SOCIAL BEHAVIOR ANALYSIS OF DROSOPHILA LARVAE VIA MOTION ACTIVITY RECOGNITION

Micah Consylman[1], Suvadip Mukherjee[1], Dipti Prasad Mukherjee[3], Barry Condron[2], and Scott T. Acton[1]

[1]Charles L. Brown Department of Electrical and Computer Engineering, University of Virginia, USA 22904.
[2]Department of Biology, University of Virginia, USA 22904.
[3]Electronics and Communication Sciences Unit, India Statistical Institute, India 700108.

ABSTRACT

We aim to analyze the social behavior of Drosophila larvae through automated identification of their motion activity pattern. Social activity recognition is an essential step in identifying the correlation between specific neuronal structures in their brain to the social behavior of the species. Larvae videos are obtained using low-cost overhead video; subsequently, the motion features of the larvae generated by computing a gradient weighted optical flow. These features are used to create a global histogram of gradient-weighted optic flow over a specific time frame. The method for estimating optical flow is based on an implementation by Gautama and Van Hulle that expands on the classical Horn and Schunck method by using spatial Gabor filter pair to find the intensity gradient and recurrent neural networks to find velocity. Activity classification is obtained by using a support vector machine and yields 82% overall accuracy in classification.

Index Terms— Activity recognition, HOG, optical flow.

1. INTRODUCTION

Understanding the social behavior of the *Drosophila* larvae serves a crucial role in analyzing their sensory and cognitive functions. State of the art research in this field has yielded many insights due to how well the relatively simple neural system of the *Drosophila* can be scaled to more complex neural systems such as the human brain [1].In the past few years, image analysis scientists have been actively involved in studying the microscopic imagery of the *Drosophila* brain [2] [3] [4] to analyze their neuronal connectivity. Whereas the microscopic analysis of the brain provides insight into the behavior of the animal, in this paper we aim to identify specific activities in a larva from a macroscopic perspective. An automated method to identify the set of activities of a larva is necessary to analyze their social behavior, which is subsequently related to the neural anatomy of the species.

The goal of our method is to allow for the automated tracking and activity classification of a larva with the imaging tools already available to most neuroscience and biology research facilities such as low cost computer video cameras and mobile devices. The downside of many low cost video solutions is that they suffer from pixel count and short distance focus limitations. Because of these limitations we aimed to provide a solution that would not require specialized lenses for focus. We also wanted to have a solution that would enable video to be shot from a distance that would encompass all larvae under test so that devices could be set up and left unattended. Fig. 1 shows an example of the video shot under our assumed conditions.

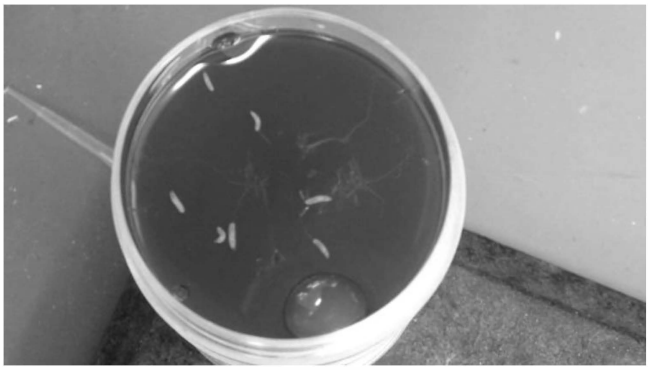

Figure 1: A single frame of a video observing *Drosophila* larvae on a petri dish.

Broadly, our approach includes two steps: an initial framework to automatically track the larva, which is allowed to move on a petri dish, and an analysis framework for the motion behavior of the tracked animal. Because of the resolution limitation and intended distance of the camera we assumed that each individual larva may only be represented by a small number of pixels. Thus, our activity recognition method does not rely on any recognizable anatomy of the larvae. Instead, we use a reduced set of optical flow information to generate features [5]. We sought to use a classifier that would provide relatively fast predictions on a feature space that may not be easily separable and thus employed the support vector machine (SVM) [6].

2. AUTOMATED TRACKING

To classify the motion activity of the *Drosophila* larvae, the first step is to be able to track the larvae from the video frames. The *Drosophila* larvae are placed on a petri dish and their actions are recorded using a simple video camera. To track the larvae, we use a local version of the Chan-Vese implementation of region based active contour [7]. The zero level set of the embedding function converges to the larva boundary at each frame. To track a given larva in next frame, the level set function is initialized to the one obtained from the previous frame, and then allowed to evolve until convergence. Correspondence resolution is performed using a simple nearest neighbor policy to determine position and boundary of the larva in successive frames (Fig. 2*)*.

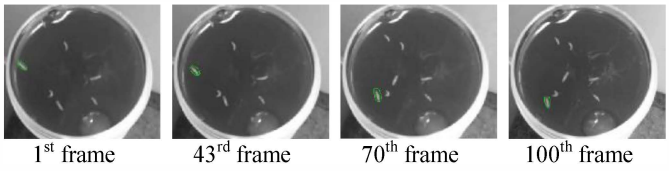

Figure 2: Tracked larva from the video frames. The tracked boundary is shown in green.

3. ACTIVITY RECOGNITION

Once the larva is tracked from the video frames, the next goal is to be able to recognize the motion activities of the larva. The feature set to represent a particular activity was computed by using a Histogram of Oriented Gradients (HOG) of the gradient weighted optical flow field. For the purpose of this study, three actions were chosen that could be performed independently by one larva and were readily identifiable by an observer over a short time period. The three actions chosen were: forward crawling, turning left and turning right. These actions are illustrated in Fig. 4.

3.1. Training Set and Test Set preparation

If two larvae collide a flag is triggered to alert the user of the possible need for manual intervention as the current tracking method only follows a single larva, however in the tested video there were no such collisions. Once the larva is tracked, a sub-image is created surrounding the larva. An individual larva is rotated such that the trailing half of the major axis faces due north in the direction of motion and is scaled to a normalized size by the total length. After rotation and scaling, 31 consecutive frames are translated and cropped in a small region around the larva. We assume that an activity can be observed in this set of 31 consecutive frames. We refer to these sets of frames as *packets.* Fig. 3 shows an example of the sub-image framing and Fig. 4 shows three examples of sub-image packets from each of the different activities.

3.2. Feature extraction

Pixel motion is calculated between consecutive frames in a *packet* using a phase based approach to find optical flow. The method for estimating optical flow is based on an implementation by Gautama and Van Hulle [8] that expands on the classic Horn and Schunck [9] method by using spatial Gabor filter pairs [10] to find the intensity gradient and recurrent neural networks to find velocity. Due to the noise incurred in imaging, optical flow vectors are weighted by the sum of the gradient magnitude between frames in order to eliminate optical flow that is generated from pixel noise or small variations in illumination.

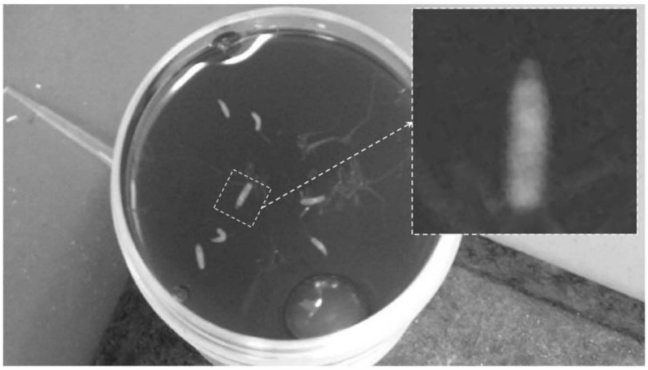

Figure 3: An example of the sub-image framing process from a test video.

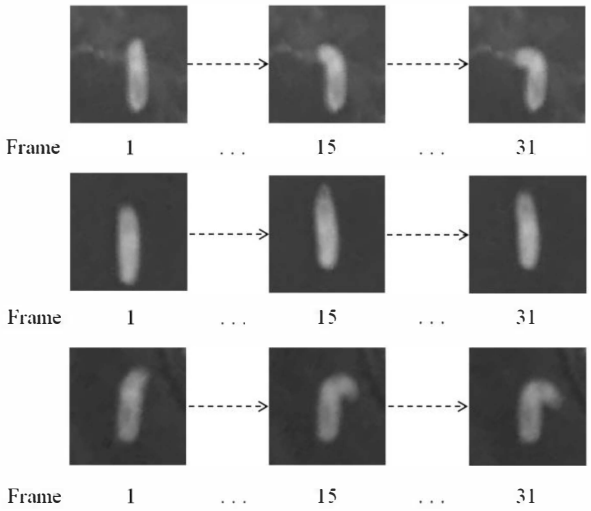

Figure 4: Three different actions (from top to bottom): left turn, forward crawl and right turn.

The resulting gradient weighted optical flow vector can be computed using

$$O^f_{filtered}(i,j) = O^f_{raw}(i,j)\left[G_f(i,j) + G_{f+1}(i,j)\right]$$

where (i,j) is the pixel position, f is the current frame index, $O^f_{filtered}$ is the filtered optical flow vectors, and O^f_{raw}

is the optical flow vectors over the entire image. The gradient magnitude is defined as $G_f = |\nabla I_f|^2$ where I_f is the image at frame f. This process is illustrated in Fig. 5. The set of feature vectors is obtained by quantizing the angle of every gradient vector to one of eight directions.

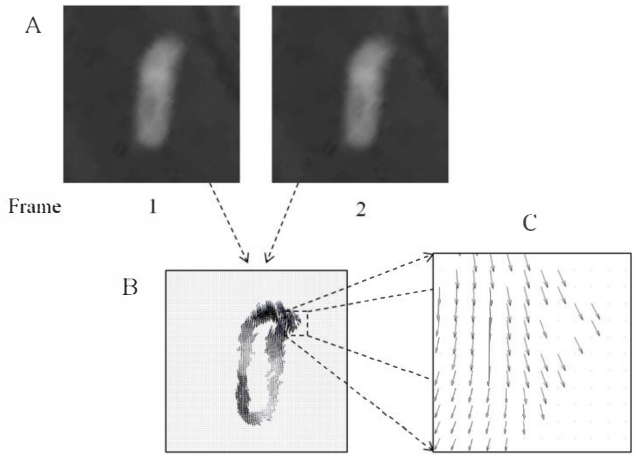

Figure 5: (A) Two frames from a right turn action, (B) gradient weighted optical flow and (C) a close up of the region over the tip of the head showing optical flow vectors.

Following the quantization, a global histogram of gradient-weighted optic flow [11] is generated by adding the magnitude of each direction quantized optical flow vector over the entire packet, resulting in a histogram containing eight bins. An example of this process is shown for one optical flow vector in Fig. 6.

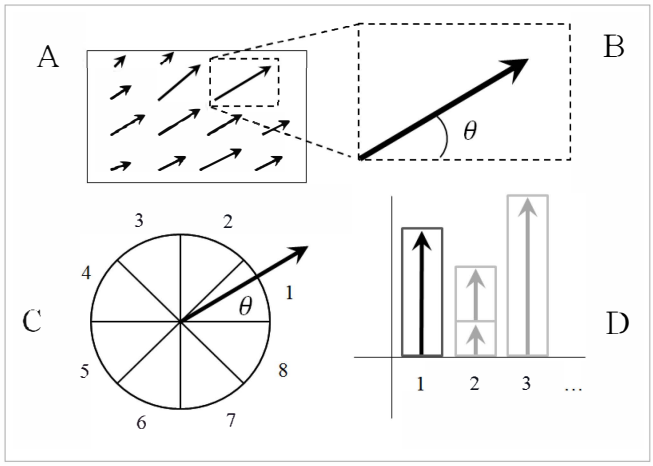

Figure 6: Feature extraction process: (A) Each optical flow vector is examined at every point in the image, (B) the angle is determined and (C) quantized to 1 of 8 directions and (D) the magnitude of the vector is then added to the corresponding bin in the histogram.

3.3. Activity training and prediction

Three specific larva motion activities, viz. forward crawling, left turn and right turn, were identified by a human expert, which are important to analyze their social behavior. The activities are encapsulated in packets of 31 consecutive frames. Each activity is trained with the HOG feature vector from individual packets. Packets from each action are used in order to find maximal separation over the many variations of one particular action. The Support Vector Machine (SVM) classifier was chosen for its established flexibility in kernel choice [6] [12] [13] and its ability to make better predictions as more data is acquired and the rapid speed at which classification can be made for new data once training is complete [14]. For our testing, we chose a linear kernel for training the SVM. Since the SVM is inherently a binary classifier, our method uses a simple "one-against-all" approach [15] to allow for multiple activities to be compared. In order to implement a "one-against-all" approach each of the three activities is tested against the other two. The prediction that has the highest confidence value is considered the winner.

3.4. Activity recognition overview

The general algorithm for activity recognition is as follows:

1. Create a normalized sub-image packet from the tracked larva.
2. Compute the optical flow between each frame in the packet.
3. Filter the optical flow vectors by taking the product of the weighted gradient magnitudes from surrounding frames.
4. Create a HOG by adding the magnitude of each filtered optical flow vector into 1 of 8 bins.
5. Use the HOG feature vector to train the SVM classifier.
6. Use the SVM classifier to predict the activity of new test data.

4. TESTING DATA

Testing was performed on video data that provided a single stationary overhead view of a petri dish containing seven *Drosophila* larvae from an off the shelf, low cost video camera (as seen in Fig. 1). The video was shot in an open laboratory setting so some minor shifts in illumination occur throughout the length of the video. The overall area of each larva covered less than 0.1% of the total imaged area of the camera frame.

5. RESULTS

The method was tested using 30 packets from each activity, for a total of 90 packets from the test video. The results of the test were compared to manual classifications made by an expert to determine accuracy. Due to the relatively small available data set we used leave-one-out cross validation to

check each video packet in the data set against a SVM trained with the rest of the packets.

Our method produced correct predictions for 82% of all tested data. Of the total predictions crawling forward was predicted correctly 93% of the time, turning left was predicted correctly 90% of the time while turning right was only predicted correctly 66% of the time. Right turn activities were less common than the other two activities and so actions with less optical flow information were accepted for use in the training set. This seems to have contributed to the significantly lower accuracy of classification for right turn activities. The results of the confusion matrix can be seen in Fig. 7, where each row shows the actual behavior and each column shows the behavior predicted by the classifier. Correct predictions are seen along the diagonal of the table.

To analyze the statistical behavior of our method 5-fold cross-validation was performed. So that all actions were tested against a training set of the same size, two of the six samples from each of the five subsets were randomly selected from each action. The result of twenty 5-fold cross-validation tests yielded correct predictions 81% of the time with a variance of 0.018%.

		Predicted Behavior		
		Forward Crawl	Left Turn	Right turn
Actual Behavior	Forward Crawl	93%	7%	0%
	Left Turn	0%	90%	10%
	Right Turn	3%	33%	66%

Figure 7: The confusion matrix for the tested data with three activity classifications.

6. CONCLUSIONS AND FUTURE WORK

In this paper we have presented a simple framework to automatically identify the motion activity in *Drosophila* larvae. Initial experiments show promising results. However, the relatively lesser number of occurrences of right turns in the acquired video meant that, many occurrences used to train the SVM had to consist of subtle motions. In these cases it was possible for a human to determine what motion was occurring but overall pixel noise and illumination differences resulted in the automated method failing. For this reason future testing will attempt to eliminate more spurious details from the image in the tracking phase. Additionally, kernel variations will be investigated for the training of the SVM with the hope that it may provide better separation in feature space between actions that contain similar optical flow information. Future work will also address more robust means of multiple larvae to eliminate the possible need for manual intervention.

As our method improves we seek to realize a software implementation that can be used by neuroscientists and biologists on preexisting or low cost hardware. We conclude that such a system will be useful in analyzing the social behavior of the *Drosophila* and of other organisms of similar scale.

7. REFERENCES

[1] K. Rein, M. Zöckler, M. T. Mader, C. Grübel and M. Heisenberg, "The *Drosophila* Standard Brain," *Current Biology,* vol. 12, no. 3, pp. 227-231, 2002.

[2] S. Basu, B. Condron and S. T. Acton, "Path2Path: Hierarchical path-based analysis for neuron matching," in *International Symposium on Biomedical Imaging*, Chicago, 2011.

[3] D. Mayerich, C. Björnsson, J. Taylor and B. Roysam, "Metrics for comparing explicit representations of interconnected biological networks," *BioVis,* pp. 79-86, 2011.

[4] S. Basu, A. Aksel, B. Condron and S. Acton, "Tree2Tree: Neuron Segmentation for generation of neuronal morphology," in *ISBI*, 2009.

[5] S. Mukherjee, S. K. Biswas and D. P. Mukherjee, "Recognizing Human Action at a Distance in Video by Key Poses," *IEEE Trans. Circuits and Systems for Video Technology,* vol. 21, no. 9, pp. 1228-1241, 2011.

[6] J. A. Suykens and J. Vandewalle, "Least squares support vector machine classifiers," *Neural processing letters,* vol. 9, no. 3, pp. 293-300, 1999.

[7] T. F. Chan and L. A. Vese, "Active Contours Without Edges," *IEEE Trans. on Image Processing,* vol. 10, 2001.

[8] T. Gautama and M. M. Van Hulle, "A phase-based approach to the estimation of the optical flow field using spatial filtering," *IEEE Trans. Neural Networks,* vol. 13, no. 5, pp. 1127-1136, 2002.

[9] B. Horn and B. Schunck, "Determining optical flow," *Artificial Intell.,* vol. 17, p. 185–203, 1981.

[10] D. Gabor, "Theory of communication. Part 1: The analysis of information," *J. Inst. Elect. Eng.,* vol. 93, no. 26, pp. 429-441, 1946.

[11] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition,* vol. 1, pp. 886-893, 2005.

[12] V. Kecman, Learning and Soft Computing, Cambridge, MA: MIT Press, 2001.

[13] B. Scholkopf and A. Smola, Learning with Kernels, Cambridge, MA: MIT Pres, 2002.

[14] K. M. Chathuramali and R. Rodrigo, "Faster human activity recognition with SVM," *2012 International Conference on Advances in ICT for Emerging Regions (ICTer),* pp. 197-203, 12-15 Dec 2012.

[15] C.-W. Hsu and C.-J. Lin, "A comparison of methods for multiclass support vector machines," *IEEE Trans. Neural Networks,* vol. 13, no. 3, pp. 415-425, 2002.

[16] S. Basu, A. Aksel, B. Condron and S. T. Acton, "Segmentation and tracing of neurons in 3D," *IEEE Transactions on Info. Tech. Biomedicine,* 2013.