Robustness and Real-Time Performance of an Insect Inspired Target Tracking Algorithm Under Natural Conditions

Zahra M. Bagheri[1,2], Steven D. Wiederman[1], Benjamin S. Cazzolato[2], Steven Grainger[2] and David C. O'Carroll[1,3]
[1] School of Medicine, [2] School of Mechanical Engineering, The University of Adelaide, Adelaide, Australia
[3]Department of Biology, Lund University, Lund, Sweden

Abstract— **Many computer vision tasks require the implementation of robust and efficient target tracking algorithms. Furthermore, in robotic applications these algorithms must perform whilst on a moving platform (egomotion). Despite the increase in computational processing power, many engineering algorithms are still challenged by real-time applications. In contrast, lightweight and low-power flying insects, such as dragonflies, can readily chase prey and mates within cluttered natural environments, deftly selecting their target amidst distractors (swarms). In our laboratory, we record from 'target-detecting' neurons in the dragonfly brain that underlie this pursuit behavior. We recently developed a closed-loop target detection and tracking algorithm based on key properties of these neurons. Here we test our insect-inspired tracking model in open-loop against a set of naturalistic sequences and compare its efficacy and efficiency with other state-of-the-art engineering models. In terms of tracking robustness, our model performs similarly to many of these trackers, yet is at least 3 times more efficient in terms of processing speed.**

I. Introduction

Many target tracking algorithms have been developed over the last decade for a diverse range of applications, e.g. surveillance, human assistance robots, wildlife monitoring and smart cars. An ideal visual tracker accounts for different problems such as illumination changes, rapid changes in target appearance, non-smooth target trajectories, occlusion and background clutter. Many engineering methods developed for target tracking simplify these scenarios, ensuring a more tractable tracking problem. Moreover, these methods involve complex computation (e.g. particle filters), that require large, high-powered processors. Consequently, these solutions are often impractical in real-time applications, particularly where probability is desirable. These issues highlight the need for an alternative and more efficient approach to solving at least a subset of the target tracking problems.

Studies of insect visual systems suggest there is a solution contained within a 'simple' neuronal architecture (~ 1 million neurons). For example, the dragonfly is a remarkable aerial predator which detects, selects and then chases prey or mates within a visually cluttered surround even in the presence of other distracting stimuli, such as swarms of prey and conspecifics [1], [2]. The dragonfly performs this task despite its light-weight and low-power brain and its low-resolution visual system (acuity of ~1°). The neuronal algorithms behind such a robust and efficient target tracking behaviour (the envy of engineers) is currently being elucidated by our lab and other neuroscientists in the field.

Our approach to engineering a solution to this target tracking problem is to model the neuronal pathway that underlies the dragonfly pursuit behaviour. We record from 'small target motion detector' (STMD) neurons of the insect lobula in response to different visual stimuli. These neurons are size selective, velocity tuned, contrast sensitive, and respond robustly to small moving targets even in the presence of background motion [3-6]. Inspired by such electrophysiological recordings from STMD neurons, we previously developed an algorithm for local target discrimination [7]. This 'elementary' small target motion detector (ESTMD) model provides nonlinear spatiotemporal matched filtering for small moving targets embedded in natural scenery [7]. Recently, we elaborated this model to include the recent observations of response 'facilitation' [8,9] (a slow build-up of response to targets that move on long, continuous trajectories) [10-14]. We implemented this elaborated model in a closed-loop target tracking system that uses an active saccadic gaze fixation strategy inspired by insect pursuit behaviour [10-14]. Using this closed-loop model we showed that facilitation improves the robustness of pursuit [14]. We also investigated the effect of different environmental variables (background clutter, target contrast, target velocity) and model parameters (spatial and temporal components of facilitation) on pursuit success. Our model predicted an optimal, dynamic behaviour for a temporal component of facilitation that was dependent on background clutter [14].

Although our model showed robust performance in a constrained virtual-reality environment, natural conditions such as illumination changes, local flicker and target occlusion could affect model behaviour. In this paper we test the robustness of our model in open-loop, using videos recorded from natural scenes [15]. This allows us to compare the processing speed and tracking performance of our insect-inspired model with several state-of-the-art engineering algorithms.

978-1-4799-7560-0/15 $31.00 © 2015 IEEE
DOI 10.1109/SSCI.2015.24

Figure 1 A single frame 'snapshot' of the videos [15] used to test both the performance of our insect-inspired model as well as other previously published tracking models [23-28].

II. Methods

A. Dataset

We used 15 different image sequences downloaded from a publicly available dataset [15]. These sequences had different lengths ranging from 80 to 3000 frames (with an average of 558 frames). Fig. 1 shows a snapshot of these videos at the midpoint of each sequence. All of these videos included camera motion.

B. Insect-Inspired Target Tracking Model

Fig. 2 shows an overview of the insect inspired target tracking model implemented in MATLAB. The optics of insect compound eyes are limited by diffraction of the facet lenses [16]. We modelled this optical blur with a Gaussian function of full-width at half maximum of 1.4° [16]. We selected only the green channel of the RGB input to simulate the sensitivity of typical insect motion sensitive pathways to green light [17]. Further sub-sampling was applied to the blurred image to model the average inter-receptor angle between photoreceptors [18]. In biological systems, early visual processing by the photoreceptors themselves and 1st order interneurons remove redundant information in space and time, using neuronal adaptation and center-surround antagonism. These properties of visual system were simulated with spatiotemporal bandpass filtering matched to properties observed in insect vision [19].

The ESTMD subsystem starts with modelling the response properties of rectifying, transient cells as observed in several insect species [20, 21] by separating the ON and OFF contrasts via temporal high pass filtering (τ=40 ms) and half-wave rectification [7]. These independent ON and OFF channels were further processed through a fast adaptive mechanism. The state of adaptation was determined by a nonlinear filter which switches its time constant [7, 11]. Time constants are 'fast' (τ=3 ms) when channel input is increasing and 'slow' (τ=70 ms) when decreasing. This adaptation state causes subtractive inhibition of the unaltered 'pass-through' signal. Additionally, we implemented strong spatial centre-surround antagonism, with each channel surround inhibiting its next-nearest neighbours. This strong surround antagonism conveys selectivity for local edge features. Sensitivity to both dark and light targets was provided by multiplying each contrast channel (ON or OFF) with a delayed version of the opposite polarity (via a low-pass filter, τ=25 ms) and then summing the outputs [12-14]. The neuron-like soft saturation of each resulting ESTMD was modeled with a non-linear saturation using a hyperbolic tangent function. This serves to ensure all signals lie between 0 and 1. A simple form of the competitive selection observed in dragonflies [2] was modeled by choosing the maximum ESTMD output as the target location.

The slow build-up of facilitation as observed in several dragonfly STMD neurons [8,9] permits the extraction of the target signal from noisy (cluttered) environments. This facilitation mechanism was modeled by building a weighted 'map' dependent on the location of the winning feature but shifted by the target velocity vector [14]. The directional component of this velocity vector was provided using a traditional bio-inspired direction selective model; the Hassenstein-Reichardt elementary motion detector (HR-EMD) [23]. The HR-EMD uses two spatially separated contrast signals and correlates them after a delay (via a low-pass filter). Additionally, the output of the HR-EMD was segmented into three equal intervals to estimate the range of the spatial component of the target velocity [14]. We multiplied the ESTMD model output with a low-pass filtered version of this 'facilitation map' (Fig. 2). The time constant of this filter controls the duration of the enhancement around the predicted location of the winning feature.

C. Benchmarking Algorithms

To establish the computational efficiency of our insect-inspired tracker (IIT) model, we compared its performance with six recent highly-cited algorithms for which code is publicly available. For a fair comparison with respect to processing speed we chose MATLAB implementations of these algorithms.

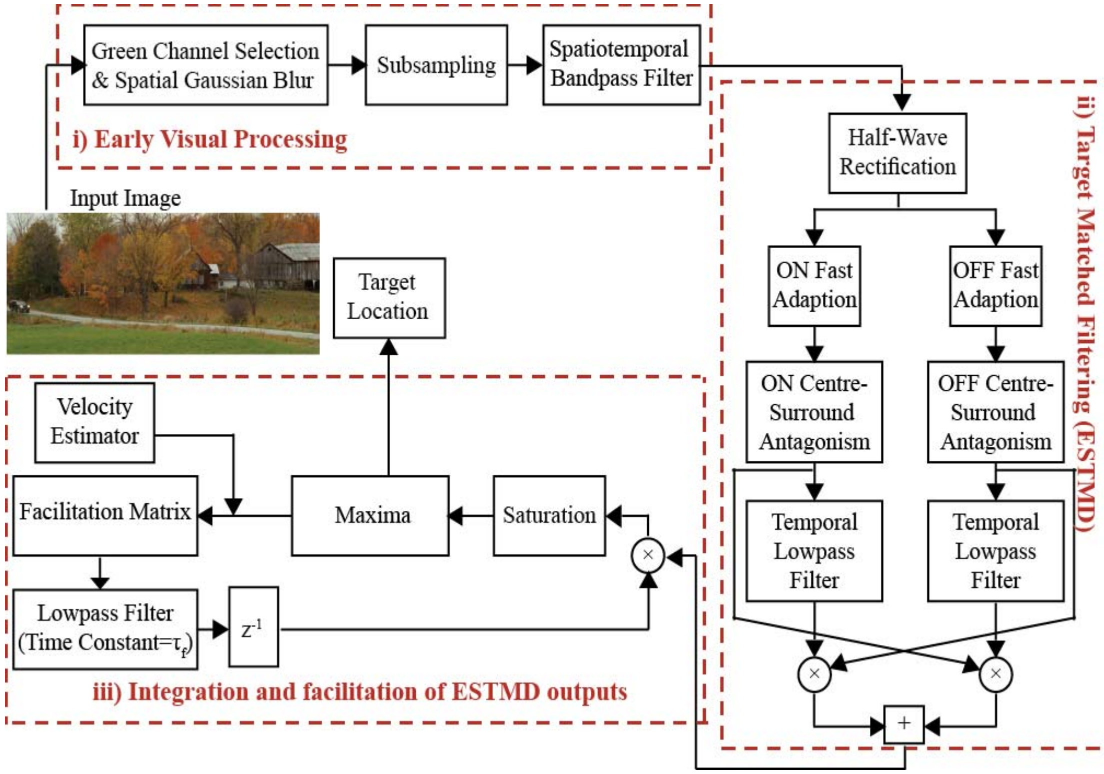

Figure 2 The overview of the insect-inspired target tracking algorithm. This model is composed of three main subsystems: i) early visual processing, ii) target matched filtering (ESTMD) and iii) integration and facilitation of ESTMD outputs.

1- **Incremental visual tracker (IVT)** [24] proposes an adaptive appearance model which stores the latest eigenvectors of the target image and deletes the old observations.
2- **L1-minimization Tracker (L1T)** [25] employs sparse representation by L1 to provide an occlusion insensitive method. This method ignores the target image samples with small probabilities to reduce the cost of computation associated with L1 minimization.
3- **Locally Orderless Tracker (LOT)** [26] proposes a joint spatial-appearance adaptive mechanism to calculate the extent of local disorder in the target. This allows the algorithm to track both rigid and non-rigid targets.
4- **Super Pixel Tracker (SPT)** [27] embeds a discriminative classifier in super pixel (group of pixels which have similar characteristics) clustering to handle changes in scale, motion and occlusion.
5- **Tracking, Learning and Detection (TLD)** [28] is ranked as one of the most resilient available trackers. It combines a discriminative learning method with a detector and an optical flow tracker.
6- **Compressive Tracking (CT)** [29] proposes an appearance model based on features extracted in the compressed domain.

All models were tested in Matlab (R2012b) on the same PC with an Intel i7 3770 CPU (3.4 GHz) and 16 GB RAM. The location of a target bounding box in the initial frame was provided for the benchmark algorithms. Likewise, in the initial frame, we biased our IIT model toward the location of the target by allowing the facilitation to build up in the target region for 40 ms prior to the start of the experiment.

III. Results

Comparing the robustness of different target tracking algorithms is a challenging task since different metrics could be analysed (e.g. scale, shape representation). Here we limit our measure of tracking robustness to correctly locating the target position in each frame. We used different metrics to compare the robustness of the algorithms as well as the processing speed of the trackers.

A. Success Plot

The engineering algorithms represent the target with a bounding box. Therefore, we scored each frame as a successful detection of the target if the center of the bounding box was within the ground truth box. Similarly, for our IIT algorithm, if the location of the winning feature was within the ground truth box it was considered a successful detection of the target.

Fig. 3 shows the box-and-whiskers plots summarizing the success of all 7 trackers for the 15 different test sequences. On each box, the central mark is the median success rate, the edges of the box are the 25th and 75th percentiles, and the whiskers extend to the most extreme data points that are not considered outliers (n=15). The IVT algorithm has the highest median which shows it was capable of correctly locating the target in all frames in half of the sequences. However, the 25 percentile and lowest value are at 30% and 5% respectively, indicating a lack of flexibility of this model under certain circumstances.

Among all algorithms, TLD performs more reliably under different conditions (i.e. it has the highest 25th percentile). Unlike our 'simple' feed-forward computations, these trackers contain several complexities (as described in the method section). Despite this difference, the median of our algorithm (IIT) indicates a performance on par with these other algorithms. Additionally, our model has the lowest inter-quartile range (distance between the 25th and 75th percentiles) showing that our model can perform as robustly as the state-of-the-art engineering algorithms under different natural conditions.

B. Precision Plot

The Precision plot is an evaluation method recently adopted to measure the robustness of tracking [15, 30, 31]. It shows the percentage of the frames where the Euclidean distance between the center of the tracked target and the ground truth is within a given 'location error' threshold. Fig. 4 shows the precision plot for all trackers. A higher precision at low thresholds means the tracker is more accurate.

Fig. 4 shows that our algorithm (IIT) has the best precision at the threshold of zero. Between thresholds of 0 and 10 pixels its precision increases rapidly, however is still below the ultimate precision of TLD, L1T and IVT. The main reason behind this behavior is likely the size selectivity of our model; i.e., it is tuned to small sized objects. Large objects are composed of small parts allowing our model to lock on to these sub-features of the larger object. The result is effective target tracking, but with the location offset from the center of the object. Our model's precision increases, catching up to those of other algorithms in the threshold range of 10 and 20 pixels. By a location error threshold of 20, our IIT exceeds the precision of all trackers except TLD. The precisions at the 20 pixel threshold widely used as a performance benchmark in the computer vision literature [15], [30], [31] are given as the representative precision score in Table I.

C. Overall Performance

Table I provides a descriptive summary of performance averaged across all 15 videos. In addition to the average success rate of the 15 sequences, we also calculated the weighted success which shows the percentage of the successful frames out of all the 8374 tested frames. This normalization accounted for the difficulty of 'long term' tracking, where it is easier for the trackers to lock on to the target in a short sequence than a long one.

Table I shows that the average success for our model is below that of TLD and IVT and close to LOT and L1T. However, when it comes to weighted success, our model takes second place, indicating very good long term tracking performance. Our facilitation mechanism (based on the recently observed facilitatory behavior of target-detecting neurons [8,9]) builds up slowly in response to targets that move

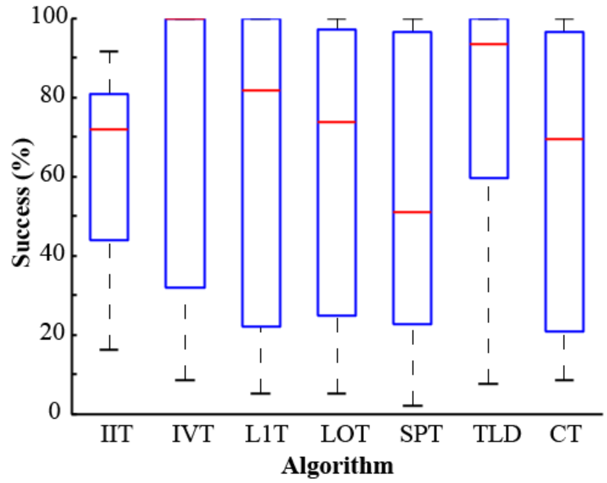

Figure 3 Box and whiskers plot for successful target tracking of different algorithm for all 15 different image sequences.

in long continuous trajectories, thus improves target detection as tracking progresses.

D. Processing Speed

Although comparable in tracking performance, our model excels in processing efficiency, a critical concern in target tracking applications. Indeed, many trackers are considered impractical in real-time scenarios due to their long processing duration. Fig. 5 shows the processing speed of the tested algorithms, with the IIT exceeding all other trackers (note the logarithmic scale). Our model performs approximately 12 times faster than IVT and TLD and 3 times faster than CT.

IV. Conclusion

We have demonstrated the robustness and efficiency of a target tracking algorithm inspired directly by insect neurophysiology. Our data clearly shows that this model can perform robustly under natural conditions. Despite the relatively simple mechanism we implemented, the robustness

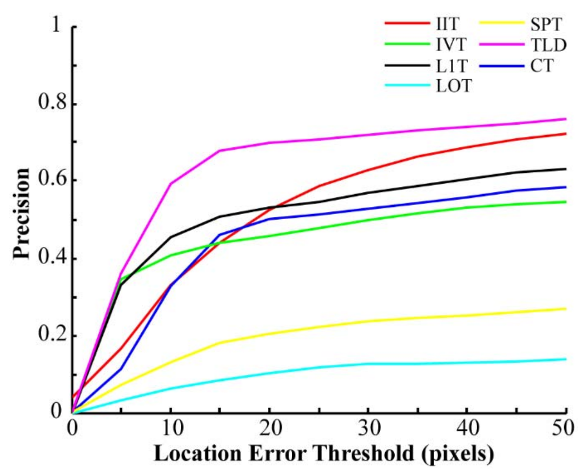

Figure 4 Precision plot for all 15 sequences.

TABLE I. SUMMARY OF EXPERIMENAL RESULTS ON THE 15 VIDEO DATASET

Performance Measure	Algorithm						
	IIT	*IVT*	*L1T*	*LOT*	*SPT*	*TLD*	*CT*
Average Success (%)	62.7	74.0	63.8	62.2	55.6	74.5	56.6
Weighted Success (%)	73.0	62.3	57.6	34.4	24.6	86.9	48.7
Precision (20 px)	0.53	0.46	0.53	0.01	0.21	0.70	0.50

of our model can compete with the state-of-the-art engineering trackers. A limitation of our model is that it was primarily designed to detect and track small moving targets. Therefore, it only tracks larger objects composed of smaller moving parts (within the size tuning range of our model). This limits its overall performance robustness compared with the best of the engineered trackers (such as TLD). Nevertheless, in terms of processing speed, our model outperforms all of the engineering trackers, mimicking the remarkable efficiency of the insect visual system upon which it is based. As such, it may be well suited to applications where efficiency is paramount.

Here, we tested our algorithm in open-loop, however, active vision may be a key to exploiting visual information by the simple insect brain for complex tasks such as target tracking. Future research will attempt to implement this model along with insect active vision system in a robotic platform to test the performance of them together under real-world conditions.

ACKNOWLEDGMENT

This research was supported under Australian Research Council's Discovery Projects (project number DP130104572) and Discovery Early Career Researcher Award (project number DE15010054) funding scheme. We thank the manager of the Botanic Gardens in Adelaide for allowing insect collection and video capture.

REFERENCES

[1] P. S. Corbet, Dragonflies: Behaviour and ecology of Odonata. Cornell University Press, 1999.

[2] S. D. Wiederman and D. C. O'Carroll, Selective attention in an insect visual neuron. Curr. Biol., vol. 23, pp. 156-161, 2013.

[3] D. C. O'Carroll, Feature-detecting neurons in dragonflies. Nature, vol. 362, pp. 541-543, 1993.

[4] K. Nordström, P. D. Barnett and D. C. O'Carroll (2006) Insect detection of small targets moving in visual clutter. PLoS. Biol., vol. 4, pp. 378-386, 2006.

[5] K. Nordström and D. C. O'Carroll DC, Feature detection and hypercomplex property in insects. Trends Neurosci., vol. 32, pp. 383-391, 2009.

[6] D. C. O'Carroll and S.D. Wiederman, Contrast sensitivity and the detection of moving patterns and features. Philos. Trans. R. Soc. B-Biol. Sci., vol. 369, pp. 20130043, 2014.

[7] S. D. Wiederman, P. A. Shoemaker, and D. C. O'Carroll, A model for the detection of moving targets in visual clutter inspired by insect physiology. PLoS ONE, vol. 3, pp. 1-11, 2008.

[8] J. R. Dunbier, S. D. Wiederman, P. A. Shoemaker and D. C. O'Carroll, Modelling the temporal response properties of an insect small target motion detector. IEEE international conference on Intelligent Sensors, Sensor Networks and Information Processing, pp. 125-130, 2011.

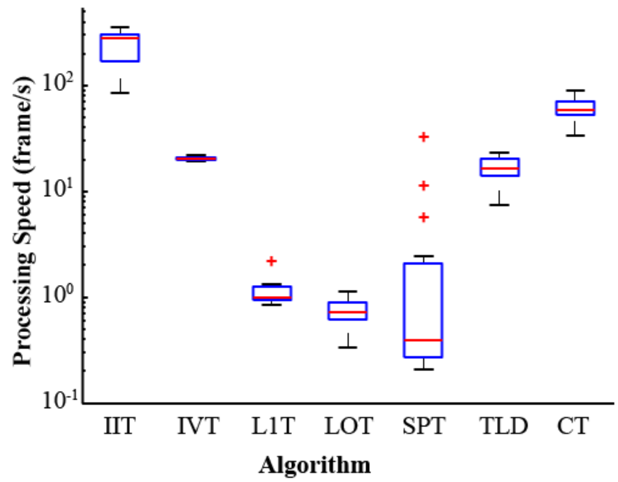

Figure 5 Processing speed of trackers.

[9] J. R. Dunbier, S. D. Wiederman, P. A. Shoemaker and D.C. O'Carroll, Facilitation of dragonfly target-detecting neurons by slow moving features on continuous paths. Front. Neural Circuits, vol. 6, 79, 2012.

[10] K. J. Halupka, S. D. Wiederman, B. S. Cazzolato and D. C. O'Carroll, Discrete implementation of biologically inspired image processing for target detection. IEEE International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP), pp. 143-148, 2011.

[11] K. J. Halupka, S. D. Wiederman, B. S. Cazzolato BS and D. C. O'Carroll, Bio-inspired feature extraction and enhancement of targets moving against visual clutter during closed loop pursuit. IEEE International Conference on Image Processing (ICIP), pp. 4098-4102, 2013.

[12] Z. M. Bagheri, S. D. Wiederman, B. S. Cazzolato, S. Grainger, and D. C. O'Carroll, A biologically inspired facilitation mechanism enhances the detection and pursuit of targets of varying contrast. International Conference on Digital lmage Computing: Techniques and Applications (DICTA) , pp. 1-5, Sep 2014.

[13] Z. Bagheri, S. D. Wiederman, B. S. Cazzolato, S. Grainger, and D. C. O'Carroll, Performance assessment of an insect-inspired target tracking model in background clutter. International Conference on Control Automation Robotics & Vision (ICARCV), pp. 822-826, Dec 2014.

[14] Z. M. Bagheri, S. D. Wiederman, B. S. Cazzolato, S. Grainger and O'Carroll DC, Properties of neuronal facilitation that improve target tracking in natural pursuit simulations. J. R. Soc. Interface vol. 12, 20150083, 2015.

[15] Y. Wu, J. Lim and M. H. Yang, M, Online object tracking: A benchmark. IEEE International Conference on Computer vision and pattern recognition (CVPR), pp. 2411-2418, 2013.

[16] D. G. Stavenga, Angular and spectral sensitivity of fly photoreceptors. I. Integrated facet lens and Rhabdomere optics. J. Comp. Physiol. A, vol. 189, pp. 1-17, 2003.

[17] M. V. Srinivasan and R. G. Guy, Spectral properties of movement perception in the dronefly Eristalis. J. Comp. Physiol. A, vol. 166, pp. 287-295, 1990.

[18] A. D. Straw, E. J. Warrant and D. C. O'Carroll DC (2006) A bright zone in male hoverfly (Eristalis tenax) eyes and associated faster motion detection and increased contrast sensitivity. J. Exp. Biol., vol. 209, pp. 4339-4354, 2006.

[19] S. D. Wiederman, R. S. A. Brinkworth and D. C. O'Carroll, Bio-inspired small target discrimination in high dynamic range natural scenes. International Conference on Bio-Inspired Computing: Theories and Applications, pp. 109-116, 2008.

[20] D. Osorio, Mechanisms of early visual processing in the medulla of the locust optic lobe-how self-inhibition, spatial-pooling, and signal rectification contribute to the properties of transient cells. Visual Neurosci., vol. 7, pp. 345-355, 1991.

[21] N. Jansonius and J. van Hateren, Fast temporal adaptation of on-off units in the first optic chiasm of the blowfly. J. Comp. Physiol. A, vol. 168, pp. 631–637, 1991.

[22] S. D. Wiederman and D. C. O'Carroll, Correlation between OFF and ON channels underlies dark target selectivity in an insect visual system. The Journal of Neuroscience, vol 33, pp. 13225-13232, 2013.

[23] B. Hassenstein and W. Reichardt, Analyse der zeit-, reihenfolgen-und vorzeichenauswertung bei der bewegungsperzeption des rüsselkäfers Chlorophanus. Z. Naturforsch, vol. 11b, pp. 513–524, 1956.

[24] D. A. Ross, J. Lim, R. S. Lin and M. H. Yang, Incremental learning for robust visual tracking. Int. J. Comput. Vis., vol. 77, pp. 125-141, 2008.

[25] X. Mei, H. Ling, Y. Wu, E. Blasch and L. Bai, Minimum error bounded efficient $\ell1$ tracker with occlusion detection. IEEE International Conference on Computer vision and pattern recognition (CVPR), pp. 1257-1264, 2011.

[26] S. Oron, A. Bar-Hillel, D. Levi and S. Avidan, Locally orderless tracking. IEEE International Conference on Computer vision and pattern recognition (CVPR), pp. 1940-1947, 2012.

[27] S. Wang, H. Lu, F. Yang, M. H. Yang, Superpixel tracking. IEEE International Conference on Computer Vision (ICCV), pp. 1323-1330, 2011.

[28] Z. Kalal., K. Mikolajczyk and J. Matas, Tracking-learning-detection. IEEE Trans. Pattern Anal. Mach. Intell.,vol. 34, pp. 1409-1422.

[29] K. Zhang, L. Zhang and M. H. Yang. Real-time compressive tracking. Springer Computer Vision–ECCV, pp. 864-877, 2012.

[30] B. Babenko, M. H. Yang and S. Belongie, Robust object tracking with online multiple instance learning. IEEE Trans. Pattern Anal. Mach. Intell, vol. 33, pp. 1619-1632, 2011.

[31] J. F. Henriques, R. Caseiro, P. Martins and J. Batista, Exploiting the circulant structure of tracking-by-detection with kernels. In Springer Computer Vision–ECCV, pp. 702-715, 2012.