

# Computational model of cortical neuronal receptive fields for self-motion perception

Chen-Ping Yu<sup>1</sup>, Charles Duffy<sup>2</sup>, William Page<sup>2</sup>, and Roger Gaboriski<sup>3</sup>

<sup>1</sup>Department of Computer Science and Engineering  
The Pennsylvania State University, University Park, PA 16801 USA  
cxy171@cse.psu.edu

<sup>2</sup>Department of Neurology  
University of Rochester Medical Center, Rochester, NY 14623 USA  
Charles\_Duffy@urmc.rochester.edu, bp@cvs.rochester.edu

<sup>3</sup>Department of Computer Science  
Rochester Institute of Technology, Rochester, NY 14623 USA  
rsg@cs.rit.edu

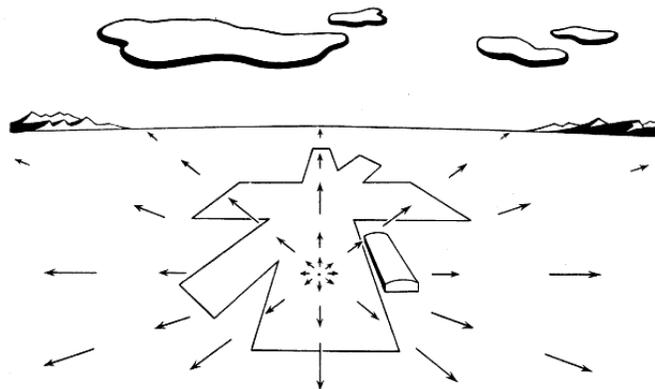
**Abstract**—Biologically inspired approaches are an alternative to conventional engineering approaches when developing complex algorithms for intelligent systems. In this paper, we present a novel approach to the computational modeling of primate cortical neurons in the dorsal medial superior temporal area (MSTd). Our approach is based on a spatially distributed mixture of Gaussians, where MST's primary function is detecting self-motion from optic flow stimulus. Each biological neuron was modeled using a genetic algorithm to determine the parameters of the mixture of Gaussians, resulting in firing rate responses that accurately match the observed responses of the corresponding biological neurons. We also present the possibility of applying the trained models to machine vision as part of a simple dorsal stream processing model for self-motion detection, which has applications to motion analysis and unmanned vehicle navigation.

**Index Terms**—biologically plausible system; mixture of Gaussians; MST single neuron receptive field model; genetic algorithm, motion detection, self-motion analysis.

## I. INTRODUCTION

As the development of intelligent system advances, numerous automated systems and algorithms have been introduced to enhance artificial intelligence and computer vision. In computer vision for motion analysis, many engineering algorithms attempt to understand optic flow in order to monitor self-movement and track moving objects [8][13]. A biologically inspired algorithm is an alternative to conventional engineering methods. Biologically inspired algorithms can be developed by simulating perceptual and neuronal process in primates' brains, and applying principles derived from those simulations to understand the world in ways that are more related to biological intelligence.

The human visual cortex consists of two primary pathways that process visual information, commonly called the ventral stream and the dorsal stream [18]. The ventral stream, also known as the what-pathway, processes the objects in a scene to support detection and recognition. The dorsal stream, also



Adapted from Gibson (1950)

Fig. 1. This commonly seen figure illustrates the perception of optic flow when moving forward.

known as the where pathway, processes location and motion information, including the self-motion cues in full-field optic flow [19][20]. Fig. 1 illustrates optic flow as the visual motion that is created by a person's self-motion [17]. In this paper, we propose a computational model of MST (medial superior temporal area) neurons, that are an advanced stage in the dorsal stream.

The dorsal stream processes visual motion data from V1 (primary visual cortex) to V2, MT (medial temporal area), and then MST [18] as illustrated in Fig. 2. V1 and V2 preprocess the visual input's multi-dimensional information and relay its output to MT for local motion and small pattern motion analysis. MT further consolidates and transforms visual motion information and transfers the results of its analysis to MST, where full-field optic flow is processed for self-motion detection.

We used an array of mixture of Gaussians models, derived

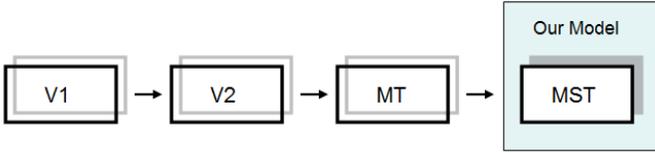


Fig. 2. The dorsal stream information flow process.

from genetic algorithms, to characterize the local motion sensitivities of MST neuronal receptive fields and explore the relations of those responses to the optic flow sensitivities of MST neurons. Finally, we apply the single neuron receptive field models with a feed-forward neural network as a biologically plausible machine vision self-motion detection system.

## II. RELATED WORK

A variety of computational models of visual cortical function have been considered. Ventral extrastriate cortical neuronal receptive fields have been modeled using Gabor filters to enable the simulation of object recognition [10]. The vast range of object shapes can be detected and segmented by convolving a group of Gabor filters with specific orientations and angles.

Connections between the neurons and neuron hierarchies in the ventral system have been modeled as Visnet [2], Visnet simulates the neurons structured in layers that propagate information in a hierarchical format. Bayesian probabilistic learning rules are used to sustain or discard connections between randomly initialized nodes, the trained Visnet can be used to recognize objects and patterns.

Dorsal extrastriate visual cortical neurons have been the subject of several efforts to develop biologically inspired models of visual motion processing. Gaussian derivative models have been used to model striate cortical neurons that are thought to be the first steps in cortical motion sensing [15][16], where three orthogonal Gaussian derivative models of different orientations can be combined to detect the velocity and the direction of a moving edge from input video frames. Summed Gabor filters of different orientation and shift are used in a separate instance to detect the motion energy of moving bars [21], and such energy models mimic many aspects of the physiology of primate motion perception. MT neurons have been modeled using sets of von Mises functions to analyze the mixture of sinusoidal grating plaid stimulus [5].

Our work follows the spirit of biologically inspired models. We propose a mixture of Gaussians model of MST single neuron receptive fields. These models are trained by a genetic algorithm using neuronal response data obtained during the presentation of local motion stimuli, and then tested with 16 full-field optic flow stimuli. The trained models should simulate the behavior of the MST neuronal receptive fields, and be able to detect self-motion in a manner that is consistent with the putative primary function of MST.

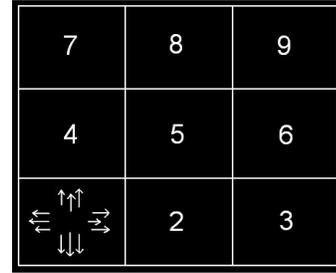


Fig. 3. The setup of single segment recordings. We record a neuron's response one segment at a time, presenting 1 of the 4 planar motion directions at a time. The segment numbers are displayed here, and the 4 planar motion directions is shown in segment 1.

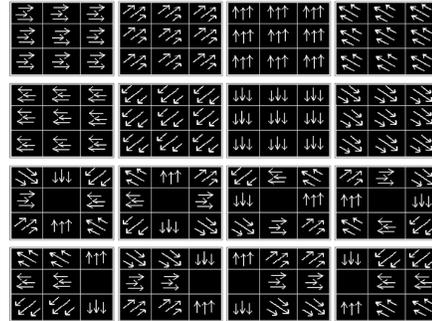


Fig. 4. The local motion patterns summarizing the directions of dot movement in the 16 optic flow stimuli that represents different self-motion sensation. Notice that the 3x3 white dividing lines for each stimulus does not actually display on screen, it is drawn here only for presentation clarity.

## III. DATA

The neurophysiological receptive field data that we model are recordings of 52 MST single neurons' firing rate responses from an adult Rhesus monkey. Each neuron's evoked responses (spikes per second) is recorded after a specific visual motion stimulus shown on a  $90^\circ \times 90^\circ$  rear projection screen while the monkey maintained centered visual fixation. Full-field optic flow is simulated by displaying a radial pattern composed of 500 moving white dots on a black background moving at an average velocity of  $40^\circ$  per second.

### A. Single Segment Recordings

The display screen is partitioned in to a 3x3 array of nine segments. A neuron's responses to each single segment stimulus is recorded by presenting one of the four planar motions ( $0^\circ$ ,  $90^\circ$ ,  $180^\circ$ , and  $270^\circ$ ) at a time (Fig. 3). Since we record 4 planar motions per 9 segments, there is a total of 36 local motion stimuli. Multiple trials of the same stimulus are recorded to derive an average response to each stimulus.

Baseline firing rate was recorded during blank screen periods of centered visual fixation that were interleaved between the local motion or full-field optic flow stimuli. The blank screen intervals provided information on how each neuron responds when there is no visual motion stimulus present. Recording the blank screen periods allowed us to obtain the

baseline firing rate of each neuron for comparison to the stimulus evoked responses.

### B. Full Field Flow Recordings

Our second type of neurophysiological data consists of recordings of single neuron responses to full-field optic flow stimuli simulating 16 different directions of observer self-movement. The optic flow stimuli can be approximated as the combination of 8 different motion directions (4 planar and 4 diagonal directions). This perspective views each direction of moving white dots as occupying one of the 3x3 array of segments with motion that is approximately uni-directional to construct the optic flow stimuli. Fig. 4 shows the 16 different optic flow stimuli that are generated from combining the 8 different motion directions at each of the 3x3 segments.

Each of the 8 motion directions is quantitatively labeled as an integer 1 through 8 corresponding to the directions of motions of 0°, 45°, 90°, 135°, 180°, 225°, 270°, and 315°, with the singularity (all directions combined to yield no net motion) being set to the value 0. This arrangement of the motion labeling enables us to process each recording of the data as an 1x11 vector: the first value is the mean firing rate as spikes per second, the 2nd through 10th values are the motion direction integers for segments 1 through 9, and the 11th value represents the firing rates' standard deviation derived from the recordings' multiple trials. This row vector represents the input stimulus as well as the resulting mean firing rate with its standard deviation in a convenient way, where a single neuron's response data is then a 16x11 matrix containing response data from all 16 optic flow visual stimulus recordings. The data for the single segment recordings follow the same convention.

## IV. PROPOSED METHOD

To model the neurophysiological data, we proposed the use of a generative model of mixture of Gaussians, trained on the optic flow response data. The Gaussian mixtures' parameters are optimized using a genetic algorithm, maintaining the link to biological plausibility.

### A. Dual-Gaussian Model

We propose using the mixture of two Gaussian shaped functions to model the directional selective firing rate data, it is chosen as a reasonable and widely employed approximation to single neuron response directionality functions [4]. Dual functions are used to accommodate the excitatory and inhibitory effects that are commonly observed from local motion mechanisms, and also from either two excitatory or two inhibitory mechanisms. Due to the local motion selectivity that is implied by the single segment recordings, we allow each segment to have its own distinct dual-Gaussian model: for each of the nine segments, we model the local motion selectivity by utilizing two Gaussian functions. For each Gaussian, disregarding the normalizing constant would yield the following form:

$$G(\mu, \sigma) = \exp\left(\frac{-\mu^2}{2\sigma^2}\right) \quad (1)$$

where  $\mu$  represents the direction of preferred motion, ranging from 0 to 359 degrees, and  $\sigma$  reflects the variance of the motion selectivity. We remove the normalizing constant because the y-axis of our coordinate system represents the magnitude of the firing rate data, and it can be any positive value (the Gaussians can be negative to represent inhibitory responses). Therefore, the dual-Gaussian model per segment's local motion selectivity would require two additional parameters  $c$  and  $p$  representing the gain constant that modulates the height of the Gaussian curve, and the polarity parameter as to indicate the excitatory or inhibitory of its associated Gaussian, respectively and shown as follows,

$$G(x, c, p, \mu, \sigma) = c \cdot p \cdot \exp\left(\frac{-(x - \mu)^2}{2\sigma^2}\right) \quad (2)$$

We denote  $G(x|\Omega)$  as the Gaussian with trained parameters  $\Omega$ , and  $\phi(x|\Omega)$  as the final and combined form of dual-Gaussian segmental model for generating segmental firing rate responses. Variable  $x$  is the direction of input visual motion stimulus, where  $x = 0, 45, 90, 135, 180, 225, 270, 315$ , and the trained parameter space  $\Omega = (\mathbf{C}, \mathbf{P}, \mathbf{M}, \mathbf{\Sigma})$ , where  $\mathbf{C} = \{c_i \in (0, \dots, 200); i = 1, 2\}$ ,  $\mathbf{P} = \{p_i \in (-1, 1); i = 1, 2\}$ ,  $\mathbf{M} = \{\mu_i \in (0, \dots, 359); i = 1, 2\}$ ,  $\mathbf{\Sigma} = \{\sigma_i \in (0, \dots, 90); i = 1, 2\}$ , with  $i$  corresponds to one of the two Gaussians that the parameters are associated with.

$$\phi(x|\Omega) = \begin{cases} \max(G_1(x|\Omega_1), G_2(x|\Omega_2)), & \text{if } p_1 = p_2 > 0 \\ \min(G_1(x|\Omega_1), G_2(x|\Omega_2)), & \text{if } p_1 = p_2 < 0 \\ \sum_{i=1}^2 G_i(x|\Omega_i), & \text{otherwise} \end{cases}$$

In the next section, the precise training of  $\Omega$ , using a genetic algorithm, is explained in detail. Each of the nine receptive field segments consists of an independent dual-Gaussian model  $\phi(x|\Omega)$ . The receptive field model is trained by placing the nine segmental dual-Gaussian models in accordance to the 3x3 segmental layout, and summing the nine segmental responses to generate the firing rate response given the full-field optic flow motion stimulus.

$$\hat{r} = b + \sum_{j=1}^9 \phi_j(x_j|\Omega_j) \quad (3)$$

We denote  $\hat{r}$  as an MST neuron's receptive field model's firing rate response given a full-field optic flow motion stimulus, and  $b$  is the neuron's baseline firing rate recorded when the monkey is fixating on the screen in the absence of a visual stimulus other than the fixation point. The summation of Eq. 3 combines the 9 segment local responses in an additive model, which follows the generally accepted convention that receptive fields are additive [22][23].

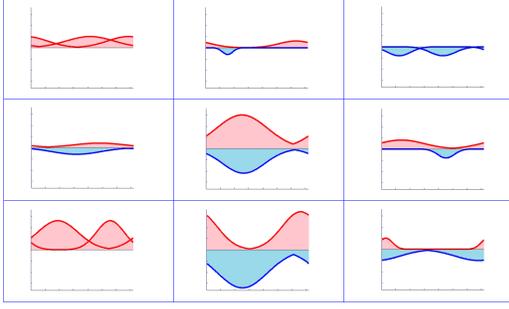


Fig. 5. A sample neuron's trained receptive field dual-Gaussian model, raw dual-Gaussian mixtures representation. The x-axis is the direction of motion selectivity and spans from 0 to 359 in degrees; the y-axis is the amplitude of the firing rate in spikes per second that spans from -18 to 18 in this case. Red implies positive Gaussian (excitatory responses) whereas blue represents negative Gaussian. The Gaussian curves wrap around both x directions that are consistent with the continuity of angles.

### B. Model Training using Genetic Algorithm

The nine pairs of segmental dual-Gaussian models are organized in the 3x3 array of receptive field segments, as shown in Fig. 5 and the corresponding arrow representation in Fig. 6. All parameters are tuned to generate firing rate responses for comparison to the recorded neuronal responses. We used a genetic algorithm as our model training method to optimize the parameters  $\Omega = (\mathbf{C}, \mathbf{P}, \mathbf{M}, \Sigma)$ , where the genetic algorithm is an optimization method that consists of several stages of processing, based on principles reflecting the course of evolution and survival of the fittest [1][12]. Although the genetic algorithm only produces an approximation to the search of optimal solutions and requires a number of iterations, it is an evolutionary computing method that is highly biologically plausible [9].

1) *Initialization*: To begin the process of finding the best fitting receptive field model, we randomly initialize the parameters  $\Omega$  of 1300 *individual* candidates, where each *individual* is the 9-segment model that contains a dual-Gaussian for each segment. Therefore, an individual model contains 18 Gaussian functions  $G(x, c, p, \mu, \sigma)$  defined from Eq. 2 with its 4 parameters  $c, p, \mu,$  and  $\sigma$  optimized, where  $x$  is the observation of recorded firing rate for a particular stimulus presentation trial. Each of the 4 parameters are randomly initialized according to their allowable range, where  $c = 0, \dots, 200, p = 1, \dots, 15$  with integer larger than 7 representing positive polarity (+1) and negative (-1) otherwise,  $\mu = 0, \dots, 359,$  and  $\sigma = 1, \dots, 90$  as defined in the previous section.

2) *Selection*: After 1300 first generation individuals are randomly initialized, we select the top performing individuals as the *parents* for crossover stage, based on their fitness score. Each individual is evaluated for its *fitness*, which represents how similarly the models' generated firing rates approximate those obtained from the recorded neuronal responses. We define two fitness criteria that represent how closely related a candidate model is to its biological counterpart:

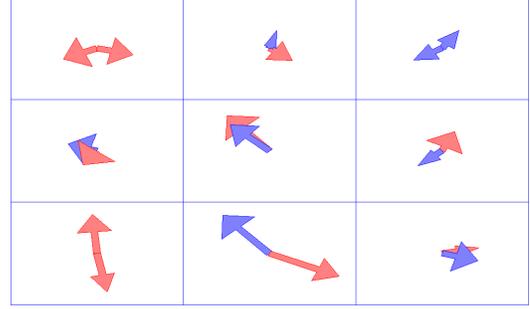


Fig. 6. A sample neuron's trained receptive field dual-Gaussian model, arrow representation. Each arrow represents a Gaussian, the orientation corresponds to its tuned motion selectivity angle ( $\mu$ ) in polar coordinate; red signifies positive Gaussian which implies excitatory response while blue represents negative (inhibitory) Gaussian. The length of the arrow is proportional to its magnitude ( $c$ ) and the width of the arrowhead denotes its motion selectivity variance ( $\sigma$ ).

$$d_e = \sum_{i=1}^{36} \frac{|\hat{r}_i - r_i|}{se_i} \quad (4)$$

$$d_g = \sum_{i=1}^{36} |kmc(\hat{r}_i, k) - kmc(r_i, k)| \quad (5)$$

We defined Eq. 4 as the *total error* fitness score over 36 distinct local motion stimuli as shown in Fig. 3, which is the distance between the model's response  $\hat{r}_i$  from the recorded firing rate response  $r_i$ , where  $se_i$  denotes the standard-error calculated from the multiple trials of biological recording given the  $i^{th}$  local visual motion stimulus. Eq. 5 is the *group error* fitness score, in which we denote  $kmc(\mathbf{x}, k)$  as the k-means clustering of a given set of observations  $\mathbf{x}$  and  $k$  clusters. In this case, we cluster the model responses  $\hat{r}_i$  as well as recorded firing rate observations  $r_i$  into groups of  $k = 3$  in a k-means cluster analysis where the product of that analysis represents inhibitory, no response, and excitatory effects by sorting the final centroids' amplitude.

Each individual's total error score  $d_e$  and its group error score  $d_g$  are computed to minimize the final  $d_e$  by utilizing both  $d_e$  and  $d_g$  error criteria in order to search for an optimally tuned model for each neuron. We then pick the best fitting 25 candidate models with the lowest  $d_e$  and the 25 with the lowest  $d_g$ .

3) *Crossover*: The crossover stage of the genetic algorithm combines the selected elite parents, in the hope of producing more viable offspring in the next generation by proliferating the better fitting genes from the parents. Crossover is done by merging the parents' genes for each of the Gaussian parameters:  $\mathbf{C}, \mathbf{P}, \mathbf{M},$  and  $\Sigma$ . As the parameters have been initialized within their allowable range, they were converted into its binary bit string representation where each bit is a chromosome in order for crossover to be possible.

From the 2 selected elite groups each containing 25 individuals, crossover is done by merging the genes from 2 parents,

where the first parent is chosen from the total error list while the second parent is from the k-means group error list. This type of hybrid crossover enables the training to include both traits as the model converges: best fitting as well as correct grouping, which allows the model to capture two plausible views of the systems implications of neurophysiological responses. When 2 parents are chosen, for each gene that is to be crossed over a random index  $z$  is selected, and it is within the size of the gene’s bit string. The index  $z$  serves as the cut-off point that splices the genes from both parents into 2 left and right subsets. By concatenating the parent 1’s left portion to parent 2’s right portion, and parent 2’s left portion to parent 1’s right portion we create 2 offspring genes, and repeating this step for all 4 parameters result in 2 individuals that are the product of crossing over a pair of parents. Since there are 25 elites from each of the 2 selected groups, the crossover stage yields  $25 * 25 * 2 = 1250$  offspring individuals.

4) *Mutation*: Genetic algorithms are a directed search method that attempts to locate the global minimum in terms of parameter errors. Mutation is applied by randomly altering the genes of some offspring individuals to overcome the algorithm’s potential resistance to escaping from any local minimum in the search space. Each offspring individual has a probability for being mutated, in this case we set our mutating probability as 0.05 meaning mutation happens about 5% of the time. When a mutation occurs, we execute a bit toggling operation, where the mutating individual toggles one bit for all 4 of its genes, at one randomly selected index per bit string. An alternative perspective might view this as a quadruple mutation within each affected gene, and hence a higher mutation rate. Mutation may increase or decrease an individual’s fitness, but the decrease may just be what it takes to bring the search out of the local minimum in order to continue the search for the actual global minimum.

5) *Next Generation*: After the previous phases have completed in sequence, the 50 elite parents are put back into the pool of their 1250 offspring individuals, making it a grand total of 1300 individuals that compose the population for the next generation. Each generation of the genetic algorithm operates through the selection, crossover, and mutation process until the individuals converge to a globally optimized fitness score. Various stop criteria have been used with genetic algorithms. We observed that the imposition of a limit of 75 generations served our modeling efforts as well as other approaches and provided a consistent limit across all models and neurons, thus we applied it exclusively.

## V. EXPERIMENT

Our goal was to train dual-Gaussian models on the single segment recordings, and test the capacity of those models to predict neuronal responses to the full-field optic flow stimuli. In doing so, we are evaluating the hypothesis that local planar motion responses sum to create the responses to full-field stimuli. We set up the dual-Gaussian model and let the genetic algorithm training take its course. A total of 10 randomly initialized and independent trials of genetic algorithm trainings

takes place for every single neuron neuron. The one model out of all 10 trials with the lowest total error  $d_e$  is picked as the final receptive field model when the training sessions finish. It is important to note that although we use a hybrid crossover method that utilizes two fitness measures to combine both best fitting as well as correct grouping criteria, in the end the model that yields the lowest  $d_e$  is selected as the final optimized model.

Our data was recorded using 36 distinct local motion stimuli, as shown in Fig. 3. Therefore, each neuron model’s total error  $d_e$  and group error  $d_g$  were derived from the 36 training stimuli, as defined from Eq. 4 and Eq. 5. We found that the training of the models converges at around 40<sup>th</sup> generation by both  $d_e$  and  $d_g$  fitness measures (Fig. 7), well within our stopping criteria of 75 generations. The figure also displays the clear superiority of the selected elites from the overall population.

The genetic algorithm training method that we use for our experiment is a hybrid-selection and crossover process, intended to keep the models successively approximating the neuronal data. To demonstrate that our hybrid-selection and crossover approach is viable, we set up a total of 4 different genetic algorithm variants that include conventional approaches: tournament selection using total error only, hybrid tournament selection (using both  $d_e$  and  $d_g$ ), top selection using total error only, and hybrid top selection.

Tournament selection is a type of genetic algorithm selection method, in which selection is done by randomly picking two pairs of individuals, and keeping the individual with the lower error from each pair as the parents to undergo crossover to create the next generation. We compared tournament selection to our top selection because of the popularity of tournament selection. We also compared the single criteria selection vs. our hybrid selection to determine whether one or the other approach conveyed a substantial advantage.

The comparison of selection/crossover methods keeps all parameters the same, that includes the test neurons, the number of randomly initialized individuals, parameter ranges, the number of individuals selected, generation count, and mutation rate. We randomly selected 10 neurons for this experiment, and 3 trials of each genetic algorithm variant is conducted over 75 generations. Fig 8 shows the result of the training session from the four selection/crossover methods that were tested, with the results averaged across tested neurons over all trials. It is apparent that top selection is the method of choice for these data, where top selection of  $d_e$  is able to converge to a net error level that is about 3 times lower than that achieved with the tournament selection method.

Hybrid selection yielded a more interesting result in which we see almost no difference between hybrid and total error only approaches with respect to  $d_e$ . However, hybrid selection is able to produce  $d_g$  as low as 0 early into the training session. Therefore, it is apparent that hybrid top selection and crossover yields the best results in this case by taking both  $d_e$  and  $d_g$  into account.

The resulting neuron models achieves a mean total error  $d_e$

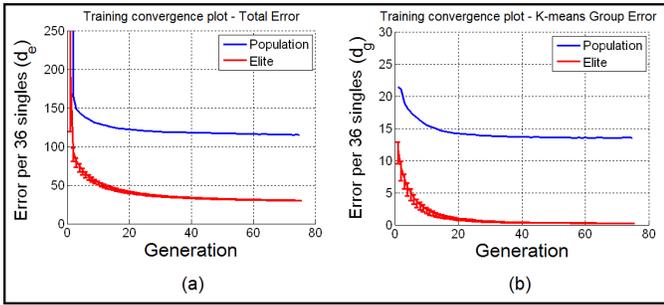


Fig. 7. Genetic algorithm training plot, averaged across all trials and neurons. The error bars from the elites curve represents the variance of a given generation. The plots display apparent convergence for the elites, as well as its variance.

of 33.95 sp/s (spikes per second) over each of the 36 local motion stimulus set, which computes to 0.94 sp/s for every local motion stimulus. This is a robust result, in the context of our mean firing rate of our recorded neurons is 12.26 sp/s, therefore our trained models are within 92% overall accuracy.

In Fig. 9, we show 4 sample neuron’s singles data training response profile, its full-field optic flow data prediction, and the resulting model visualized in arrow representation. The figure shows that training on the singles data is highly accurate, and that dual-Gaussian neuron models are can predict the full-field optic flow data, suggesting that the receptive fields of neurons are linearly additive. However, the 4th neuron in the figure tells a different story: it was not able to predict the optic flow responses. This raises the issue of segmental interaction effects, which is the focus of ongoing work.

With the total of 52 models that are trained from hybrid top selection and crossover, we may construct a self-motion detector based on the trained dual-Gaussian models by using Gaussian derivative models [15][16], our trained dual-Gaussian models, and a feed forward neural network.

To build a simple model for the dorsal stream that detects self-motion, we refer back to the specifications of Fig. 2, and implement the goal of detecting the global pattern of optic flow from an input video sequence, divide the global motion into 3x3 local motion fields, then use the trained dual-Gaussian models to differentiate which self-motion the video contains.

We maintain biological plausibility by using the approach that optic flow from a given input video sequence can be detected by utilizing the Gaussian derivative models introduced by Young et. al. (2001). Gaussian derivative models are 3D Gaussian filters that are differentiated and organized into specific orientations that mimic the receptive fields in V1 and V2 to detect motion. More details of the approach can be found in [15] and [16]. For our purposes, we situate 4 different Gaussian derivative filters at each pixel for detecting the 4 planar motions of left, right, up and down. We extract Harris corner features on each video frame, and convolve the Gaussian derivative filters with the frames from the resulting Harris corner responses.

The mean responses of the 4 Gaussian derivative filters

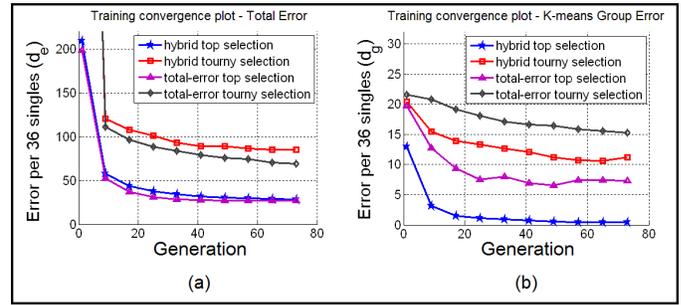


Fig. 8. Genetic algorithm variants training comparison plot, averaged across all trials and the tested neurons. It is apparent that although all four variants of genetic algorithm training sessions converge, the hybrid top selection setup achieves both the lowest total-error  $d_e$  and k-means group error  $d_g$ .

from each of the 3x3 local area of the video frames are computed and by subtracting the mean leftward responses from the mean rightward responses, and subtracting the mean downward responses from the mean upward responses of each 3x3 local segments of the resulting frames, we are able to obtain each local segment’s x and y components that describes the segmental motion.

We use the Gaussian derivative detected motion responses as the input to the trained dual-Gaussian models to get the firing rate responses. One way to make sense of the responses from the 52 neuron models is to train a neural network that learns how to classify self-motion directions based on the firing rate responses from all 52 neuron models.

We used a feed-forward neural network to train our models. It has a 52 neuron input layer that takes the responses of the 52 neuron models as inputs, 3 hidden layers that are 40, 30, and 20 neurons toward the 4 neuron output layer. The 4 neuron output layer is trained to output as a column vector of either [1 0 0 0], [0 1 0 0], [0 0 1 0], or [0 0 0 1] that represents the four self-motion directions tested: rightward global motion (leftward self-motion), leftward global motion (rightward self-motion), inward global motion (backward self-motion), and outward global motion (forward self-motion) respectively.

We tested our dorsal stream model on a virtual reality video sequence that was captured during a forward motion followed by a right turn, then a final brief forward motion. We choose to use a virtual reality motion sequence because the VR environment eliminates the shaky movements that complicate the acquisition of video signals. This simulates the stabilizing effects that are provided by oculomotor reflexes in biological systems. Fig. 10 shows an example of the results from the processing of images from the virtual reality walk through.

This preliminary examination of self-motion detection yielded an overall accuracy rate of 60% correct discriminations between the four directions of self-motion presented, in which random guesses would be correct 25% of the time. We conclude that our models may provide the basis for automated self-motion detection.

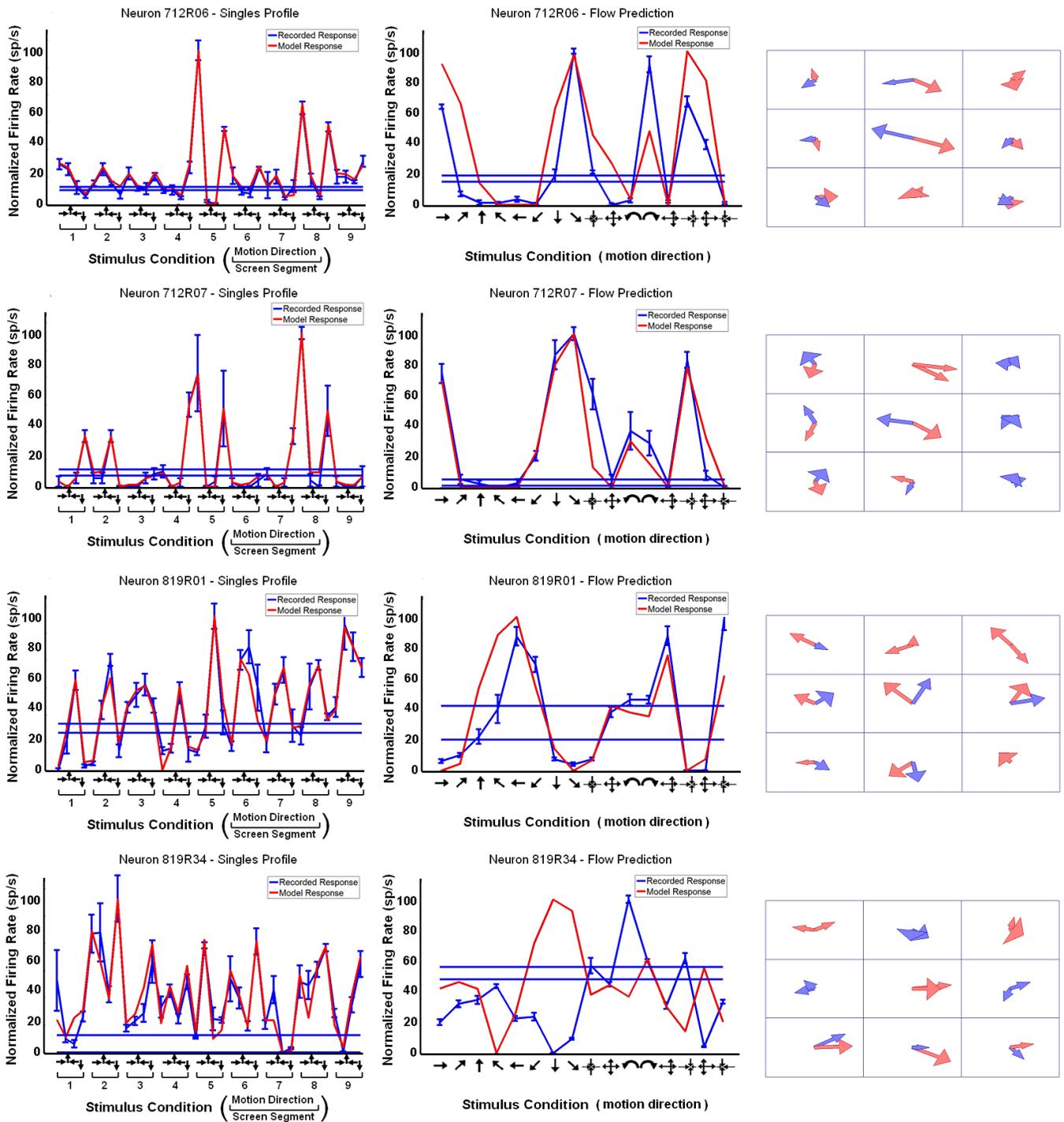


Fig. 9. Sample results for 4 neurons. The first column is the singles data training response profile with the 2 horizontal lines from each profile representing the range of observed baseline firing rates, and the error bars represent the standard error from the multiple recording trials. The center column is each neuron's full-field optic flow data prediction using the trained model, and the right column is the neuron model visualization in arrow representation. The training profiles indicate high training accuracy, but only the first 3 neurons are able to predict their optic flow data responses, suggesting that there is more than linearly additivity for MST neuron receptive fields. (All neuron responses are normalized to the range of [0 100])

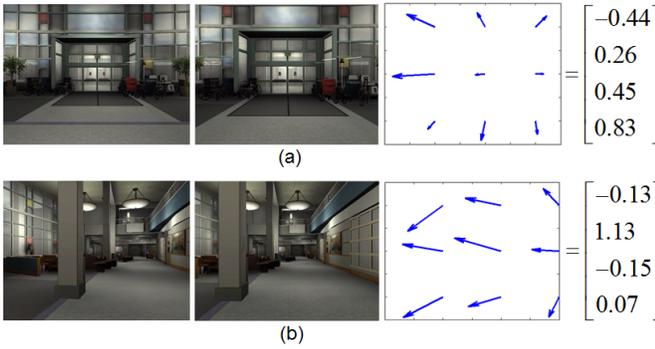


Fig. 10. Sample test video sequences within the lobby of Strong Memorial Hospital virtual reality environment. Sequence (a) shows a forward self-motion followed by the detected global flow field by the Gaussian derivative model, and the 52 dual-Gaussian neural network classifies the motion as forward self-motion ( $[0\ 0\ 0\ 1]$  after rounding to the nearest integer). Sequence (b) shows a right turn self-motion, with the corresponding flow field and the decision vector rounded into  $[0\ 1\ 0\ 0]$  that is also correctly classified as a rightward self-motion.

## VI. CONCLUSION

We have developed a mixture of Gaussians method to model MST neuron receptive fields, and tested the resulting models on full-field optic flow data. Our hybrid selection for the training of the genetic algorithm optimized the parameters of our dual-Gaussian models in 75 generations. We conducted this experiment as 10 independent trials that all yielded converging results. The models' optic flow predictions revealed that some models are able to predict optic flow responses whereas others could not. This may imply interaction between screen segments, where parts of a neuron's receptive field interact with each other non-linearly to produce an optimal full-field optic flow response.

Finally, we applied the models to computer vision as a biologically inspired self-motion detection system that represents a simple pipeline in dorsal stream processing. Although we would like the accuracy to be higher, it is the first real neuron model that is applied to computer vision motion detection to the best of our knowledge, therefore we are excited about the results from our experiments.

We are currently working on exploring more of the possibilities of segmental interactions that are suggested from our experiments, and we are also analyzing the design of data collection, in which we hope to incorporate more diversity into our data while reducing the recording time. Future work includes finding the potential segmental interaction effects by designing a reduced dual-Gaussian model, and construct a more robust self-motion detector using interaction effects.

## VII. ACKNOWLEDGMENTS

We gratefully acknowledge Dr. David J. Logan's role in the design and execution of the single neuron recording experiments and Ms. Sherry Estes' assistance in monkey care and surgery. This work was supported by National Eye Institute grants R01-EY10287 and P30EY01319.

## REFERENCES

- [1] M. Battarra, B. Golden and D. Vigo, "Tuning a parametric Clarke-Wright heuristic via a genetic algorithm." *Journal of the Operational Research Society*, Vol. 59, No. 11, pp. 1568-1572, 2008.
- [2] S. Stringer, E. Rolls and J. Tromans, "Invariant object recognition with trace learning and multiple stimuli present during training." *Network: Computation in Neural Systems*, Vol. 18, Iss. 2, pp. 161-187, 2007.
- [3] K. Deep and M. Thakur, "A new crossover operator for real coded genetic algorithms." *Applied Mathematics and Computation*, Vol 188(1), pp. 895-911, 2007.
- [4] J. Kulkarni and L. Paninski, "Common-input models for multiple neural spike-train data." *Network: Computation in Neural Systems*, Vol. 18(4), pp. 375-407, 2007.
- [5] N. Rust, V. Mante, E. Simoncelli and J. Movshon, "How MT cells analyze the motion of visual patterns." *Nature Neuroscience*, Vol. 9, No. 11, pp. 1421-1431, 2006.
- [6] P. Achard and E. De Schutter, "Complex Parameter Landscape for a Complex Neuron Model." *PLoS Computational Biology*, 2006.
- [7] K. Grill-Spector, R. Henson and A. Martin, "Repetition and the brain: neural models of stimulus-specific effects." *Trends in Cognitive Sciences*, Vol. 10(1), pp. 14-23, 2006
- [8] E. Andrade, S. Blunsden and R. Fisher, "Hidden Markov Models for Optical Flow Analysis in Crowds." *International Conference on Pattern Recognition*, Vol. 1, pp. 460-463, 2006.
- [9] W. Chang, "An improved real-coded genetic algorithm for parameters estimation of nonlinear systems." *Mechanical Systems and Signal Processing*, Vol. 20(1), pp. 236-246, 2006.
- [10] T. Serre and T. Poggio, "Object Recognition with Features Inspired by Visual Cortex." *Massachusetts Institute of Technology*, 2005.
- [11] N. Rust, O. Scharz, J. Movshon and E. Simoncelli, "Spatiotemporal Elements of Macaque V1 Receptive Fields." *Neuron*, Vol. 46(6), pp. 945-956, 2005.
- [12] F. Friedrichs and C. Igel, "Evolutionary tuning of multiple SVM parameters." *Neurocomputing*, Vol. 64, pp. 107-117, 2005.
- [13] S. Baker and I. Matthews, "Lucas-Kanade 20 Years On: A Unifying Framework." *International Journal of Computer Vision*, Vol. 56(3), pp. 221-255, 2004.
- [14] J. Verbeek, N. Vlassis and B. Krose, "Efficient Greedy Learning of Gaussian Mixture Models." *Neural Computation*, Vol. 15(2), pp. 469-485, 2003.
- [15] R. Young, R. Lesperance and W. Meyer, "The Gaussian Derivative model for spatial-temporal vision: I. Cortical model" *Spatial Vision*, Vol. 14, No. 3,4, pp. 261-319, 2001.
- [16] R. Young, R. Lesperance and W. Meyer, "The Gaussian Derivative model for spatial-temporal vision: II. Cortical data" *Spatial Vision*, Vol. 14, No. 3,4, pp. 321-389, 2001.
- [17] M. Lappe, "A model of the combination of optic flow and extraretinal eye movement signals in primate extrastriate visual cortex: Neuronal model of self-motion from optic flow and extraretinal cues." *Neural Networks*, Vol. 11, pp. 397-414, 1998.
- [18] D. Hubel and T. Wiesel, "Eye, Brain and Vision." *Scientific American Library*, No. 22, 1995.
- [19] C. Duffy and R. Wurtz, "Sensitivity of MST Neurons to Optic Flow Stimuli. I. A Continuum of Response Selectivity to Large-Field Stimuli." *Journal of Neurophysiology*, Vol. 65, No. 6, pp. 1329-1345, 1991.
- [20] C. Duffy and R. Wurtz, "Sensitivity of MST Neurons to Optic Flow Stimuli. II. Mechanisms of Response Selectivity Revealed by Small-Field Stimuli." *Journal of Neurophysiology*, Vol. 65, No. 6, pp. 1346-1359, 1991.
- [21] E. Adelson and J. Bergen, "Spatiotemporal energy models for the perception of motion." *Journal of Optical Society of America*, Vol. 2, No. 2, pp. 284-299, 1985.
- [22] J.D. Victor and R.M. Shapley, "Receptive field mechanisms of cat X and Y retinal ganglion cells." *The Journal of General Physiology*, Vol. 74, pp. 275-298, 1979.
- [23] C. Enroth-Cugell and P. Lennie, "The control of retinal ganglion cell discharge by receptive field surrounds." *The Journal of Physiology*, Vol. 247, pp. 551-578, 1975.