concavity Detection

CSS representation is useful but not sufficient to classify DNA pliers. For instance, anti-parallel DNA pliers and crossing DNA pliers have a similar CSS representation with four large peaks. Therefore, further information must be extracted to distinguish them. Here, we used convexity-concavity values of the outer contours as a measure.

In general, various attributes including colors, contours and textures can be used for object recognition. Among them, contours are widely believed to be the most important [10, 7, 11, 12, 13]. In case of DNA pliers recognition in the AFM image, it is difficult to obtain inner contours due to the resolution limit. Thus, we use outer contours to recognize DNA pliers.

Similar to other approaches, we acquire the outer contour of the target shape as a chain of linked pixels of which directions are ordered. In order to extract landmarks from the linked pixels, we use the controlling landmarks number (CLN) method [12]. The CLN method approximates an outer contour with many straight lines jointed at landmarks. Thus, a DNA pliers can be represented by a polygon whose number of vertexes is controlled. The number of vertexes is not larger than 13 for the DNA pliers. The polygon vertexes are defined in anti-clockwise order. Every polygon vertex has two angles, which are inner angle and outer angle, respectively. Here, outer angles of all vertexes are used to calculate convexity-concavity value.

Suppose a_i represents the outer angle of a vertex i and D_{ai} is convexity-concavity value corresponding to a_i. If a_i is larger than π, D_{ai} is 1; conversely, if a_i is not smaller than π, D_{ai} is 0; a_i does not equal to π, since vertexes are obtained by the CLN algorithm. The convexity-concavity value N_d, is defined by the following equation where N_v is the number of vertexes.

$$N_d = \sum_{i=0}^{N_v-1} (\mid D_{a_{i+1}} - D_{a_i} \mid) + \mid D_{a_{N_v}} - D_{a_0} \mid \qquad (1)$$

Figure 6 shows the ideal polygons obtained from model DNA pliers. The number of vertexes on outer contour is 9, 9, 13 and 5, corresponding to two anti-parallel DNA pliers, cross DNA pliers and parallel DNA pliers, respectively. The corresponding

convexity-concavity N_d values are 4, 4, 8 and 0. This suggests that N_d values represent useful information to distinguish anti-parallel DNA pliers and cross DNA pliers.

On actual AFM image, N_d may be altered by the effects of noises and twisted shapes of DNA structures. Therefore, we set a threshold 5 to distinguish the types of cross DNA pliers and anti-parallel DNA pliers; and a threshold 4 to judge whether a form is a parallel DNA or not in our study.
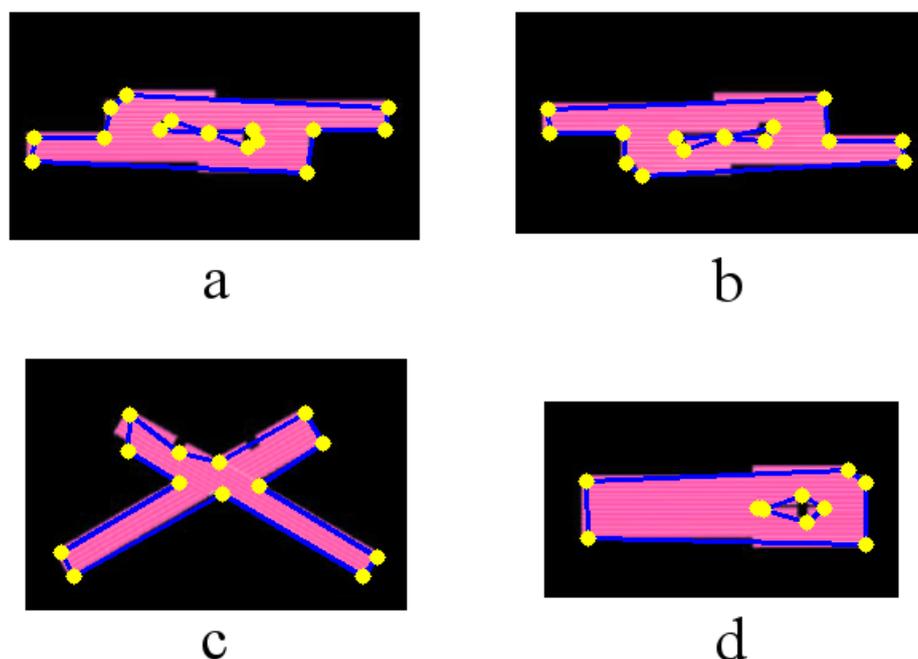
Figure 6. Polygons and their vertexes representing contours of DNA pliers with the CLN method

5. Classification Algorithm

The classification algorithm of DNA pliers is summarized in Figure 7. After obtaining outer contours of DNA nanostructures in image preprocessing, the DNA nanostructures whose longest length (L) and area (A) are not within the range of [L_{min}, L_{max}] and [A_{min}, A_{max}], respectively, are eliminated as non DNA pliers. L_{min}, L_{max}, A_{min}, *and* A_{max} are threshold values determined in empirical. Then, convexity-concavity values (N_d) and the number of large peaks (NPP) of DNA pliers are calculated with CNL method and CSS method, respectively. If NPP = 2 and $N_d < 4$, then the DNA nanostructure is classified into parallel form. In case of NPP = 4, there are two possible forms. The object is classified into cross form if $N_d > 5$, otherwise anti-parallel form. If NPP is not 2 or 4, then the DNA nanostructures are considered as non DNA pliers in this study.
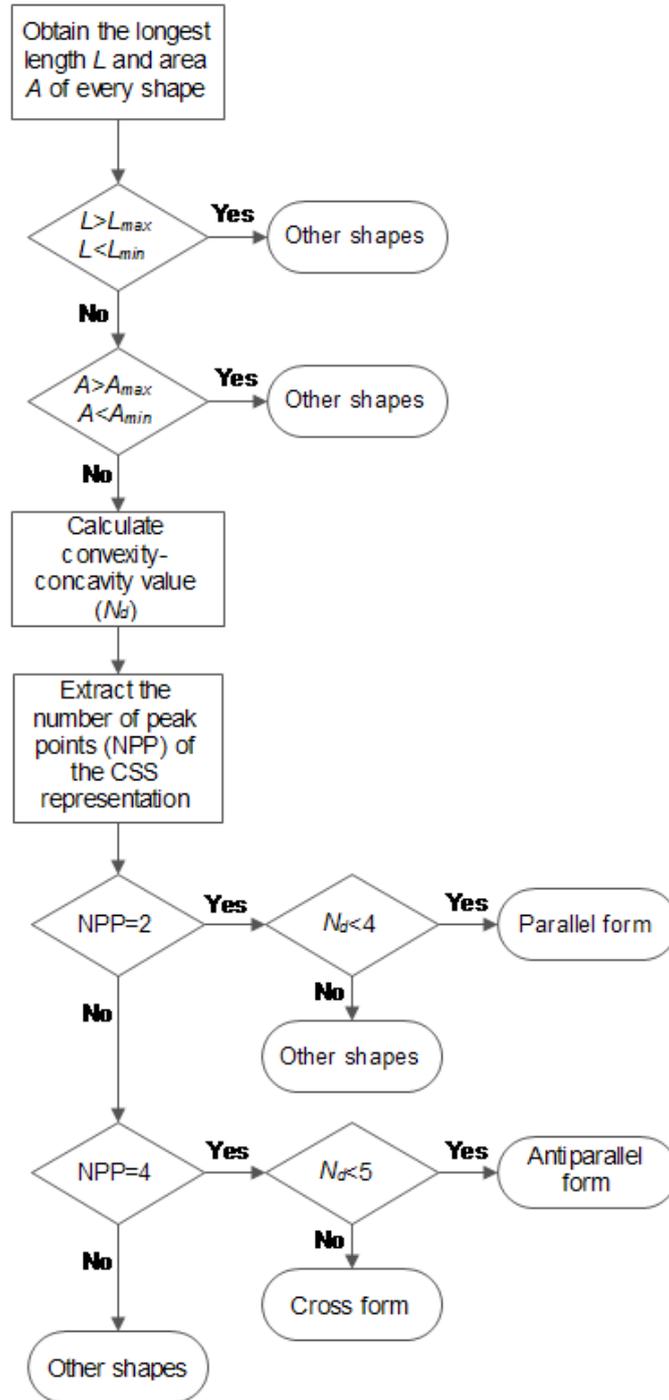
Figure 7. Classification algorithm of DNA pliers: L_{min} and L_{max}: the minimum and maximum lengths of the DNA pliers in any form with some margin, A_{min} and A_{max}: the minimum and maximum area of the DNA pliers in any form with some margin, NPP: the number of large peaks of CSS representation, N_d: convexity-concavity value.

6. Experimental Results

We used three AFM images of DNA pliers to show the effectiveness of our method. The AFM images include different forms of DNA pliers as well as a lot of noise points and areas. In Figure 8, the first row represents three original AFM images. The images are changed to binary images. Noise points and areas as well as most overlapped DNA pliers shapes were removed by the preprocessing techniques described in section 2.

Then, the polygon vertexes on the outer contour are extracted with the CLN method (shown in the second row of Figure 8). The yellow points and the blue lines in the figure represent vertexes and polygons characterized by the convexity-concavity values.
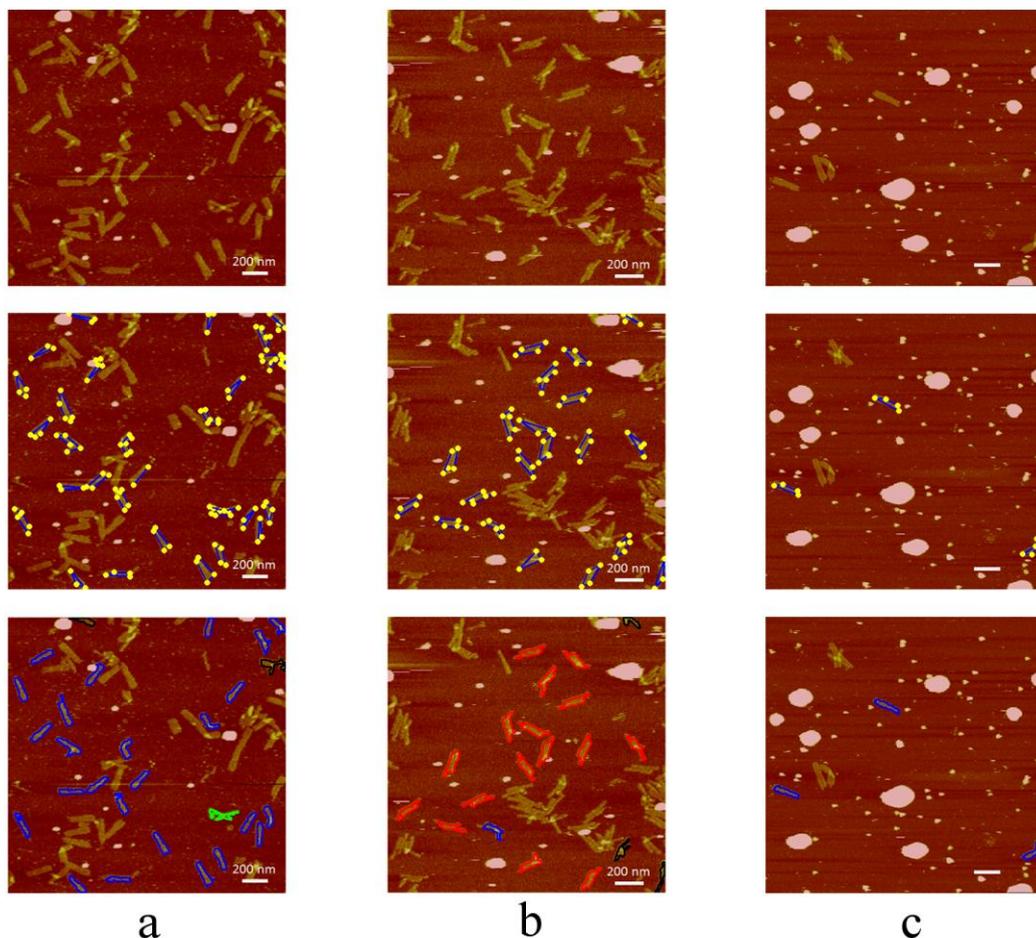
Figure 8. Three examples (a) (b) (c) of automatic DNA pliers recognition: original images (first row), landmarks obtained with the CLN method (second row), recognized DNA pliers (third row); blue (parallel form), red (anti-parallel form) and green (cross form)

The recognition results of DNA pliers are shown in the last row in Figure 8. Shapes in blue, red and green are parallel form, anti-parallel form and cross form, respectively. Ones in black are objects recognized as not DNA pliers. The execution time for DNA pliers recognition is summarized in Table 1. The execution time basically depends on the number of DNA pliers in AFM images. Average performance is 0.122 second per DNA pliers on a computer with Intel Core i5, 2.5GHz and RAMS 8GB.

It should be noted that not all single DNA pliers can be rightly recognized by the proposed method in our study. For example, an anti-parallel DNA pliers in (b) of Figure 7 is wrongly recognized as parallel one, denoted in blue color. Automatic recognition of DNA structures on AFM image is still in its infancy. Further studies are needed to improve the prediction performance as well as the recognition of other DNA nanostructures.

Table 1. The execution time of recognizing DNA pliers in AFM image

Image (see Figure 8)	a	b	c
Execution time (in second)	3.303	2.447	0.493

7. Conclusions and Future Works

A new approach to recognize DNA pliers and to classify their forms in AFM images is proposed. Our approach is characterized by its use of the CSS method and the convexity-concavity method to distinguish anti-parallel form, cross form and parallel form of DNA pliers. The experimental result with AFM images has demonstrated the effectiveness of our approach for the classification of DNA pliers' forms as well as the detection of DNA pliers from other objects including noises and aggregation of DNA nanostructures.

Automatic recognition of DNA nanostructures is still in its infancy. Although both the CSS method and the convexity-concavity method are very general and can be applied for the recognition of DNA nanostructures, further research is needed to improve the recognition performance and classification performance.

8. ACKNOWLEDGEMENTS

9. REFERENCES

[1] A. Kuzuya, Y. Ohya. DNA nanostructures as scaffolds for metal nanoparticles, *Polymer Journal, Vol.44, pp.452-460, 2-12.*

[2] A. Kuzuya, Y. Sakai, T. Yamazaki, Y. Xu, and M. Komiyama. Nanomechanical DNA origami 'single-molecule beacons' directly imaged by atomic force microscopy. *Nature Communications, Vol. 2, DOI: 10.1038/ncomms1452, 2011.*

[3] F. Mokhtarian and A. Mackworth. Scale-Based Description and Recognition of Planar Curves and Two-Dimensional Shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.PAMI-8, No. 1, Pp. 34-43, 1986.*

[4] F. Mokhtarian and A. Mackworth. A Theory of Multiscale, Curvature-Based Shape Representation for Planar Curves. *IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 14, No. 8, Pp. 789-805, 1992.*

[5] F. Mokhtarian and M. Bober. Curvature Scale Space Representation : Theory, Applications, and Mpeg-7 Standardization (Computational Imaging and Vision). Kluwer Academic Pub, United States, 2003.07.

[6] K. Lund, A. Manzo, N. Dabby, N. Michelotti, A. Johnson-Buck, J. Nangreave, S. Taylor, R. Pei, M. Stojanovic, N. Walter, E. Winfree, and H. Yan. Molecular robots guided by prescriptive landscapes. *Nature, Vol. 465, Pp. 206–210, 2010.*

[7] H. Bay, A. Ess, T. Tuytelaars and L. V. Gool. Speeded-Up Robust Features (SURF). *Computer Vision and Image Understanding (CVIU), Vol. 110, No. 3, Pp. 346-359, 2008.*

[8] P. W. K. Rothemund. Folding DNA to create nanoscale shapes and patterns. *Nature 440, Pp. 297-302, 2006.*

[9] S. Murata, A. Konagaya, S. Kobayashi, H. Saito, M. Hagiya. Molecular Robotics: A New Paradigm for Artifacts, *New Generation Computing, Vol.31, pp.27-45, 2013.*

[10] T. Yamazaki, Y. Aiba, K. Yasuda, Y. Sakai, Y. Yamanaka, A. Kuzuya, Y. Ohya, and M. Komiyama. Clear-cut observation of PNA invasion using nanomechanical DNA origami devices. *Chemical Communications, Vol. 48, Pp. 11361-11363, 2012.*

[11] Y. Han, B. Wang, M. Idesawa and H. Shimai. Recognition of Multiple Configurations of Objects with Limited Data. *Pattern Recognition, Vol. 43, Pp. 1467-1475, 2010.*

[12] Y. Han, H. Koike, and M. Idesawa. Recognizing Objects with Multiple Configurations. *Pattern Analysis and Applications, DOI: 10.1007/s10044-012-0277-7, 2012.*

[13] Y. Han. Recognize Objects with Three Kinds of Information in Landmarks. *Pattern Recognition, Vol. 46, Issue 11, Pp. 2860-2873, 2013.*