Quantitative Analysis of *C. elegans* Mutant Type Using Movement and Reversal Features

Won Nah and Joong-Hwan Baek

School of Electronics, Telecommunication & Computer Engineering
Hankuk Aviation University, Koyang City, 200-1, Korea
E-mail:{nahwon, jhbaek}@mail.hangkong.ac.kr

Abstract: *Caenorhabditis (C.) elegans* is often used in genetic analysis in neuroscience because its simple organism; an adult hermaphrodite contains only 302 neuron. So the worm is often used to study of cancer, alzheimer disease, aging, etc. To analysis mutant type of the worm, an experienced observer was able to subjectively before, but requirements for objective analysis are now increasing. For this reason, we use automated tracking systems to extract global movement coordinate of the worm. In this paper, we extract features, which are related on reversal and movement of the worm. Using these features, we quantitatively analysis 6 type mutant by movement and reversal characteristic.

Keywords: Pattern Recognition, Computer Vision, Image Processing

1. INTRODUCTION

C. elegans is a small, free-living soil nematode. It is an ideal genetic model system, which has three advantages for genetic analysis when compare with other animals. First, *C. elegans* has a relatively short life cycle. Three days are enough to observe its next generation under optimal conditions. Second, *C. elegans* has a hermaphroditic reproduction system. A hermaphrodite is a bi-sexual state that can fertilize itself and reproduce its own progenies. That is, it can produce both sperms and oocytes so that a particular strain can be maintained without male-female mating at each generation. Third, It is possible to store *C. elegans* for a long time by freezing in liquid nitrogen.

C. elegans is about as primitive an organism that exists which nonetheless shares many of the essential biological characteristics that are central problems of human biology. A result of genome project, *C. elegans* gene is similar to human gene about 40% and 75% of 5,000 known human gene is shared. In this reason, the worm is often used to study of cancer, alzheimer disease, aging, etc.

Understanding the relationship between genes and the behaviour of *C. elegans* is a fundamental problem in neuroscience. An experienced observer was able to subjectively distinguish worm types before. But it is often imprecise and unreliable. For this reason, automated classification systems using machine vision appeared for the purpose of objective classification. In the previous works [1], [2], classification was automated using the patterns from reliable egg-laying event timing data.

We use an automated tracking system, which make it possible to measure the rate and direction of movement for each worm and to compute the frequency of reversals in direction. This system is designed to follow an individual worm at high magnification. For image preprocessing, we perform the thresholding and median filtering first, and then detect the holes using the worm's thickness. Finally we obtain clean binary image. From the binary image, we extract the features related to the movement and reversals of the worm. We also analyze each worm type using the hierarchical clustering method. In this paper, wild type and its five mutant types (goa-1, nic-1, unc-36, unc-38, and egl-19) are used to analyze.

2. IMAGE FEATURE EXTRACTION

In order to obtain quantitative values of the features related with worm's movement and reversal, we use a computer vision system, which can track a worm automatically by PC. Here image preprocessing is performed to get clean binary image and extract movement and reversal features.

2.1 Image Acquisition System

C. elegans locomotion is tracked with a Zeiss Stemi2000-C Stereomicroscope mounted with a Cohu High Performance CCD video camera (See Fig. 1). A computer-controlled tracker (Parker Automation, SMC-1N) is used to put the worms in the center of optical field of the stereomicroscope during observation. *C. elegans* is small (1mm) worm. Thus, high magnification (50×) requires to image processing and feature extraction. But in high magnification, the worm can quickly outside the field of view. So our tracking algorithm typically processes 14-15 frame/second and rarely loses track of the worm during image acquisition.

The frame grabber captures 640×480 images with 256 gray scale values. Since resolution is not critical for locating the center of the worm, we first down-sample the original image to get 160×120 image. By this procedure, we can save a great deal of computation time. Then, we apply a threshold and get a binary image. First, we calculate the mean (m) and standard deviation (σ) of the current gray level image. The threshold is chosen as

$T=m-2\sigma$. Using this threshold, we can get a binary image. Then we apply a flood-fill operation [3] to fill any noise, which is caused by transparent part of the worm. Second, we calculate the centroid of the worm using binary image, which is a black (background) and white (worm body) image. With a clean binary image, it is then possible to calculate the centroid of the worm.

The centroid coordinates are then multiplied by 4 and mapped onto the original gray level image. Using this centroid the computer sends a command to the stage controller to re-center the field of view at these coordinates.

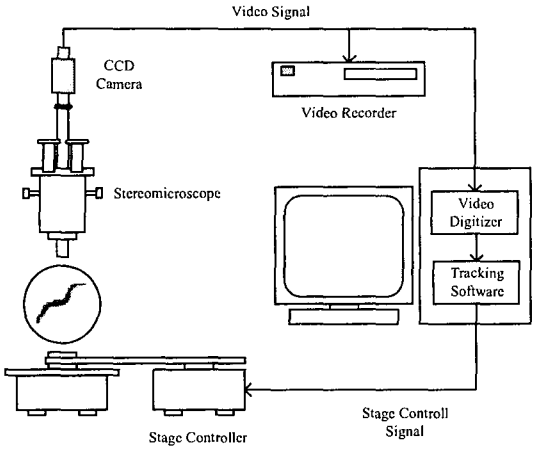


Fig. 1. Image acquisition system.

2.2 Image Processing

Image preprocessing is the most important factors for the successful analysis of *C. elegans* [4]. In this paper, we propose a new binarization method using hole detection. The method is performed with the following 3 procedures: Thresholding, median filtering, and hole detection.

To decide upon the threshold, a 5×5 moving window was used for scanning over the experimental image, and the mean and standard deviation of the 25 pixels inside the window were computed at every pixel position. The pixel is estimated experimentally to be on the worm's body when the intensity value of the pixel is about 70% of the background intensity. Note that the background intensity level is higher than that of the worm's body. Also, pixels on the worm's body tend to have a larger variance in their intensity values than pixels on the background. So when the standard deviation of the pixels within the window is over 30% of the mean, we consider the pixel as a part of the worm's body.

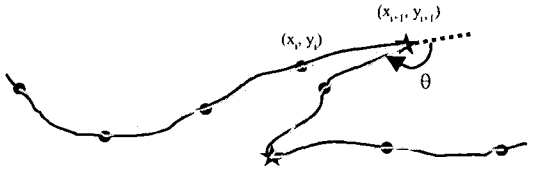
And we performed median filtering. A median filter has a superior effect for removing impulse noise [5]. Median filtering can preserve small sized holes and remove impulse noise, which is caused by reflecting on the worm's body. In binary image, median filter is easily performed by comparison between the number of '1' in the window, *count*, and the window size, $m \times n$. Even after applying the median filter to the binary worm image, some noise occasionally remains on the worm's body. To remove the remaining noise, we propose a method that can distinguish between hole and noise. A hole is created when the worm loops or touches itself, while noise is located on the worm's body. Therefore, we can determine whether the region is a hole or noise by measuring the total thickness of the body enclosing the region.

In order to measure the thickness, we define 16 vectors. We traverse and count the number of pixels from the centroid of a region in each vector direction until we reach the background. Then we compute the thickness by multiplying the magnitude by the number of pixels traversed. The total thickness is the sum of two opposite direction thicknesses. So, among the 8 total thicknesses, if the minimum is less than 25, the region is considered as noise, because the thickness of the worms used in this work is not larger than 25. If a region is determined as noise, we fill the region with body pixels. Otherwise, we preserve the region as a hole. These procedures are repeated for all of the labeled regions. After detecting holes, we remove the remaining noise using a closing morphological operation.

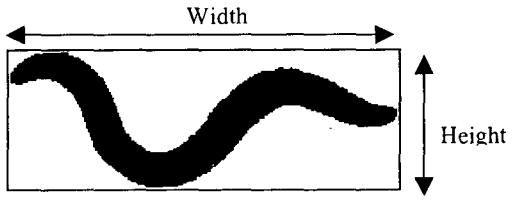
2.3 Feature Extraction

After binarization and preprocessing such as skeletoning, feature extraction is performed. Features for the large-scale movement are global moving distance and reversal frequencies during some intervals. To measure moving distance, the centroid position data are sampled over a constant time interval (0.5 sec). And to measure reversal, the trajectory of the centroid is sampled at intervals of constant distance (30 pixel, which is one-tenth of the normal worm length). Fig. 2 (a) shows directional change detection method. The trajectory of the worm's centroid (black solid line) is sampled at intervals of 30 pixels. The directional change position (mark with a star) is found by computing the angle deviation at every vertex of the polygon (gray line). If the angle (θ) is greater than 120°, then the position is considered to be a reversal.

$$\theta = \left| \frac{\pi}{180}\arctan(y_{i+1}-y_i) - \frac{\pi}{180}\arctan(x_{i+1}-x_i) \right|$$
$$\text{if} \quad \theta > 180, \qquad \theta = 360 - \theta \qquad (1)$$
$$\text{else} \quad \theta = \theta$$



(a)



(b)

Fig. 2. (a) Directional change detection method , (b) MER (Minimum Enclosing Rectangle).

We also compute the moving distance of normalized centroid (CNTMV). The coordinates of the worm's centroid are normalized with the MER (See Fig. 2(b))

width and height. Normalization and moving distance are computed with the following equation:

$$\mathbf{C}_n = (\frac{C_w}{width}, \frac{C_h}{height}), \quad \mathbf{C} = (C_w, C_h) \qquad (2)$$
$$CNTMV = |\mathbf{C}_{n,t} - \mathbf{C}_{n,t-1}|$$

We subdivide the interval for measuring the reversals. The number of reversals is measured during 10, 20, 30, 40, 50, and 60 sec.

3. EXPERIMENTAL RESULT

C. elegans locomotion is tracked with a stereomicroscope mounted with a CCD camera. A computer-controlled tracker is used to put the worms in the center of the optical field of the stereomicroscope during observation. To record the locomotion of a worm, an image frame is snapped every 0.5 seconds for 5 minutes. So the video clip of each worm consists of 600 frames. All of the software for binarization and feature extraction is coded in C++ and implemented on a PC with a 1.7GHz CPU.

In this experiment, we use 6 different worm types (wild, goa-1, nic-1, unc-36, unc-38, egl-19). Each worm type has 100 worms. So a total of 600 worms are used in this experiment. Primary features are extracted from each frame after binarization and preprocessing.

Fig. 3 shows the typical trajectory of a wild type worm. Reversal event is indicated by the rectangle symbol. Fig. 4 shows that there are big differences between each mutant in movement and reversal characteristic. Wild and goa-1 have relatively high values in average moving distance in 0.5 sec and total reversals, while nic-1 has the lowest value. And unc-36, unc-38 and egl-19 have moderate value. These mean that wild or goa-1 type worms are tend to be hyper active, while nic-1 is very sluggish.

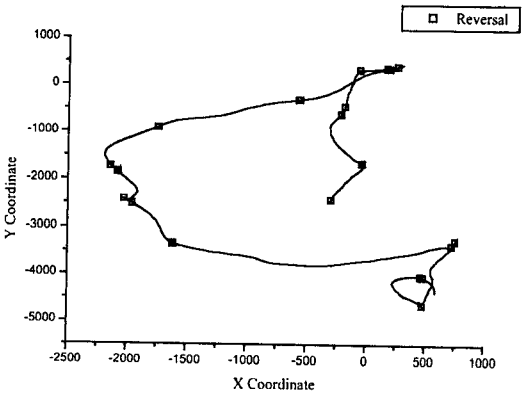
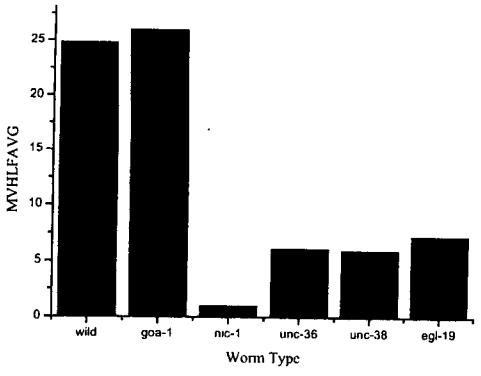
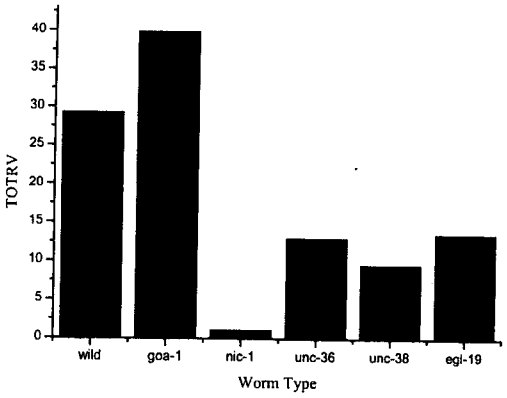


Fig. 3. Trajectory of a wild type worm.



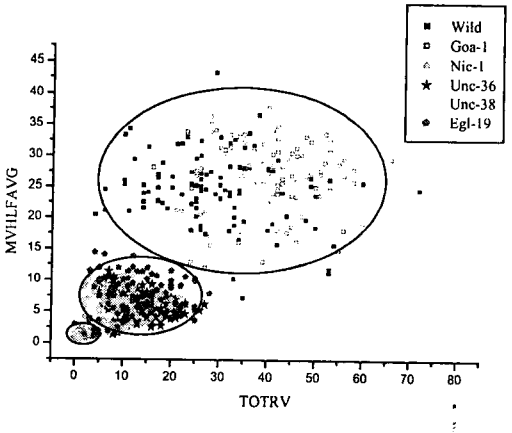
(a)



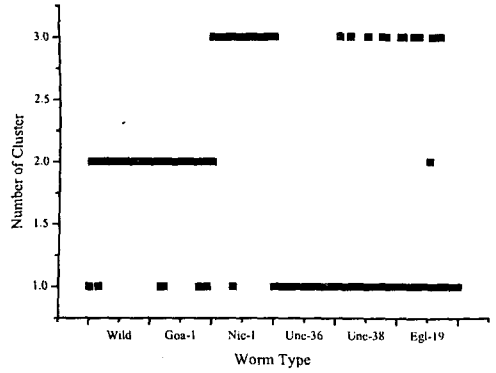
(b)

Fig. 4. (a) Average moving distance in 0.5 sec, (b) total reversals.

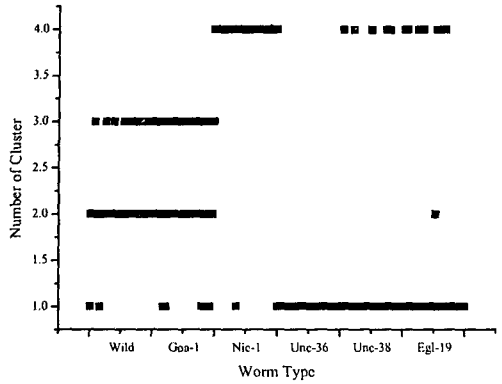
To analysis the movement and reversal of each mutant, we use a hierarchical clustering. Each worm type is assigned to its own cluster and then the algorithm proceeds iteratively, at each stage joining the two most similar clusters, continuing until there is just a single cluster. At each stage distances between clusters are recomputed by the Lance-Williams dissimilarity update formula according to the particular clustering method being used. And we use Ward's minimum variance method aims at finding compact, spherical clusters.



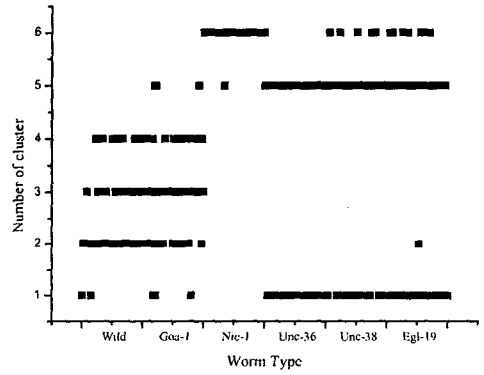
(a) Worm type plot using two features



(b) Hierarchical clustering result (k=3)



(c) Hierarchical clustering result (k=4)



(d) Hierarchical clustering result (k=6)

Fig. 5. Hierarchical clustering results using two features (TOTRV, MVHLFAVG).

In Fig. 5 (a), we can see intuitively that there are 3 rough clusters. As the result of Fig. 5 (b)~(d), each of them implements hierarchical clustering tree cut by parameter k, we can see appropriate cluster number (k=3). Wild and goa-1 has relatively high values both the number of total reversal and average distance moved in 0.5 sec. Unc-36, unc-38 and egl-19 mutant type has middle values and nic-1 has low values. In this result, we can see that each mutant type can be roughly classified into 3 clusters by the movement and reversal features. And it means that mutant type of the worm, in any cluster, has similar characteristics in reversal and movement.

4. CONCLUSION

We have described a system that allows automated recording and real-time analysis of several parameters of C. *elegans* nematode behavior. By following changes in the animal's position and using this positional information to direct the movement of a motorized microscope stage, it is possible to record the animal's behavior at high magnification for indefinitely long time periods. In addition, the system is capable of gathering real-time data on the rate and direction of the animal's movement. Using this movement information, it is possible to automatically detect a reversal event.

Finally, using these features, which are related on movement and reversal, we can analysis behavioral characteristics of each mutant. And we can roughly classify into 3 clusters (wild, goa-1; unc-36, unc-38 and egl-19; nic-1). The mutant in same cluster is similar to each other and that can't easily distinguish using only movement and reversal features described here. For further work, more distinguishable features for unc-36, unc-38 and egl-19 should be extracted.

Acknowledgements

This research was partially supported by IRC (Internet Information Retrieval Research Center) in Hankuk Aviation University. IRC is a Kyounggi-Province Regional Research Center designated by Korea Science and Engineering Foundation and Ministry of Science & Technology.

References

[1] Waggoner, L., et al., "Control of behavioral states by serotonin in Caenorhabditis elegans", *Neuron* pp. 203-214, 1998.

[2] Zhou, G.T., Schafer, W.R., Schafer, R.W., "A three-state biological point process model and its parameter estimation", *IEEE Trans On Signal Processing*, pp. 2698-2707, 1998.

[3] J.D. Foley, A. van Dam, S.K. Feiner, and J.F. Hughes., *Computer Graphics Principles and Practice, Addison-Wesley Publishing Company*, 1997.

[4] Baek, J.H., et al., "Using machine vision to analyze and classify Caenorhabditis elegans behavioral phenotypes quantitatively", *Journal of Neuroscience Methods*, Vol. 118. pp. 9-21, 2002.

[5] Gonzalez, R.C., Woods, R.E., *Digital Image Processing, New Jersey, Prentice Hall Inc.*, 2002.

[6] Jain, R., Kasturi, R., Schunck, B.G., *Machine Vision, New York, McGraw-Hill Inc.*, 1995.