31st Annual International Conference of the IEEE EMBS
Minneapolis, Minnesota, USA, September 2-6, 2009

Toward Automatic Behavioral Screen: A Computational Model for Analyzing *Caenorhabditis elegans* Locomotion

Baitao Zhou and Joonghwan Baek

Abstract— **Genetic screen has been facilitating molecular geneticists to analyze mutants that produce certain phenotypes. However traditional methods for behavioral phenotype screen of mutant *Caenorhabditis elegans* rely on human observers and therefore are subjective and imprecise. This work dedicates a model to quantify and analyze the worm behavior using automatically-tracking and time-coded images. We have delved into the following questions: (1) how to achieve simplified worm-shape representation, (2) how to describe worm locomotion, (3) how to obtain frequent locomotion patterns and then representative behavioral patterns, and (4) how to discover interesting behaviorial actions within a representative behavioral pattern. Since the methodologies focus on rigorous image-based behavioral screening and phenotyping, the proposed methods should be trustworthy for behavior analysis of tiny organisms based on their microscopic video frames.**

I. INTRODUCTION

Since it is introduced by Sydney Brenner [1] in 1976, *C.(Caenorhabditis) elegans*, as a model organism, has been widely employed for solving fundamental issues in neuroscience [2][3]. *C. elegans*'s simple nervous system controls a rich variety of significant behaviors such as uncoordinated locomotion, drug chemosensory response, and mechanosensory response, which provide important channels to understand the relationship between genes and behaviors. However, distinctive abnormal behaviors among different species, particularly in some more complex behaviors, such as locomotion, are often subtle and complex, therefore making long-time naked-eye observation imprecise and subjective.

In the last a half decade, schools of researchers have been dedicating their works to computerized tool design for image-based automatic *C. elegans* analysis. One group [4] proved that a subset of discrete features abstracted from a comprehensive feature set (94 morphological and locomotion features) is sufficient to discriminate 6 (1 wild type, 5 mutants) types of animals. Later the same group [5], enlarged the feature set (253 features) and demonstrated that patterns of phenotypic similarity identified by natural clustering closely paralleled the functional similarities of the mutant gene products. Further refinements [6] of the system even allowed distinct behaviors, such as foraging, to be included in the parameter set. The typical output of these systems is a classifier supported by a subset of features and their calibrated measurements. However, those discrete features, like 'Avg head brightness', 'Maximal tail thickness', and 'Local head and tail movement ratio', are more suitable to be biological keys for taxonomy than direct behavioral phenotype for neurobiology. Another group [7] proposed an automated system for measuring nematode sinusoidal movements with such parameters as body bend extent, amplitude and wavelength. Behavioral phenotypes are demonstrated by comparing of those less perceptive parameters, which need further explanations and interpretations. Our work targets on designing an automatic behavioral screen model by comprehensively depicting and analyzing worm shape and locomotion over a prolonged time period using intuitive homogeneous parameters. These parameters can be used either in scoring a population of worms for a particular phenotype, might be done in large scale chemical or genetic screens, or for quantifying phenotypes or even describing phenotypes hitherto thought of as being distinctive from the wild type. To achieve the above target, we carried out the following designs:

- A worm Auto-tracking and image processing subsystem, which captured and processed image sequences into simplified worm-shape form, as shown in Fig. 1 (A), (B), and (C). The details can be seen in our previous work [4][8].
- A worm shape model and a locomotion model, as shown in Fig. 1 (D) and Fig. 2, which define a concise worm-shape representation and a hierarchical worm locomotion model. It will be discussed in section 2.
- Locomotion pattern mining and behavioral phenotype screening, which provide methodologies of quantitative worm locomotion mining and behaviorial analysis. It will be discussed in section 3.

II. MODELING WORM SHAPE MORPHOLOGY AND LOCOMOTION DYNAMICS

A. An (Orientation, Scale, and Translation)-impervious Shape Descriptor

Although it is a simplified form, a worm Skeleton Point Sequence (SPS) is still not efficient and convenient for shape representation and comparison due to animal's orientation arbitrariness and length variability. In addition, it is computationally burdensome to directly mine skeletons, each of

---

B.T. Zhou is with the School of Electronics, Telecommunication and Computer Engineering, Korea Aerospace University, 412791, Goyang, South Korea `zhoubaitao@gmail.com`

J.H. Baek is with Faculty of the School of Electronics, Telecommunication and Computer Engineering, Korea Aerospace University, 412791 Goyang, South Korea `jhbaek@kau.ac.kr`

978-1-4244-3296-7/09/$25.00 ©2009 IEEE

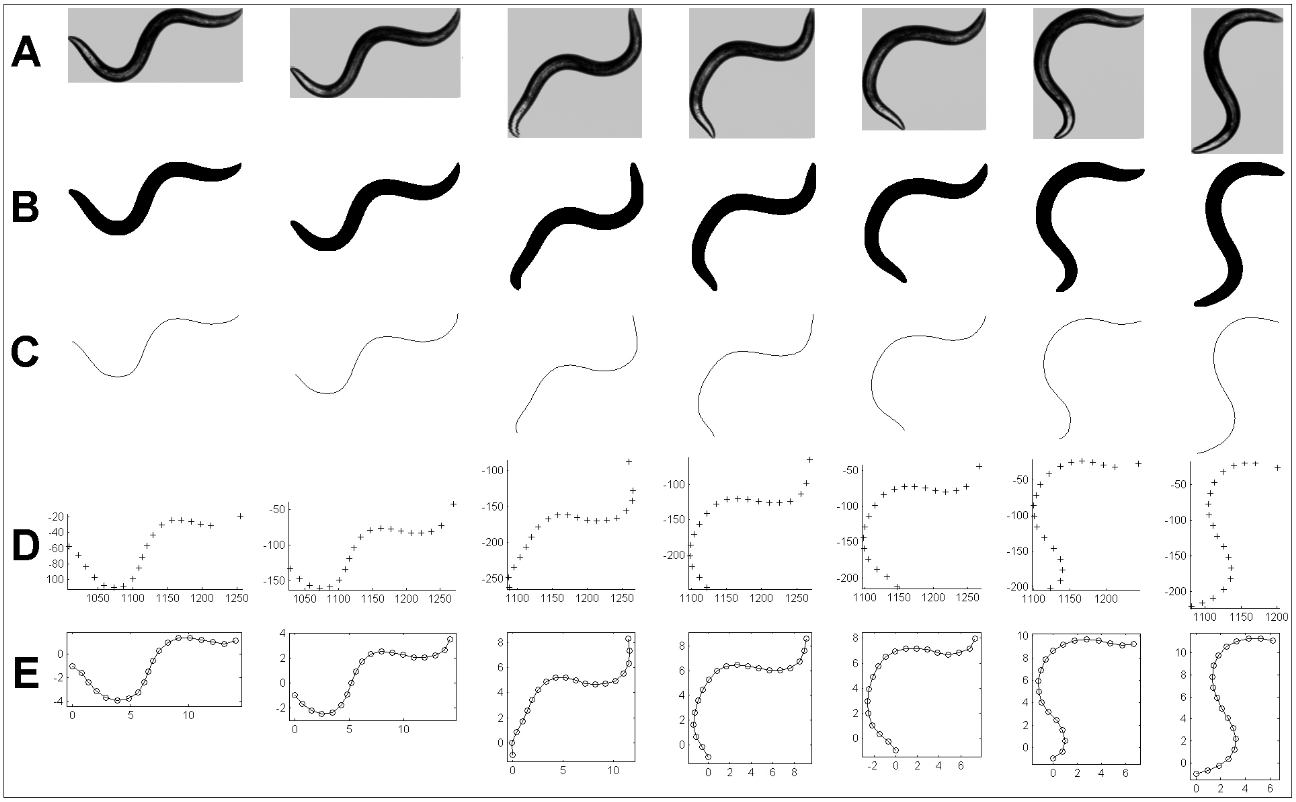

Fig. 1. The image frame sequence shows a wild-type animal is turning left. **A**. Image sequence samples captured from worm auto-tracking and imaging subsystem. The worm image was trimmed to the smallest axis-aligned rectangle, and saved as eight-bit grayscale data with frequency of 2Hz. The coordinates of the center of mass of the worm in the tracker field are also saved. **B**. Binary images after processing graylevel image with binarization, closing operation, and isolated-object removing. **C**. Skeleton images after thinning and pruning operation. **D**. Worm-shape depicted by even-sampled points from a skeleton in absolute reference frame. **E**. The regenerated worm shapes from their BASes, which shows BASe is feasible for describing worm shapes.

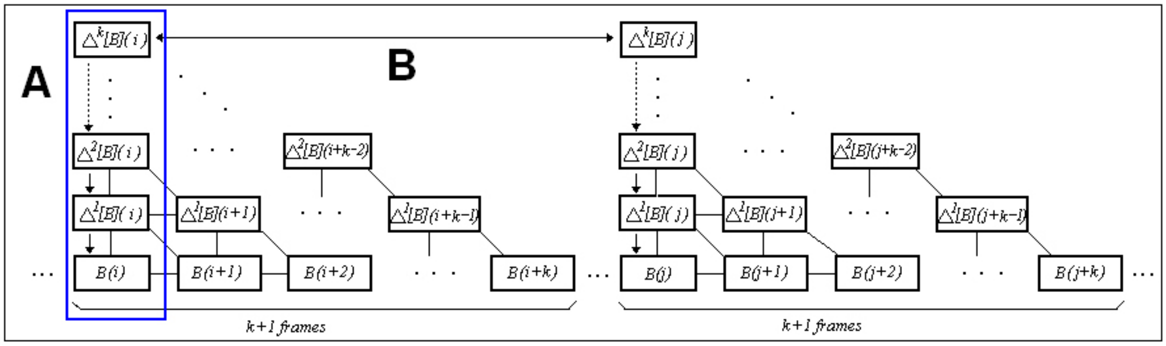

Fig. 2. Hierarchical locomotion descriptor and locomotion comparison. **A**. In the rectangle, a $(k+1)$-frame locomotion is represented by a hierarchical structure. **B**. Locomotion comparison between two $(k+1)$-frame locomotions is carried out of their counterparts in the direction of arrow.

which contains around 130 pixels in our experiment.

$$SPS_i = \begin{bmatrix} x_{i1}, y_{i1} \\ x_{i2}, y_{i2} \\ \cdots \\ x_{im}, y_{im} \end{bmatrix} \qquad ESP_i = \begin{bmatrix} \hat{x}_{i1}, \hat{y}_{i1} \\ \hat{x}_{i2}, \hat{y}_{i2} \\ \cdots \\ \hat{x}_{in}, \hat{y}_{in} \end{bmatrix} \tag{1}$$

$$L_i = \begin{bmatrix} l_{i1} \\ l_{i2} \\ \cdots \\ l_{i(n-1)} \end{bmatrix} = \begin{bmatrix} \hat{x}_{i2} - \hat{x}_{i1}, & \hat{y}_{i2} - \hat{y}_{i1} \\ \hat{x}_{i3} - \hat{x}_{i2}, & \hat{y}_{i3} - \hat{y}_{i2} \\ \cdots & \\ \hat{x}_{in} - \hat{x}_{i(n-1)}, & \hat{y}_{in} - \hat{y}_{i(n-1)} \end{bmatrix} \tag{2}$$

$$B_i = [\alpha_{i1}, \alpha_{i2}, ..., \alpha_{i(n-2)}] \text{ , where} \tag{3}$$

$$\alpha_{ij} = \lambda_{ij} \arccos \frac{l_{ij} \cdot l_{i(j+1)}}{\|l_{ij}\| \|l_{i(j+1)}\|} \text{ and}$$

$$\lambda_{ij} = \begin{cases} -1 & l_{ij} \times l_{i(j+1)} \geq 0 \\ 1 & l_{ij} \times l_{i(j+1)} < 0 \end{cases}$$

Here we design Bend Angle Series expression (BASe): an (Orientation, Scale, and Translation)-impervious shape descriptor. As shown in (1), (2), (3) and Fig. 1 (A)-(E), an animal's posture/shape of a frame i is represented by a $(n-2)$ bend-angle series as $B_i = [\alpha_{i1}, \alpha_{i2}, ..., \alpha_{i(n-2)}]$, where α_{ij}

is a deflection angle formed by lines $L_i(j)$ and $L_i(j+1)$. $L_i(j)$ is connected by points $ESP_i(j)$ and $ESP_i(j+1)$, which are evenly sampled from head to tail of the skeleton point sequence SPS_i. More generally, SPS_i can be obtained from a worm contour by traversing it from head or tail end in clockwise (or anticlockwise) manner. It is noted that the head and tail recognition scheme can be seen in our previous work [8].

B. Hierarchical Locomotion Descriptor and Comparison Scheme

Animal locomotion is a superimposition of worm-mass displacement and worm-shape variances, they can be studied by a decoupled way. Since the worm-shape variances manifest the main part of worm behaviors, it is the target of following discussion, and the term of locomotion refers to shape variance in the following parts.

1) Hierarchical Locomotion Descriptor: we propose a locomotion descriptor based on n-order finite differences, which represents continuous shape variances or locomotion in a hierarchical structure.

$$\begin{aligned} LD_i^k &= f(B_i, B_{i+1}, ..., B_{i+k}) \qquad (4) \\ &= [\triangle^k B_i, \triangle^{k-1} B_i, ..., \triangle^1 B_i, \triangle^0 B_i], \text{ where} \\ \triangle^k B_i &= \sum_{i=0}^{k} \binom{r}{k} (-1)^{k-r} B_{i+r} = \triangle^{k-1} B_{i+1} - \triangle^{k-1} B_i \end{aligned}$$

A $(k+1)$-frame locomotion starting from frame B_i can be exactly represented by a $(k+1)$-element vector, as it is expressed in (4), and shown in the Fig. 2. The locomotion can also be represented in the following forms:

$$\begin{array}{llllll} LD_i^0 & :[\Delta^0 B_{i+k}, & \Delta^0 B_{i+k-1}, & \cdots, & \Delta^0 B_{i+1}, & \Delta^0 B_i] \\ LD_i^1 & :[\Delta^1 B_{i+k-1}, & \Delta^1 B_{i+k-2}, & \cdots, & \Delta^1 B_i, & \Delta^0 B_i] \\ \cdots & & & & & \\ LD_i^k & :[\Delta^k B_i, & \Delta^{k-1} B_i, & \cdots, & \Delta^1 B_i, & \Delta^0 B_i] \end{array}$$

As shown in Fig. 2 blue rectangle, LD_i^k has the favorite property of describing locomotion from the view of global to local: $\Delta^k B_i$ determines the global variance of all the $k+1$ frames, $\Delta^{k-1} B_i$ gives changes of first k frames, . . ., $\Delta^1 B_i$ decides the differences the first two fames, and $\Delta^0 B_i$ is the original starting fame. But the other forms do not have this property.

2) Hierarchical Locomotion Comparison: The distance between two $(k+1)$-frame locomotions starting from B_i and B_j respectively can be compared through a hierarchical computing structure by calculating each similarity of the corresponding order of differences from global to local if necessary, as in (5).

If two locomotions are identical, then they must be similar from global variance to local details. Locomotion comparison procedure should start from the highest k-order differences comparison, if they are same, then the comparison goes to lower $(k-1)$-order of comparison; otherwise they do not have the same trend from that level, which means they are different locomotions. These kinds of comparison continues until jumping out from a lower-order differences comparison for their dissimilarity, or passing the final-step comparison of 0-order differences for their similarity. The latter declares the similarity of the two locomotions.

$$\begin{aligned} & Dist(LD_i^k, LD_j^k) \\ = \;& f(B_i, B_{i+1}, ..., B_{i+k}) - f(B_j, B_{j+1}, ..., B_{j+k}) \\ = \;& \begin{bmatrix} \triangle^k B_i & - & \triangle^k B_j \\ \triangle^{k-1} B_i & - & \triangle^{k-1} B_j \\ \cdots & & \\ \triangle^0 B_i & - & \triangle^0 B_j \end{bmatrix} \qquad (5) \end{aligned}$$

III. MINING FREQUENT LOCOMOTION PATTERN AND DISCOVERING INTERESTING BEHAVIORAL ASSOCIATION

A Frequent Locomotion (FLm) pattern is described as a segment of time-indexed frames (a video clip) that observed from a image sequence (a video unit), and the segment (clip) may also be found in the same or other image sequences (video units) with a certain frequency. A FLm pattern is defined as following:

Definition 1: (Neighbor Locomotion) Given a locomotion LD_i^k, its neighbor locomotion within a radius of r_ε is defined by $N_{r_\varepsilon}(LD_i^k) = \{LD_j^k : Dist(LD_i^k, LD_j^k) \leq r_\varepsilon\}$.

Definition 2: (Frequent Locomotion) A locomotion LD_i^k is a Frequent locomotion if it meets the following condition: $\mid N_{r_\varepsilon}(LD_i^k) \mid \geq Minpt$.

A. Mining Frequent Locomotion Patterns

This task aims at breaking the locomotion dataset of all assay species into locomotion patterns, each of which is a FLm pattern, and the members of a pattern may come from one or multiple worm species.

For clustering the locomotion dataset into FLm patterns, we proposal a top-down hierarchical separating and regulating scheme, as shown in Fig. 3. For instance of grouping 5-frames' behaviorial patterns, 1) the highest order, here 4-order, differences are compared, and the locomotions are clustered into different sub-clusters, each of which has its own global changing tendency, 2) the step 1 continues on the 3-order differences, until the 0-order of differences. With this process going on, the locomotion are regulated from global to local, from coarse to fine into different pattern groups.

We use a density-based clustering method to implement the hierarchical separating procedures. The FLm pattern here is the locomotion with $Minpt$=4 neighbor locomotions within the radius of r_ε=3, which are decided by using heuristic k-dist(LD_i^k) method.

B. Ranking Frequent Locomotion Pattern

The rank to a FLm pattern is an uniqueness or distinctiveness measurement of a FLm pattern to a given species. The uniqueness measurement is carried on with respect to both intra-species frequency and inter-species scarcity as shown in (6). ω_{i,C_j} represents the rank of a FLm pattern i to a class (species) j , and it is determined by two factors. One factor is LDf_{i,C_j}^k, which is the normalized frequency of locomotion LD_i^k in the class C_j. The other is scarcity factor measured by inverse ratio between the number of classes

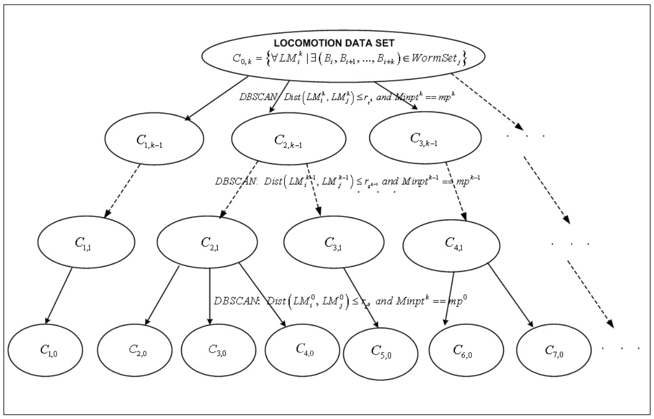

Fig. 3. Hierarchically clustering locomotions into frequent locomotion patterns. $Minpt$ and r_ε of each level are decided by heuristic statistics.

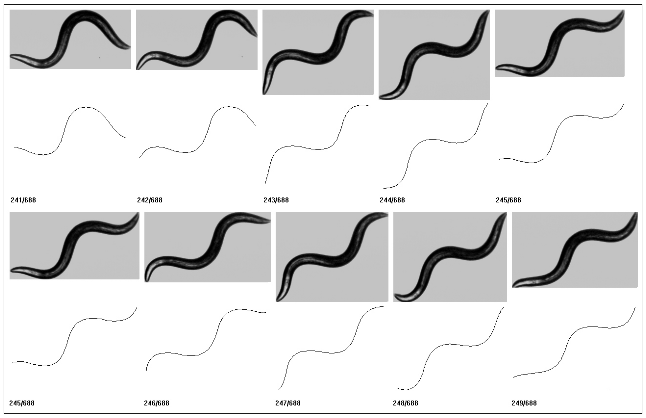

Fig. 4. A representative behavior abstracted from *n2*.

Cf_i that contain LD_i^k and the total number of classes n_c in a given experiment. A locomotion pattern with high rank to a species, is called a representative behavior to that species. A representative behavior of wild type is shown in Fig.4. Our temporal datatset contains 5 species, around 30,000 5-frame locomotions.

$$w_{i,c_j} = LDf_{i,c_j}^k \times \log(\frac{n_c}{Cf_i}), \text{ where} \tag{6}$$

$$LDf_{i,c_j}^k = \frac{n_{LDf_{i,c_j}^k}}{\sum_{l \varepsilon C_j} LDf_{l,C_j}^k}$$

C. Behavior Screen and Correlation Analysis

Continued to above issues, behavior screen and correlation analysis here are designed to find, within a representative behavior, which sub-behaviors (actions) leads and dominates the following sub-behaviors. For a (k+1)-frame representative behavior P_{LB}^{k+1}=$\{B_i, B_{i+1}, ..., B_{i+k}\}$, there are potentially k leading sub-behavior L_{SBhv}=$\{ \{B_i\}, \{B_i, B_{i+1}\}, ..., \{B_i, B_{i+1}, ..., B_{i+k-1}\}\}$, and the k correspondingly following sub-behavior $F_{SBhv} = \{\{B_{i+1}, B_{i+2}, ..., B_{i+k}\}, ..., \{B_{i+k}\}\}$. Such two measurements as $Confidence(L_{SBhv} \Rightarrow f_{SBhv})$ and $Correlation(L_{SBhv} \Rightarrow F_{SBhv})$ are used to extract the interesting rules, and the j-th rule $\{L_{SBhv}(j) \Rightarrow F_{SBhv}(j)\}$ is measured as in (7) and (8). The rules with high confidence and positive correlation are regarded as interesting rules, and only the leading sub-behaviors from interesting rules are the dominating sub-behaviors. Similarly, the consequences between sequential behaviors also can be discovered.

Behavior screen and correlation analysis here are used to pinpoint those sub-behaviors (key actions), which is hoped to directly mirror the known similarities in molecular function and cellular site of action of the mutant gene products.

IV. CONCLUSION

We have presented a model to quantitatively analyzing locomotion patterns of the *C. elegans* using automatically tracking and time-coded image sequences. The intuitive worm shape and locomotion representation and their hierarchical comparing structure are proved to be efficient for high-throughput worm locomotion mining purpose. And the locomotion ranking and behavior screen schemes set a step toward automatic behavioral screen to *Caenorhabditis elegans*.

$$\begin{aligned} &Confidence(L_{SBhv}(j) \Rightarrow F_{SBhv}(j)) \\ &= \frac{P(\{B_i, B_{i+1}, ..., B_{i+k}\})}{P(\{B_i, B_{i+1}, ..., B_{i+j-1}\})} \end{aligned} \tag{7}$$

$$\begin{aligned} &Correlation(L_{SBhv}(j) \Rightarrow F_{SBhv}(j)) \\ &= \frac{P(L_{SBhv}(j) \bigcap F_{SBhv}(j))}{P(L_{SBhv}(j))P(F_{SBhv}(j))} \\ &= \frac{P(\{B_i, B_{i+1}, ..., B_{i+k}\})}{P(\{B_i, B_{i+1}, ..., B_{i+j-1}\})P(\{B_{i+j}, B_{i+j+1}, ..., B_{i+k}\})} \end{aligned} \tag{8}$$

V. ACKNOWLEDGMENTS

This research is partially supported by the Korean Air General Electric Research Institute for International Collaboration (KAGERIIC).

REFERENCES

[1] S. Brenner, The Genetics of *Caenorhabditis elegans*, *Genetics* vol. 77, 1974, pp. 71-94.

[2] W. Zhong and W. Sternberg, Genome-wide Prediction of *C. elegans* Genetic Interactions, *Science* Vol.311, no.5766, 2006, pp.1481-1484.

[3] J. Chronis, M. Tsunozaki, J.M. Gray, D. Ramot, M.B. Goodman and C.I. Bargmann, Dissecting a Circuit for Olfactory Behaviour in *Caenorhabditis elegans*, *Nature* Vol.450, 2007, pp.63-70.

[4] J.H. Baek, P. Cosman, Z. Feng, J. Silver and W.R. Schafer, Using Machine Vision to Analyze and Classify *C. elegans* Behavioral Phenotypes Quantitatively, *J. Neurosci. Methods* Vol.118, 2002, pp. 9-21.

[5] W. Geng, P. Cosman, C.C. Berry, Z. Feng and W.R. Schafer, Automatic Tracking, Feature Extraction and Classification of *C. elegans* Phenotypes, *IEEE Trans. Biomedical Engineering*, Vol.51, 2004, pp.1181-1820.

[6] Z. Feng, C.J. Cronin, J.H. Wittig, P.W. Sternberg and W.R. Schafer, An Imaging System for Standardized Quantitative Analysis of *C. elegans* Behavior, *BMC Bioinformatics* Vol.5, 2004, pp.115.

[7] Cronin, J. E. Mendel, S. Mukhtar, Y. Kim, R.C. Stirbl, J. Bruck and P.W. Sternberg, An Automated System for Measuring Parameters of Nematode Sinusoidal Movement, *BMC Genetics*, 2005, http://wormlab.caltech.edu/.

[8] B.T. Zhou and J.H. Baek, Using Machine Vision to Detect Distinctive Behavioral Phenotypes of Thread-shape Microscopic Organism, *Book Series of Studies in Computational Intelligence: Applications of Computational Intelligence in Biology Current Trends and Open Problems*, vol.122, 2008, pp.161-182.