Bio-inspired speed detection and discrimination

Mauricio Cerda[1], Lucas Terissi[2], and Bernard Girau[1]

[1] Loria - INRIA Nancy Grand Est, Cortex Team
Vandoeuvre-lès-Nancy - France,
[cerdavim,girau]@loria.fr,
[2] Laboratory for System Dynamics and Signal Processing,
Universidad Nacional de Rosario - CIFASIS - CONICET - Argentina,
terissi@cifasis-conicet.gov.ar

Abstract. In the field of computer vision, a crucial task is the detection of motion (also called optical flow extraction). This operation allows analysis such as 3D reconstruction, feature tracking, time-to-collision and novelty detection among others. Most of the optical flow extraction techniques work within a finite range of speeds. Usually, the range of detection is extended towards higher speeds by combining some multi-scale information in a serial architecture. This serial multi-scale approach suffers from the problem of error propagation related to the number of scales used in the algorithm. On the other hand, biological experiments show that human motion perception seems to follow a parallel multi-scale scheme. In this work we present a bio-inspired parallel architecture to perform detection of motion, providing a wide range of operation and avoiding error propagation associated with the serial architecture. To test our algorithm, we perform relative error comparisons between both classical and proposed techniques, showing that the parallel architecture is able to achieve motion detection with results similar to the serial approach.

Key words: motion perception, optical flow, speed discrimination, MT

1 Introduction

The visual capabilities in humans have motivated a large number of scientific studies. The performance to perceive and interpret visual stimuli in different species including humans is outstanding: the wide range of tolerance to different illumination and noise levels are just a few characteristics that we are aware of, but that are still barely understood.

An important aspect in visual processing is the perception of motion. Motion is a key step in several computer vision tasks such as 3D reconstruction, feature tracking, time-to-collision estimation, novelty detection, among others [1]. Motion is also one of the features that many species can perceive from the flow of visual information, and its detection has been observed in a large number of animals [2], from invertebrates to highly evolved mammals. From optical engineering and experimental psychology we already know the main features of human motion discrimination [3]. In this work, we are particularly interested in taking inspiration from biology in order to design a parallel algorithm with similar discrimination capabilities as obtained by classical serial architectures.

Our work begins by presenting an overview of techniques to detect motion in machine vision to continue with the available experimental results and their procedures in human psychophysics. Section 3, presents our algorithm for speed detection with which we perform our simulations, and compare with experimental data. The results are analyzed in section 4, and in the last two sections, we present the discussion and conclusions about our work.

2 Overview

In this work we are interested in the detection of motion, specifically in the coding and retrieval of speed ($\mathbf{v}$), and in the link between the idea of selecting a range of speed to work with, and providing wider ranges of discrimination as observed in human psychophysics experiments [4]. We focus on two features: the multi-scale architecture of the speed detection, and the relation between the number of multi-scale levels and the range of speeds the system is sensitive to.

2.1 Motion detection in computer vision

The detection of motion is a widely used operation in computer vision. Commonly called "optical flow extraction", the main objective is to assign a vector $\mathbf{v} = (u, v)$ to each frame pixel from a given sequence of frames (at least two). In this section, we explain the basic technique to increase the motion range of an optical flow extraction which the method is sensitive to. We ground our explanation on the well-known Lucas & Kanade's method [5, 1] (the basic multi-scale technique similarly applies to other methods for optical flow extraction).

Optical flow Many optical flow extraction methods are based on the initial assumption of brightness conservation, that is,

$$\frac{dI(x,y,t)}{dt} = \frac{\partial I}{\partial x}u + \frac{\partial I}{\partial y}v + \frac{\partial I}{\partial t} = 0 \tag{1}$$

where $\mathbf{v} = (u, v)$ is the velocity vector. A well known technique following this approach is the Lucas & Kanade algorithm [5], that minimizes the following cost function in a small fixed region Ω, *i.e.*

$$\mathbf{v} = \arg\min_{\mathbf{v}} \left\{ \sum_{\mathbf{x} \in \Omega} W^2(\mathbf{x}) \left[\nabla I(\mathbf{x}, t) \cdot \mathbf{v} + \frac{\partial I}{\partial t}(\mathbf{x}, t) \right]^2 \right\} \tag{2}$$

where W is a two-dimensional Gaussian function used to give more importance to the central points and Ω is a square region of a few pixels. This minimization estimates $\mathbf{v}$ with sub-pixel precision after a few iterations. This method achieves good optical flow extraction in regions where $|\nabla I(\mathbf{x}, t)| > 0$, such as corners [6].

Serial multi-scale optical flow The Lucas & Kanade method for optical flow extraction considers a small region Ω. The use of this region Ω is not particular to this method: it is used in most algorithms [1]. As the computation is performed in small windows, the detection of motion is constrained to detect speeds up to ω pixels per frame, where ω stand for the diameter of Ω. To overcome this limitation, a multi-scale representation of the images can be performed, usually by considering Gaussian pyramids [7]. A Gaussian pyramid representation of an image is computed by recursively smoothing (using a Gaussian kernel) and sub-sampling the original image. In this way, the original image is represented by a set of smaller images. The representation at scale level $l = 0$ is the original image itself. The image at level l is obtained by sub-sampling a filtered version of the image at level $l-1$ with a downsampling factor equal to 2. Thus, the size of the image at each level l is $N_l = N_{l-1}/2$ with $l = 1, 2, \ldots, (L-1)$, where L is the number of levels of the representation.

In the serial multi-scale optical flow estimation, speed is computed by sequentially projecting the estimation obtained at level l to level $l-1$, until level $l = 0$. There are complex strategies for computing the optical flow with a multi-scale approach [6]. A simple solution for optical flow computation is implemented in the widely used computer vision library OpenCV [8, 7]. In this case, the multi-scale estimation starts from the highest level ($l = L-1$) and it propagates to the next one:

$$\mathbf{v}_{l-1} = 2 * \mathbf{v}_l + d_{l-1}(\mathbf{v}_l) \tag{3}$$

where d_{l-1} is the estimation of velocity at level $l-1$ after projecting the estimation v_l by warping the image by $-v_l$ at level $l-1$. Computing the optical flow from the highest level and then projecting the solution to the lower level [6, 7] increases the range of detectable speeds. This range is wider when more scales are used. On the other hand, the sequential projection between levels also propagates the error introduced at each level. Thus, in terms of precision, increasing the number of scales in the representation increases the error introduced in the estimation.

2.2 Biological elements

This section sketches out the current experimental knowledge in biology, focusing on studies of speed coding in the human brain [9] and on higher level descriptions of speed discrimination from experimental psychophysics [3, 10, 11].

Parallel architecture In the human brain the main area that is responsible for coding different speeds is area MT [2]. It is located in the occipital region (back of the head). Neurons in this area are selective to stimuli moving at a given speed [12]. Their spatial organization is retinotopical [13]: each neuron has a reduced visual field, and neurons who share the same local visual field are grouped together in some macro-column that contains cortical columns that are selective for different orientations. This configuration allows a complete mapping of the visual field with a group of cortical columns that codes for all possible directions of local motions. The spatial organization is less known with respect to the speed selectivity. Nevertheless, it has been found that (1) the average detected speed increases with eccentricity (with respect to the retinotopical organization of MT), (2) similar speeds are detected by neurons closer than for distinct speeds, and (3) for each eccentricity, there are neurons for different speeds [13]. The interactions between different units is not completely understood, but there is evidence that units sensitive to different speeds could be coding a range of speeds in parallel [9]. It has been observed that the range of detectable speeds is not uniformly covered [12], but in this work we are interested in the simultaneous existence of speed selective units in MT that could be accountable for a parallel architecture dealing with different speeds.

Speed discrimination In the work of McKee et al. [3], two subjects were exposed to several stimuli, one of those being a horizontal scaled single bar vertically moving at different eccentricities. The goal of this experiment was to determine the minimal relative detectable variation in speed for every subject with the sight fixed at a certain location and for each stimuli eccentricity.

It is important to mention how this was actually measured, because the subject cannot assign a precise velocity at each location. Instead, given a reference velocity, the subject was asked to indicate whether the next presented stimuli moves faster or slower. The minimal detectable variation was then statistically inferred. Related experiments were performed by others [4, 10, 11], showing that the measurements are not affected by different contrast conditions, and that they do not depend on binocular or monocular sight.

The described experiments study the speed discrimination at several eccentricities[1], see Fig. 1. In this work we are interested in each one of these eccentricities and their related discrimination properties, and not in the relations between different eccentricities. In order to model these discrimination functions, we need to generate a given discrimination percentage in a range of speed $[v_1, v_2]$. We also point out that the left side of the experimental curves, see Fig. 1, is related to the eccentricity but the same idea holds: for each eccentricity there is a wide range of speed discrimination, where the relative error (rather than the absolute error) remains stable (5%-15%).

3 Proposed parallel multi-scale speed detection

Multi-scale speed detection is based on the fact that a particular speed detection algorithm can be used to estimate slower speeds at lower levels and to estimate faster speeds at higher levels. This information is used in the above described serial multi-scale optical flow algorithm to detect speeds in a wide range of velocities by projecting the information at level $l+1$ to estimate speed at level l, *i.e.* in a serial manner. As it is described in [12, 9], it seems that human motion perception is based on a parallel multi-scale scheme. Based on this idea, the speed

[1] Distance to the center of the eye in foveated visition (humans, primates and others).

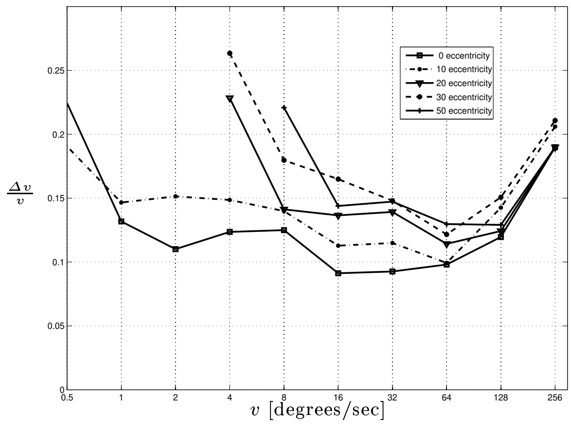


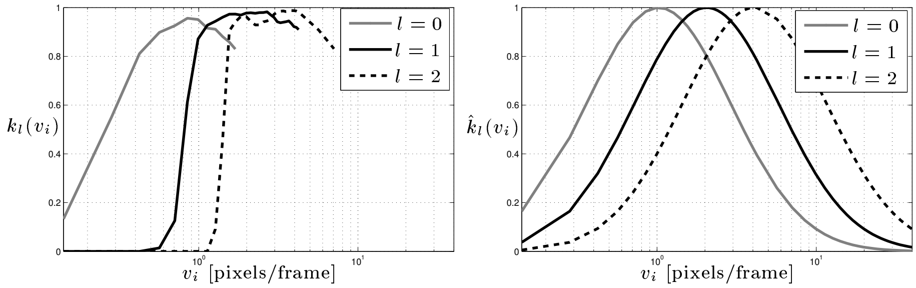
Fig. 1. Weber fraction (minimum threshold of perceived change in speed or $\Delta v/v$) as function of velocity, for different eccentricities in the human subject B.D.B. (condition 2), in the differential motion experiment by [4] (reproduced from their work), each curve is at eccentricities 0,10,20,30 and 50, respectively (left to right). The speed axe is in logarithmic scale.

detection algorithm proposed in this paper estimates the speed by combining the information computed at each level independently, *i.e.* using the multi-scale information in a parallel manner. In this case, there is no error propagation on the computation of speeds at each level because it does not depend on the estimation performed for other levels. At each level l, we compute speeds using the optical flow estimation algorithm described above, see subsection 2.1. As explained before, this choice does not bias our results, since our work is to provide a bio-inspired parallel speed detection instead of the standard serial approach, for any optical flow extraction method.

As expected, the speed detection algorithm estimates speed with a certain error at each multi-scale level l. The confidence in the estimation of speed v_r at level l, denoted as $k_l(v_r)$, can be defined as

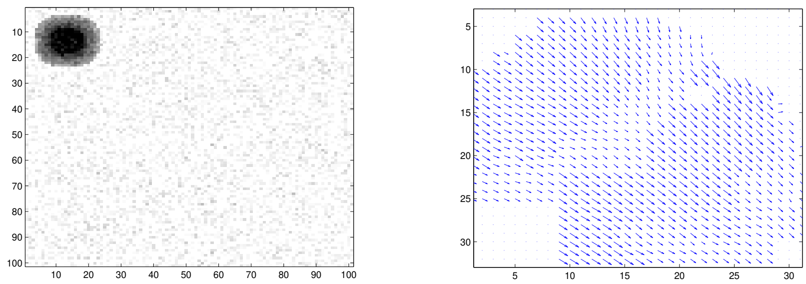
$$k_l(v_r) = 1 - \left| \frac{v_r - v_e}{v_r} \right| \tag{4}$$

where v_r is the magnitude of the object's real speed ($v_r = \|\mathbf{v}_r\|$) and v_e is the magnitude of the average estimated speed on the object pixels location. It can be noted, that this computation only takes into account the magnitude of the speed, ignoring its direction. Figure 2(a) shows the confidence $k_l(v_r)$ for three different multi-scale levels. These distributions were computed using an input image sequence containing an object moving at different speeds in a range from 0.5 pixels per frame to 20 pixels per frame. To statistically determine the confidence at each level l and speed v_r, the experiments were carried out using the input image sequence with several realizations of Gaussian white noise, then the resulting confidence $k_l(v_r)$ is computed as the mean value of the ones obtained in the experiments. Figure 3(a) shows two frames of an input image sequence used in the experiments. In this sequence the object is moving at 10 frames per pixel in the bottom-right direction.



(a) (b)

Fig. 2. Confidence distribution k_l for different levels l. (a) Experimental distributions k_l. (b) Approximated distributions $\hat{k}_l$.



(a) (b)

Fig. 3. In (a) one frame of an input image sequence used in the experiments is depicted ($v = 10$ bottom-right direction). (b) is the obtained optical flow using the proposed parallel multi-scale algorithm. Note that optical flow image was zoomed around the object position.

As it may be seen in Figure 2(a), a particular speed v_r can be detected at several multi-scale levels but with different confidence values. Thus, the current speed could be estimated by taking into account the speeds computed at each level l and their associated confidence values k_l. For that reason, the experimental distributions depicted in Fig. 2(a) have to be approximated by a closed-form equation. In this work, these distributions are approximated (modeled) as Gaussian distributions in a semi-log space defined by the following equation

$$\hat{k}_l(v_r) = \exp\left(-\left[\frac{\log(v_r) - \mu_l}{\sigma_l}\right]^2\right) \tag{5}$$

$$\mu_l = \mu_0 + \log(c^{l-1}) \tag{6}$$

$$\sigma_l = \sigma_0 \tag{7}$$

where μ_0 and σ_0 are the mean and variance of the distribution at level $l = 0$ and c is the scaling factor used in the sub-sampling of the images. The approximated distributions $\hat{k}_l$ for $l = 0$, $l = 1$ and $l = 2$ are depicted in Fig. 2(b). It may be seen that a better approximation of the distributions could be obtained using a particular set of variables for each level but this would increase the model complexity. The approximation of the distributions $\hat{k}_l$ for each level in Eq. (5) only depends on μ_0, σ_0 and c. It may be noted that this approximation allows to perform the estimation of speeds using different values of the scaling factor c, which is usually set to $c = 2$, *i.e.*, the case of using Gaussian pyramids for the sub-sampling.

Finally, denoting the detected speed at each level l by $\mathbf{v}_e^l$, the proposed algorithm computes the current speed, using the speed detected at each multi-scale level with its associated confidence value $\hat{k}_l(\|\mathbf{v}_e^l\|)$, as

$$\mathbf{v}_f = \frac{\sum_{l=0}^{L-1} \mathbf{v}_e^l \hat{k}_l(\|\mathbf{v}_e^l\|)}{\sum_{l=0}^{L-1} \hat{k}_l(\|\mathbf{v}_e^l\|)} \tag{8}$$

where L is the number of levels used to compute the estimated speed $\mathbf{v}_f$. Figure 3(b) shows the obtained optical flow using the proposed parallel multi-scale algorithm. The comparison between the experimental confidence distribution of the proposed algorithm and confidence distributions for three levels is shown in Fig. 4. As expected, the confidence distribution of the parallel multi-scale algorithm with $L = 3$ is approximately the envelope of the confidence distributions of levels $l = 0$, $l = 1$ and $l = 2$.

4 Results

As it was described in subsection 2.2, speed discrimination is computed as the minimal detectable variation in speed of a particular visual stimuli. In this work, a variation in speed, from a given reference speed v_{obj}, of the moving object is considered to be noticeable if the following inequality holds

$$\left|\frac{\widehat{v}_{v_{obj}} - \widehat{v}_{v_{obj} \pm \Delta v_{obj}}}{v_{obj}}\right| > \alpha \tag{9}$$

where $\widehat{v}_{v_{obj}}$ and $\widehat{v}_{v_{obj} \pm \Delta v_{obj}}$ are the speeds estimated by the algorithm when the object is moving at velocities v_{obj} and $v_{obj} \pm \Delta v_{obj}$, respectively, and α is the

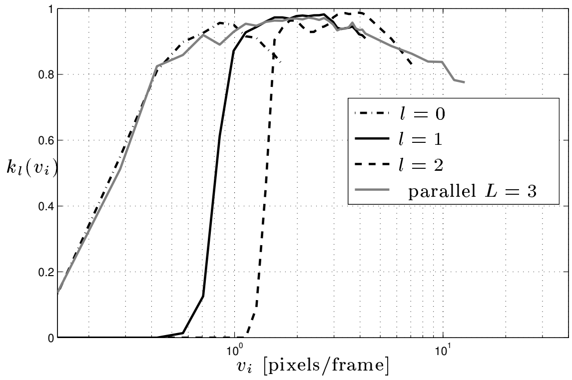


Fig. 4. Comparison between confidence distribution of the proposed algorithm (parallel) with $L = 3$ and $c = 2$, and confidence distributions k_l for levels $l = 0, 1$ and 2 without projection between them.

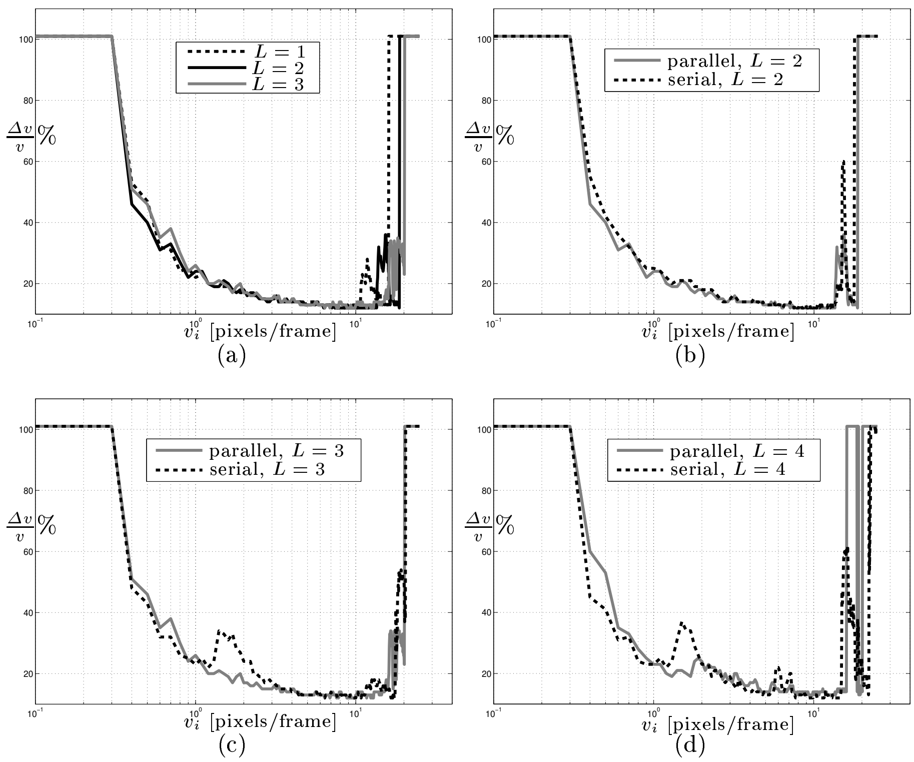


Fig. 5. (a) Discrimination of the proposed algorithm for $L = 1$, $L = 2$ and $L = 3$ ($c = 2$). In (b), (c) and (d), the comparisons between the discrimination of the proposed parallel ($c = 2$) and the serial algorithms for $L = 2$, $L = 3$ and $L = 4$, respectively, are shown.

percentage of variation from the object speed v_{obj} required to consider Δv_{obj} as detectable. It may be noted in Eq. (9) that a variation in speed is considered noticeable if it is detectable when v_{obj} is both increased and decreased in Δv_{obj}. To statistically determine the minimum value of Δv_{obj} several experiments were carried out using the input image sequence with several realizations of Gaussian white noise. Then, the minimal detectable variation in speed, from a given reference speed v_{obj}, is computed as the minimum detectable Δv_{obj} obtained in 90% of the experiments.

We summarize our results in Fig. 5. Figure 5(a) shows the discrimination of the proposed parallel multi-scale algorithm for different values of L. The range of discriminated speeds is enlarged when the number of levels used in the multi-scale representation increases. In comparison with the serial multi-scale, our method has a similar range of speed discrimination when the same number of levels are used, see Fig. 5(b), 5(c) and 5(d). Considering both mean and variance of the discrimination in the range of speeds from 1 to 15 pixels per frame, the parallel multi-scale method shows lower values. For the case of $L = 3$, parallel discrimination has mean= 14.1 and variance= 2.2, while serial discrimination has mean= 15.5 and variance= 4.2. This indicates that the proposed parallel algorithm presents a better discrimination in this range.

5 Discussion

Recent works [6, 14] have developed the idea of multi-scale estimation of speed. First, Simoncelli [6] proposes a bayesian scheme to compute the error distributions and then to estimate the velocity using a Kalman filter through the space of scales (not time). This approach builds a far more sophisticated error function, but it is still serial. Our work assumes that the error functions are fixed, while [6]assumes the error changes with respect to $\nabla I(\mathbf{x}, t)$, and this might be important in real-world scenarios. On the other hand, Chey et al. [14] propose that considering higher threshold levels for higher scales (scale-proportional thresholds) and inter-scale competition could explain human speed discrimination curves. We have presented a scheme where the response of each scale regulates the relevance of the responses of that scale. Since we handle all scales at the same time, it corresponds to a notion of threshold and competition. To our knowledge, no other work models error functions as Gaussians in the log space (this strengthen the idea that detection is not symmetrical), which seems to fit recent recordings of motion sensitivity of neurons [9].

Finally, about the time complexity of our algorithm. Lets consider the size of the image as N, the order of the optical flow algorithm as K (clearly $K(N)$) and the number of scales l. The complexity order of the serial multi-scale algorithm is $lN + lK$, and $N + lK$ for the parallel algorithm. The only difference is in the operations involved in the merge of scales. Considering the possible speed-up using p processor for the case $p = l$, then $S_p = \frac{lN+lK}{N+Kl/p}$, what can be also written as $S_p = p$. This last equation show us that the degree of parallelism (taking one level by processor) achieved by our proposed algorithm is linear.

6 Conclusions

In this work we have presented a parallel multi-scale algorithm to perform the estimation of motion using two consecutive images. This method takes bio-inspiration from human physiology and psychophysics knowledge in the sense that it achieves wide uniform relative discrimination properties by using evenly spaced logarithmic scales, and it gives results in constant time as a function of scales. With respect to the classical serial multi-scale optical flow algorithm, the error propagation among scales appears less important for our proposed algorithm in terms of relative discrimination. We now explore the idea of using more biologically plausible methods of optical flow extraction and the integration with a foveated topology.

References

1. J. L. Barron, D. J. Fleet, S. S. Beauchemin, and T. A. Burkitt. Performance of optical flow techniques. *CVPR*, 92:236–242, 1994.
2. M. Srinivasan and S. Zhang. *Motion Cues in Insect Vision and Navigation*, pages 1193–1202. MIT Press, Cambridge, MA, 2003.
3. S. P. McKee and K. Nakayama. The detection of motion in the peripheral visual field. *Vision Research*, 24:25–32, 1984.
4. G. A. Orban, F. V. Calenbergh, B. De Bruyn, and H. Maes. Velocity discrimination in central and peripheral visual field. *J. Opt. Soc. Am. A*, 2(11):1836, 1985.
5. B. D. Lucas. *Generalized image matching by the method of differences.* PhD thesis, Pittsburgh, PA, USA, 1985.
6. E. P. Simoncelli. Bayesian multi-scale differential optical flow. In B. Jähne, H. Haussecker, and P. Geissler, editors, *Handbook of Computer Vision and Applications*, volume 2, chapter 14, pages 397–422. Academic Press, April 1999.
7. M. J. Black. *Robust incremental optical flow.* PhD thesis, New Haven, USA, 1992.
8. G. Bradski and A. Kaehler. *Learning OpenCV: Computer Vision with the OpenCV Library.* O'Reilly Media, Inc., 1st edition, October 2008.
9. H. Nover, Anderson C. H., and G. C. DeAngelis. A logarithmic, scale-invariant representation of speed in macaque middle temporal area accounts for speed discrimination performance. *The J. of Neurosci.*, 25(43):10049–10060, October 2005.
10. A. B. Metha, A. J. Vingrys, and D. R. Badcock. Detection and discrimination of moving stimuli: the effects of color, luminance, and eccentricity. *J. Opt. Soc. Am. A*, 11(6):1697, 1994.
11. J. J. Koenderink, A. J. van Doorn, and W. A. van de Grind. Spatial and temporal parameters of motion detection in the peripheral visual field. *Journal of the Optical Society of America A*, 2:252–259, February 1985.
12. J. H. R. Maunsell and D. C. Van Essen. Functional properties of neurons in the middle temporal visual area (mt) of the macaque monkey: I. selectivity for stimulus direction, speed and orientation. *J. Neurophysiol.*, 49:1127–1147, 1985.
13. J. Liu and W. T. Newsome. Functional Organization of Speed Tuned Neurons in Visual Area MT. *J. Neurophysiol.*, 89(1):246–256, 2003.
14. J. Chey, S. Grossberg, and E. Mingolla. Neural dynamics of motion processing and speed discrimination. *Vision Research*, 38(18):2769–2786, September 1998.