THESIS FOR THE DEGREE OF LICENTIATE OF ENGINEERING IN MACHINE AND VEHICLE SYSTEMS

Bio-inspired retinal optic flow perception in robotic navigation

BJÖRNBORG NGUYEN

Department of Mechanics and Maritime Sciences
Division of Vehicle Engineering and Autonomous Systems
CHALMERS UNIVERSITY OF TECHNOLOGY

Göteborg, Sweden 2021

Bio-inspired retinal optic flow perception in robotic navigation
BJÖRNBORG NGUYEN

Thesis for the degree of Licentiate of Engineering 2021:01
ISSN 1652-8565
Department of Mechanics and Maritime Sciences
Division of Vehicle Engineering and Autonomous Systems
Chalmers University of Technology
SE-412 96 Göteborg
Sweden
Telephone: +46 (0)31-772 1000

Cover:
Color encoded retinal optic flow field visualized during a steering task in a road vehicle from the driver's point of view.

Chalmers digitaltryck
Göteborg, Sweden 2021

Bio-inspired retinal optic flow perception in robotic navigation
BJÖRNBORG NGUYEN
Department of Mechanics and Maritime Sciences
Division of Vehicle Engineering and Autonomous Systems
Chalmers University of Technology

Abstract

This thesis concerns the bio-inspired visual perception of motion with emphasis on locomotion targeting robotic systems. By continuously registering moving visual features in the human retina, a sensation of a visual flow cue is created. An interpretation of visual flow cues forms a low-level motion perception more known as *retinal optic flow*. Retinal optic flow is often mentioned and credited in human locomotor research but only in theory and simulated environments so far. Reconstructing the retinal optic flow fields using existing methods of estimating optic flow and experimental data from naive test subjects provides further insight into how it interacts with *intermittent control* behavior and dynamic gazing. The retinal optic flow is successfully demonstrated during a vehicular steering task scenario and further supports the idea that humans may use such perception to aid their ability to correct their steering during navigation.

To achieve the reconstruction and estimation of the retinal optic flow, a set of optic flow estimators were fairly and systematically evaluated on the criteria on run-time predictability and reliability, and performance accuracy. A formalized methodology using containerization technology for performing the benchmarking was developed to generate the results. Furthermore, the readiness in road vehicles for the adoption of modern robotic software and related software processes were investigated. This was done with special emphasis on real-time computing and introducing containerization and microservice design paradigm. By doing so, continuous integration, continuous deployment, and continuous experimentation were enabled in order to aid further development and research. With the method of estimating retinal optic flow and its interaction with intermittent control, a more complete vision-based bionic steering control model is to be proposed and tested in a live robotic system.

Keywords: Retinal optic flow fields, retinal flow path, gazing, active fixation, visual perception, optic flow, robotic navigation, bionics

Only one soul was harmed or transformed during the development of this thesis

Acknowledgements

I would like to sincerely thank my supervisors Prof. Ola Benderius and Prof. Christian Berger for their excellent support and encouragement towards my development as a researcher and foremost their faith in my work. I am forever grateful for the experiences and joyful moments we shared along our journey.

I would also like to show my sincere gratitude towards my examiner Prof. Mattias Wahde for allowing me to enter the world of academia. It is certainly a unique and stimulating place to be in.

My work would not be possible without the support provided by Dr. Fredrik Von Corswant, Arpit Karsolia, and the supporting staff at Revere. I am forever grateful for the valuable and friendly working environment that exists there.

I want to acknowledge all former, current, and future colleagues and coworkers at VEAS for contributing to a friendly, inviting, and relaxing work environment. Coffee breaks are always something to long for, thanks to you guys. A special thanks go to the Ph.D. student protector and hero, Sonja Laakso Gustafsson, for standing up and raising a voice for all Ph.D. students at VEAS and all her supporting work which enables us, Ph.D. students, to do what we need to do.

And finally, I want to thank my friends and family for supporting me in my work and encouraging me during difficult periods, and especially Molly for reminding me to sleep, take the much-needed breaks, and keeping me safe during the pandemic with her barks.

List of included papers

This thesis consists of the following papers:

Paper A B. Nguyen and O. Benderius. Retinal optic flow, active gaze, and intermittent control. *Nature Communications* (submitted)

Paper B B. Nguyen, C. Berger, and O. Benderius. “Systematic benchmarking for reproducibility of computer vision algorithms for real-time systems: The example of optic flow estimation”. *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE. 2019, pp. 5264–5269

Paper C C. Berger, B. Nguyen, and O. Benderius. “Containerized development and microservices for self-driving vehicles: experiences and best practices”. *2017 IEEE International Conference on Software Architecture Workshops (ICSAW)*. IEEE. 2017, pp. 7–12

Table of contents

Chapter 1

Introduction and motivation

Robotic systems are expected to integrate and become a part of the human society [40]. This imposes requirements of them having navigation abilities in their respective robotic application. Resulting in making navigation highly relevant in current research and development in academia and industry alike. Many modern robotic systems can already optimally solve navigation in a static environment with perfect knowledge of its surroundings. However, the general case with uncertainties is difficult to solve, and therefore there is still a need for new theories and methodologies.

Looking to biologically inspired engineering, or **bionics**, may provide another way of solving the navigation task. For example, it is not a coincidence that the shape of a modern aircraft has a striking resemblance to a bird. Biology and evolution have been finding solutions to non-trivial tasks or problems since the inception of life. Therefore finding inspiration and studying biological solutions to navigation can be exploited to the human advantage. The goals of this thesis are to understand how humans perceive, plan, and locomote and further apply this newly gained knowledge towards solving navigation task.

Unsurprisingly, creatures can consciously navigate in order to find food, mate, or avoid danger. Among microscopic life, navigation without a target is sufficient for sustaining life [42]. However, for larger animals, target-based navigation is needed to conserve energy. Despite the seemingly simple task of moving towards a target, one does not fully understand exactly how it is done from perception to action. The navigation task may be discerned into smaller sub-components: perceiving the environment, planning the actions, and performing the locomotion. There is an on-going debate in the research community on what strategies humans use to perceptually guide their locomotion such as walking, driving, or flying. One established strategy is the *you should look where you are going*-strategy [48], which details how humans use **retinal optic flow fields** to perceive the world in motion in order to control the steering movements.

In fact, retinal optic flow has been identified as an integral part of perceiving motion, localization, and motion control [51, 48, 46]. Much of the previous work related to retinal optic flow has been done in interdisciplinary fashion between psychology, behavioral science, neuroscience, computer vision, and mathematics on a qualitative level and within a theoretical framework.

Humans achieve retinal image stabilization through various biomechanical processes, for example, **active fixation** (smooth pursuit) through fine eye movement, and **vestibulo-ocular reflexes** through sensing and counter-acting head movements. This is to aid the formation of the imagery with as clear visual details as possible. However, directed eye gazing and fixations are not only for image stabilization navigation, as vision is also shared for scene perception. Currently, there is a lack of research on the interactions of these perceptual functions during locomotion. However, there has been a recent study on human gazing behavior and foot-placement during locomotion in rough natural terrain [28].

Another field of research focusing on the integration of perceptual functions and action is driver modeling. Such models often include cognitive aspects such as gazing, looming, **focus of expansion** (vanishing point), or even retinal optic flow. However, when it comes to retinal optic flow, it is often confused with optic flow and in many cases over-simplified. For the full representation of retinal optic flow, the photonic interaction between the scene and the agent has to be considered.

In computer vision, estimating the closely related **optic flow** (optical flow) is a well-established problem domain. The optic flow estimation proved to be useful in computer vision, for example, in video compression, object motion analysis, image stabilization, motion detection, simultaneous localization and mapping, and visual odometry. Unlike *retinal* optic flow, optic flow disregards biomechanical processes that produce a flow displacement to achieve, for example, retinal image stabilization as mentioned above.

This thesis investigates how retinal optic flow can be analyzed, reconstructed, and used in control with the ambition of implementation in a technical system, such as a robot. One example could be an autonomous road vehicle, as they are likely to include powerful computational capabilities and various built-in sensors. This type of system demands for modern architecture both in software and hardware. In order to process and estimate retinal optic flow in a live system, using conventional or event-based cameras will require highly efficient data pipelines. In a larger, or commercial, system this requires well-developed hardware and software architectures and software deployment.

This thesis does not consider deep learning approaches for the investigation of retinal optic flow. Even if it might perform well, the deep learning approach is unable to provide a deeper analysis or understanding of the underlying problem which is the main purpose of this work. Moreover, it has even argued that the societal implication of blindly applying end-to-end solutions might do more harm than good [39].

By studying the human visual perception of motion in a navigation context, the retinal optic flow has been identified as an important source of information. It is clear from previous research that retinal optic flow is used during locomotion, and the current research debate is on the details of *how* humans make use of it. While within computer vision, methods have been developed for estimating optic flow, but it slightly differs from the retinal counterpart as it lacks the gazing dynamics. Furthermore, road vehicle systems are evolving, equipped with sensors, and more advanced computational units. Therefore they could represent a good robotic platform for the deployment of flow-based methods of human-like locomotion.

In summary, this thesis investigates: (i) the modeling and reconstruction of retinal optic flow field as perceptual input for navigation, (ii) benchmarking of optic flow algorithms to be used in formal real-time applications, and (iii) the readiness of road vehicles for the adoption of modern software development within bionics.

1.1 Retinal optic flow field

Retinal optic flow may be regarded as an extension of the more known and well established optic flow in computer vision. The main difference between the two is that retinal optic flow captures the biomechanical dynamics which consist of complex head- and eye movements into the perception of motion, achieving for example retinal image stabilization. Thus, by combining the biologic aspects of gazing and optic flow fields, retinal optic flow fields can be computationally reconstructed.

Wann, Wilkie, Swapp, Land, and colleagues have been significant advocates of retinal optic flow as the primary source of visual cue for locomotion [52, 46]. In one of their papers [48], they argued for a steering control strategy using retinal optic flow field by continuously *nulling flow curvature* through active fixation. By intelligently choosing the gazing point, by simply *looking where you want to go*, the flow of the ideal path, i.e. retinal optic flow path, will only consist of a vertical flow component at the correct instantaneous motion. This online control proposal was shown as a mathematical plausibility using the retinal optic flow field to be exploited as guidance for circular locomotion.

1.2 Scope and contributions

The retinal optic flow field has been credited as a quintessential part of human navigation in terms of localization and steering control tasks during motion. This thesis considers the retinal optic flow as a source of perceptual information in the context of robotic navigation applications. Specifically, with the intended use-case of computational and strict deadline constraints in the example of a road vehicle scenario. Due to the very nature of the retinal optic flow, gazing behavior will be considered from cognitive and biomechanical aspects. Furthermore, some motion control will be discussed as it is tightly coupled with egomotion and plays a vital part in retinal optic flow.

The author was the main contributor to Paper A and Paper B, and one of the three main contributors to Paper C.

Chapter 2

Visual motion perception

There is a famous thought experiment from the 17$^{\text{th}}$ century in philosophy called *Molyneux's problem* which poses an interesting problem on the topic of **sensation** and **perception**. Given a human who is congenitally blind where the person can feel the differences between geometric shapes via tactile touching, could that same person also distinguish and identify between the very same objects by only visually looking if they were given the ability to see? This particular line of thinking emphasizes the differences between the fundamental sensation of different senses to the higher cognitive perception of the same object. Held *et al.* later investigated this via experimental trials [20] by letting the newly sighted patients attempt to pair the haptically touched objects to the seen ones. The answer to the thought experiment was found to be simply *no* since the newly sighted could not establish the perception of the same object via different types of sensations. The researchers were however quick to note that they were learning and developing the cross-modal mapping after the trial. This strongly confirms that perception has a significant learning aspect to it in contrast to sensation.

Visual perception as a whole is thus a complex field to study where experience and learning heavily factors in. There is arguably a difference between visually sense and visually perceive something where the latter requires higher cognitive function. Moreover, different types of visual perception like the perception of patterns, edges, objects, and motion, to name a few, form an intricate and complex cognitive system to construct the mental representation of the environment. This create an **awareness** which some researchers argue is an important, if not the most significant step, towards **consciousness** [15].

The sensation on the other hand does not arguably require higher cognitive function but instead operates on a lower cognitive level. One example would be when a photon hits one of the photo-receptors in the retina which converts the light to an electric action potential, a process called transduction. The transduction process, happening in the retina of animals, continuously collects information and further provides minimal interpretation as input to the neural perception system. The signal will propagate throughout the clusters of neural networks for various purposes like the formation of knowledge and memories, recognition of patterns and objects, and awareness. These biological processes occur without voluntary introspection and the formations are automated to a certain degree. This thesis proposes that low-level vision includes similar automatic properties to the high cognitive load perception ones for the purpose of navigation. This could specifically relate to the *nulling the flow curvature* in steering tasks which is investigated in Paper A as it does not require other higher cognitive functions such as object recognition.

By subconsciously registering visual features such as light, shape, color, or patterns in a scenery, translational and rotational movements create a visual motion or visual flow cue. It is apparent that such a sensation input cue provides crucial information for the observer about the dynamic and moving environment or the observer's locomotion. One indication of its importance can be seen in the documented condition **akinetopsia**, where the ability to process, or even perceive, visual motion is impaired. It has been reported [12, 33, 22] that the affected has noticeable challenges to perceive the world in motion and they describe it as seeing the world in a series of still images instead. To compensate for the lack of motion perception in vision, some has reportedly relied on other sensation for judging the motion, for example using the hearing to judge the motion of approaching vehicles. Interestingly, the affected can still perceive the surrounding scene in the way most people would consider to be perfect scene awareness, but they still lack the crucial parts of the subconscious vision which negatively impacting their daily life.

From the above, it can be argued the visual motion cue has a distinct pathway in the neural system with considerable impact on daily functions. When studying visual motion, it is intuitive and natural to consider the visual movement input cue, the general referred to as optic flow. The optic flow field may be thought of as a three-dimensional visual movement-vector at every visual point, projected to a two-dimensional space. This does not however consider the observer's higher cognitive or psychological functions like gazing, retinal image stabilization via biomechanical reflexes, or the tendencies of fixating the gaze to visual details. Each of the aforementioned phenomenons is an important aspect of the perception of visual motion. In the case of humans as they quickly can change the direction of their eyes, turn their heads, and according to Henderson intelligently tend to guide their gaze through the scene using knowledge-driven gaze control [21].

2.1 Optic flow

In the 1950s, Gibson published influential work [19] where the optic flow concept was formally conceived in the academic world. The author made the example of pilots guiding their aircraft to the runways during level flight using optic flow as a primary source of information for locomotion. In addition to the optic flow theory, the idea and the use of **point of aim**, **vanishing point** or **focus of expansion** were introduced and developed. The focus of expansion is the point of divergence in the optic flow field analog to the divergent source point in vector field analysis. Early interpretation and purpose of FOE were that direction or heading of locomotion could be detected and used for motion control. A large part of the early research focused on the optic flow field in general and the perception of heading. Optic flow and retinal optic flow have unfortunately been used interchangeably in the early research whereas the latter consider biological phenomenons as mentioned above.

In development and explaining the retinal optic flow phenomenon, Longuet-Higgins and Prazdny made significant contributions in modeling and interpreting moving retinal images in a theoretical mathematical framework [26]. Their work describes an image plane where a world scene texture is projected to, very similar to a pinhole camera model. By introducing a relative motion, any combination of translation and rotation, between the observer and the texture, retinal optic flow emerges and can be described mathematically. A common description for the emergent optic flow is using the pin-hole camera model which is somewhat similar to the human eye. The flow may be described as

$$\vec{Q} = \begin{pmatrix} \dot{x} \\ \dot{y} \end{pmatrix} = \underbrace{\frac{1}{Z}\begin{pmatrix} -f & 0 & x \\ 0 & -f & y \end{pmatrix}\begin{pmatrix} v_x \\ v_y \\ v_z \end{pmatrix}}_{\text{translational component}} + \underbrace{\frac{1}{f}\begin{pmatrix} xy & -(f^2+x^2) & fy \\ (f^2+y^2) & -xy & -fx \end{pmatrix}\begin{pmatrix} \omega_x \\ \omega_y \\ \omega_z \end{pmatrix}}_{\text{rotational component}} \tag{2.1}$$

where x, y represents pixel coordinates in the image space, f the focal length, $\vec{v}$ the observer's velocity, $\vec{\omega}$ the observer's angular velocity, and Z the depth of the arbitrary visual point projected to the image space. The dots denote the differentiation with respect to time. The optic flow as seen in the Eq. (2.1) may emerge from either translation, rotation, or any combination of the both (examples of these is illustrated in Figs. 2.1 and 2.2). It is to be noted that the information of the scene is required for computing optic flow in the translational component

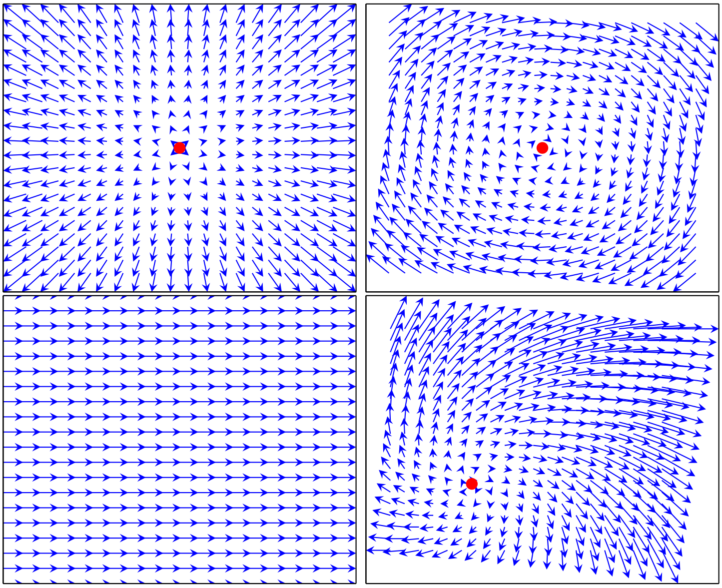

Figure 2.1: *Examples of sparse optic flow fields during different translational and rotational scenarios from a perpendicular textured plane in front of the observer.* ***Upper left panel:*** *a pure translation into the image plane is applied which creates a divergence around the focus of expansion which is depicted as a red dot.* ***Upper right panel:*** *a rotation is applied instead which creates a curled flow field around the axis of rotation through the image plane.* ***Lower left panel:*** *a translation to the left creates a uniform field to the right with a focus of expansion outside of the field of view to the left.* ***Lower right panel:*** *All three scenarios are combined, creating a field with an expanding and curling field and displaces the focus of expansion. The dense optic flow equivalent visualizations are illustrated in Fig. 2.2.*

as it is dependent on the Z depth, in contrast to the rotational component which independent of the scene. This implies making this model more suitable when perfect knowledge of the observer and the scene is accessible for example in a simulation environment. In a real-world setting, where incomplete information or insufficient accuracy is common, the optic flow model would most likely not suffice. There are alternative methods developed for estimating optic flow which will be briefly discussed in Sect. 3.3.

In contrast to the straightforward and naive manner of modeling optic flow, as shown Eq. (2.1), Prazdny made further attempts to reverse the process and extract information such as egomotion and depth from the visual optic flow alone [37]. It was concluded that the information is not conserved in the structure of optic

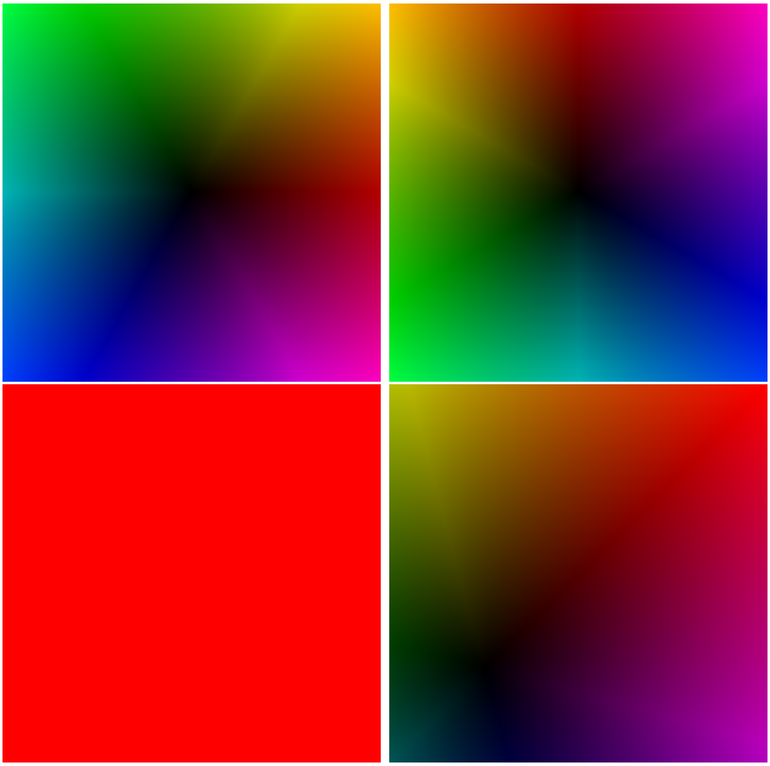

Figure 2.2: *The dense optic flow equivalent of Fig. 2.1. Color encoding is used instead of quiver plot, where the flow vector angle represents the hue and magnitude the value in HSV color encoding. The focus of expansion can be located in the point where it is the darkest.*

flow and is lost in the process of projective transformation. Neither the relative depth and local surface orientation can be retrieved since they are unambiguously stored in the optic flow. Thus, if such information is to be extracted, further assumptions and prior knowledge or experience are required to estimate or guess these parameters.

Much of the previous research has been focused on the relative motion between the observer and the scene itself, but an optic flow may also emerge from independent moving objects which Layton *et al.* investigated from a neurophysiological point of view [24]. By studying the optic flow, the closely related focus of expansion, and the optic expansion of objects, so-called looming, they developed neuro-models inspired by the biological processes. It was mainly designed to mimic the human heading perception in presence of independently moving objects in a parallel pathway fashion. They found and concluded that their models also could produce the heading biases in the same direction and magnitude as similarly found in experimental data of humans.

Throughout the early research of optic flow, it has mostly been regarded in two-dimensional image space or retinal space. However, with the advancements of complementary sensor frameworks such as **lidars** and **stereo vision systems**, or even using **sensor fusion**, three-dimensional flow space could be considered. The three-dimensional flow is referred to as **scene flow** [45], describing every visual point and its associated velocity. Although scene flow might appear just slightly different from optic flow at first glance, its implication imposes great challenges for the numerical estimation. Much of previous work for estimating optic flow would not be directly applicable for corresponding scene flow thus creating a chasm between optic flow and scene flow research.

2.2 Retinal optic flow

As emphasized above, retinal optic flow incorporates the gazing dynamics and head movements into the flow fields. A simplified schematic of retinal flow is shown in Fig. 2.3 demonstrating the visual dynamics between visual features and the projections to a non-moving human eye.

A large portion of the work related to optic flow may be transferred and extended to the field of retinal optic flow. However, some challenges are made apparent when considering retinal optic flow. For example, how does one consider and estimate gazing and fixation dynamics to incorporate them into the existing optic flow framework? Early research has successfully demonstrated in reconstructing retinal optic flow fields by Calow and Lappe [11]. In their work, they used extensive data consisting of egomotion, eye movement, and depth structure to generate, what they call, true retinal motion fields. Although this thesis does not go deep into the detailed research on a neuron-level analysis, Calow and Lappe do provide a

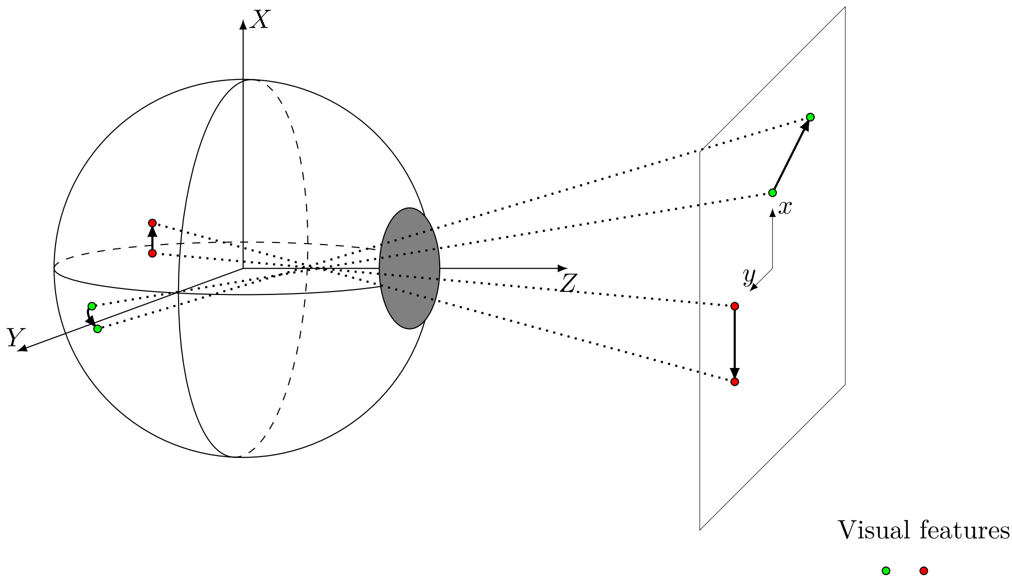

Figure 2.3: *An illustration of optic flow projected to an eye, creating the emergent retinal optic flow. By registering the visual features, green and red dots, over time, the corresponding displacement vector a flow cue emerges which the human perception of motion exploits.*

hypothetical ground on how humans on a neural level could process retinal optic flow using an encoding for the flow. In their modeling of retinal optic flow where polar transformation is applied, they were able to reproduce some properties of real motion-sensitive neurons found in the middle temporal area of the human brain. While their method of reconstructing the flow produces accurate retinal flow fields, it puts a significant dependency on various sensors and data fusion methods and would perhaps not be feasible in real-world applications.

2.2.1 Retinal image stabilization and gaze fixation behaviors

Gazing is described as the bio-mechanical coordination effort of directing one's visual sensation. By continuously directing the gaze and maintaining the fixation on a visual feature, one achieves retinal image stabilization. Many processes contribute to the stabilization of retinal image but active fixation, vestibulo-ocular reflexes, and cervico-ocular reflexes are mainly credited. Without retinal image stabilization, the vision would easily become blurry and misinterpreted by cognitive functions and would result in dizziness or nauseousness. Anecdotally, many have probably experience blurred vision with erratic eye movements and slight dizziness after abruptly stopping the spinning of an office chair; as the vestibulo-ocular reflexes are trying to counteract and compensate for the rapid change in angular velocity.

Vestibulo-ocular reflexes (VOR) correct and adjust the vision mainly for the head movements, both translational and rotational via the vestibular sensation system located in the inner ears of humans. It is the same sensation system responsible for equilibrioception and spatial orientation with respect to gravity. By sensing the fine and jerky head movements, it can inhibit or exhibit signals to the eye muscles to readjust for retinal image stabilization. **Cervico-ocular reflexes** (COR) work similarly but use sensation in the neck to compensate for head movements.

In terms of pure eye movements, there are four distinct types or classifications: saccades, passive fixations, active fixation, and post-saccadic oscillations. A saccade is the rapid and jerky eye movement where the observer temporarily becomes blind during the movement by cognitively blocking visual sensation or **saccadic masking** in order to avoid motion blur in the formation of the retinal image. The purpose of the saccade is to quickly move the gazing point. A passive fixation is when the eyes are locked relative to the eye socket thus rotating the head will also rotate the gaze. In contrast, active fixation is when the gazing point is stable and fixed on the location of a visual target, for example when tracking a moving object. Active fixation has been reported to reach angular speeds up $30\,^{\circ}\,\mathrm{s}^{-1}$, while any greater speed tends to cause saccadic behavior. Lastly, post-saccadic oscillation is a transient state of the eye caused by the rapid saccade, before reaching a steady-state fixation.

Relating these behaviors to the mathematical modeling of retinal optic flow formation can be done from Eq. (2.1). These biomechanical reflexes attempt to null $\vec{Q}$ in the origin (fovea) by adding rotation in the rotational component through a tightly closed control-loop interplay in the eye muscles, vestibulo-ocular reflexes, and active fixation. This results in the retinal optic flow field in the proximity of the point of active fixation or the **fovea** is nearly nulled where the clarity of vision or **visual acuity** is the greatest, thus achieving a retinal image stabilization.

There have been great advancements in eye-tracking technology in recent years which has stimulated and reinvigorated retinal optic flow research. There are several solutions based on different methodologies, both active and passive sensors, for detecting and estimating gaze. Furthermore, Salvucci and Goldberg made a taxonomy of different algorithms in identifying gaze fixations [41]. This enabled new types of eye-tracking technologies and improved existing ones. In their work, they investigated the main advantages and disadvantages among different methods which highlighted the suitability of the different approaches.

As mentioned above, humans tend to intelligently guide their gaze through the scene using knowledge-driven gaze control [21]. This has spurred a debate whether humans use model-free online control mechanisms or some degree of inherent internal models of visual anticipation during locomotion. Tuhkanen *et al.* investigated this issue [43] and found that predictive gazing behaviors emerged during locomotor steering control in a simulated environment. To make these predictive gazing behavior, the subject is required to internalize a model in contrast to a model-free online control where control is immediate and based on currently available information. The latter case implies that the agent can solely by the immediate available visual information guide their locomotion without any additional complex perception. Their findings suggest that human steering actions are driven by internal models and one might argue that it could be the case of general human locomotion. This implies that humans automatically choose where to fixate their gaze in the visual scene during navigation.

2.2.2 Nulling flow curvature with active fixation

In the context of navigation, it has become widely accepted that the retinal optic flow field is used in locomotor control. However, there is an ongoing debate among researchers on *how* the information is perceived, interpreted, and used. A notorious case is whether if the direction of egomotion, or heading, can be extracted and if required for locomotor control.

As mentioned in Sect.2.1, retrieval of information such as egomotion and depth perception could not be exactly determined from optic nor retinal optic flow alone. Despite this, Fajen and Warren still argued that *sufficient* estimation of extra-retinal information could be made for navigational tasks [50, 49]. Hence they proposed a simple additive control law on foot using optic flow

$$\phi - \psi_g = (\phi_{ego} - \psi_g) - wv(\phi_{of} - \psi_g) \tag{2.2}$$

where ϕ is the current heading, ψ_g the current target goal, w some optic flow measure weight, and v the observer's velocity. The proposed theory was also extended to accommodate obstacle avoidance and path routing by slightly modifying the model. This was done by introducing dynamics mimicking the differential equation aspects of stiffness and dampening factors.

On contrary to strategies requiring extra-retinal information, Wann and colleagues argued that such information, in particular heading retrieval, was not required for sufficient locomotor control. Furthermore, the model that Fajen and Warren presented could only explain the special case of walking, and could not be extended to curve bends, i.e. high-speed steering. The argument of Wann *et al.* was simply that humans could simply pivot their torso and take the straight-line trajectory to the target.

Instead, Wann *et al.* suggested and proposed retinal optic flow to be one of the primary information sources of guidance for locomotor steering control. They presented a mathematical framework using retinal optic flow and intended path projection [48]. By continuously minimizing the horizontal flow field component of the intended path, *nulling flow curvature with active fixation*, successful steering control could be realized. This strategy is based on principle *you should look where you are going*, as inspired by driving school instructions. Assuming that the agent locomotes in a circular motion and fixates its gaze onto a point above the current circular path, parallel to the ground plane i.e. towards the horizon. Then, the retinal optic flow field of the intended path can be described as a fraction of horizontal and vertical components, according to

$$\frac{\dot{U}}{\dot{V}} = \frac{\sqrt{2}(r^2 - X_p^2 - Z_p^2)}{(Y_e - Y_p)((X_p + r)\sqrt{1+\cos\theta} + Z_p\sqrt{1-\cos\theta})} \tag{2.3}$$

as further explained elsewhere [47]. In this mathematical derivation, they could conclude that all points on the intended path, i.e. $X_p^2 + Z_p^2 = r^2$, result in the numerator equal zero, implying that no visual horizontal flow is present. Eq. (2.3) as presented here is slightly different from the one presented by the authors, due to a minor derivation error which can be detected through dimensional analysis. Despite the error, it does not affect the overall analysis made by Wann and Swapp.

Recent research shows that both retinal flow and extra-retinal information could support locomotor control through a flexible combination, which Wilkie *et al.* [53] proposed as

$$\ddot{\theta} = k_1(\beta_1\dot{P}_{RF} + \beta_2\dot{P}_{ERD} + \beta_3\dot{P}_{VD}) + k_2(\beta_4 P_{ERD} + \beta_5 P_{VD}) - b\dot{\theta} \tag{2.4}$$

where k_i are gains, b dampening, β_i weights (relative reliance) balancing the perception variables, and $P_{\{RF,VD,ERD\}}$ perception variables for retinal optic flow, visual direction and extra-retinal target direction. This model attempts to incorporate all available signals of perception to then generate a control signal.

In parallel with developing the nulling flow curvature-strategy as presented above, Wann and Land proposed the following five principles for the emergent retinal optic flow during steering locomotor control [46]:

R1. If the flow-lines are curved, then the agent is not on a path to where it is looking and the steering error is in the direction opposite to the flow curvature.

R2. If the agent is on a curved path to the point of gaze, then the flow-lines for ground elements are straight but distributed non-radially.

R3. If the agent is on a linear path to the point of gaze, then the flow-lines are straight and distributed radially.

R4. In both R2 and R3, points that move with pure vertical motion in retinal flow indicate the future path for the current steering angle.

In Paper A, these principles could occasionally be identified in the experimental data, with the slight adjustment of studying the instantaneous retinal optic flow instead of the so-called flow lines. This is mainly due to lacking a formal definition of flow-lines. However, further statistical analysis requires a larger data sets, in order to conclusively confirm if these principles hold in the general case.

Chapter 3

Flow-based vision in robotics

In this chapter, robotic systems in terms of hardware and software will be discussed, as was specifically studied in Paper B and Paper C. The text will start by introducing common vision hardware sensors used for optic flow estimation and proceeding with the software with emphasis on real-time computing and deployment. Finally, an overall description and overview of optic flow estimation and computation will be covered along with established common benchmarking metrics used in the optic flow computer vision community.

3.1 Vision sensors

The most common sensor used for robotic vision is the conventional three or single-channel video camera prevalent in industrial and surveillance settings. In the case of three channels, the typical technical solution is to use a well-established Bayer filter on top of the image sensor to produce the three base colors: red, green, and blue; often referred to as RGB or BGR in reverse. The underlying technique is to exploit how the human visual perception functions, to represent many colors through only three base colors. The additive superposition of the fundamental colors can therefore create a new perceived color in humans. Despite that, even though three-channel vision sensors are common in many applications, single-channel sensors are frequently used in robotic applications, including optic flow estimation, due to faster response time (shorter exposure time), simplification of colors, and less data volumes.

In traditional single or multi-channel vision sensors, the fundamental measurement procedure is to measure illumination over the image sensor at a fixed time, referred to as the exposure time. This can be done either for the entirety of the frame using a global image sensor shutter, or a row-by-row basis using a rolling shutter. The former ensures that all parts of the image originate from the same time, while the latter allows for more efficient use of the hardware and related computational resources. However, in contrast to the traditional way to interpret vision using three or single-channel images generated from fixed-length exposure time, there is another way to do it via so-called **event-based camera** vision or **dynamic vision sensor**. These bio-inspired event-cameras function fundamentally differently in contrast to the traditional vision sensor as they do instead of measuring collected illumination in a fixed time, rather measure the per-pixel-time to fill an illumination buffer, after which they send an event (i.e. in effect measuring the illumination dynamics in the time-domain). Thus, the challenges of rolling shutter, motion blur, and incorrect light exposure are not present in event-based vision technology.

This event-based vision sensor technology is likely to be more suited as a visual motion sensation which has successfully been demonstrated in the related work of Scarramuzza and colleagues. An example of this is the related work of Gehrig *et al.* where an event-based vision was used with a machine learning approach to estimate optic flow [17]. Their result showed an improvement of 12% in the errors metrics compared to other state-of-the-art methods. Another example where intensity estimation (image reconstruction) and optic flow estimation were simultaneously computed from an event-based sensor was successfully demonstrated even during challenges of rapid motion and high dynamic range scenes [4].

Another strong argument for this new technology is that it enables an asynchronous data pipeline compared to the synchronous data flow. Less data has to be transferred and filter through the system and further only important data is being sensed at detection (an event). An example of traditional data in contrast to event-based, a Full-HD 1080p ($1920 \times 1080 \times 3$) at 144 Hz raw video stream uses a bandwidth of roughly $0.89\,\mathrm{GB\,s^{-1}}$ assuming an 8-bit channel. The bandwidth is in reality much less thanks to various video and image compression techniques at a cost of computational resource. Further a higher resolution of Ultra-HD 2160p ($3840 \times 2160 \times 3$) at same frame rate would be roughly $3.58\,\mathrm{GB\,s^{-1}}$, an increase fourfold. These magnitudes of data bandwidth pose significant challenges for data processing, both on software and hardware levels, which often negatively affect overall performance in robotics.

3.2 Robotic software

There are many challenges in designing a software architecture when considering robotics and further their applications. Some of these challenges are highlighted and addressed in this thesis. In many developed research robotic platforms, in particular, road vehicles, a pipes-and-filters-based data processing approach is often utilized. While this design architecture is a straight-forward and simpler method to implement, it may be argued that it quickly becomes complicated, vulnerable, and monolithic. One way to mitigate these shortfalls, as investigated and demonstrated in Paper C, is to adopt containerization and microservice design paradigm which was done with special consideration towards road vehicles. This was demonstrated in various vehicles of different scales with emphasis on high demand on reliability, traceability, and deployment.

3.2.1 Microservice design, real-time computing, and embedded system deadlines

According to Fowler, a microservice design paradigm may be described as *a suite of small services, each running in its own process and communicating with lightweight mechanisms.* To support such a paradigm, a middleware handling the communication and message specifications have been used as demonstrated in Paper C in the example of a fleet of vehicles. Due to the very design of the communication protocol, a microservice can be regarded as a program block with well-defined input and output signals. An implementation of this is clearly shown in Paper B where the defined inputs and outputs are sequential images and optic flow data respectively. By understanding this process, a structured and ease-of-use methodology was created where executable binaries were easily swapped when generating the results. Furthermore, since the microservice binaries can be packed in sufficient containers, they can be deployed on a wide range of host machines. This then creates run-time measurements that are normalized to the host machine thus producing valuable and comparable results (see Table I in Paper B).

Relating to the execution and run-time, when machines are programmed to perform a task, they are expected to provide a response before a time-constraint, or **a deadline**. Depending on the importance of the task, or severity of task outcome, deadlines are often classified as hard, firm, or soft where the severity is ranging from the harshest in the first, and to the least in the last. This is easily imagined in a critical situation where a subsystem is expected to perform a task, such as anti-lock braking wheels systems in road vehicles where fatality is a possible outcome. For such reason, embedded real-time systems are designed and constructed for executing programs that comply with set deadlines. Moreover, it is challenging to guarantee predictability or real-time computing as software systems grow larger and more advanced. Robotic locomotion should in general be considered safety-critical, so if the optic flow should be implemented in a robotic system. However, in Paper B it was found that no optic flow estimator is fully capable of real-time execution, nor suitable to be used in a live real-time robotic system at its current state.

It is important to note that commonly available personal computers with general-purpose operating systems are not real-time computing machines. These machines are primarily designed for adopting broader support for hardware, shared computing resources, and better user experience. For example, graphics processing units (GPUs) are designed to parallel-process data for graphics pipelines, which is often time-sensitive and is of significant importance for the user experience. For this reason, to maximize the utilization, the operating system kernel allows for variable and tighter computational timing for smoother graphics rendering in contrast to well defined fixed deadlines. However, it is possible to make personal computers comply with deadline computing through software techniques such as a customized and specialized scheduler or efficient locking protocols.

Optic flow estimators of today, like many other computer vision algorithms, typically require high-end features found in general-purpose computation nodes, such as *general-purpose computing on graphics processing units* (GPGPU). The maturity and suitability of using graphics processing units in real-time computing settings are improving and may be considered for soft deadline applications [13, 14]. However, specialized features like GPGPU often require closed-source user-space drivers to operate. Therefore, for future applications where optic flow should be further exploited for locomotion in robotics, new and more specialized estimator implementations are needed especially for real-time embedded systems for hard and firm deadlines.

3.3 Optic flow estimation computation

The neurocognition and behavioral aspects of optic flow have been considered in the previous Sect. 2. However, methods to numerically estimate the optic flow in a sequence of images has not been discussed. In this section, a very brief overview of optic flow estimation theory will be considered.

The related work of Raudies, Andreev, Wildes, and colleagues, detail a taxonomy overview of optic flow constraints that are commonly available and described in research literature [38, 2, 51]. Many of the optic flow estimation computation routines can according to their taxonomy and notations be generalized and described with the help of the well-known continuity equation from physics. Thus it may be described as

$$\partial_t \rho + \nabla_{xy} \cdot (\vec{v} \cdot \rho) = \sigma \tag{3.1}$$

where ρ is the density of some quantity, ∇_{xy} the divergence operator with respect to x, y, $\vec{v}$ the optic flow, and σ the generation or destruction of the quantity ρ.

The equation is kept to a generalized quantity ρ since different algorithms use a different quantity in their applications to estimate the optic flow $\vec{v}$. It is noted that the majority of optic flow estimations assumes that the quantity ρ is conserved thus simplifying the expression to

$$\partial_t \rho + \nabla_{xy} \cdot (\vec{v} \cdot \rho) = 0. \tag{3.2}$$

This form serves as a common stint for the more traditional naive optic flow estimations and further constraints further transform the equation to a more approachable and computationally possible form.

Optic flow constraint equation

Before detailing the optic flow constraint equation form, a few assumptions are made. Let the quantity be a scalar valued intensity quantity $\rho = g(x, y, t)$ and assume that $\nabla_{xy}\vec{v} = \vec{0}$ where x, y are spatial dimension variables for example in an image space. The assumption motivated by the argument that in a contained local pixel domain in an image, the brightness can not be accumulated nor lost caused by flux of brightness. If a *brightness constancy* assumption is made then

$$g(x + \Delta x, y + \Delta y, t + \Delta t) \approx g(x, y, t) \tag{3.3}$$

Further a linear approximation of the $g(x, y, t)$ in the local proximity would be

$$g(x + \Delta x, y + \Delta y, t + \Delta t) \approx g(x, y, t) + (\nabla_{xyt} g(x, y, t)) \cdot (\Delta x, \Delta y, \Delta t). \tag{3.4}$$

Combining these two equations results in

$$(\nabla_{xyt} g(x, y, t)) \cdot (\dot{x}, \dot{y}, 1) = 0, \tag{3.5}$$

which is a just further simplification of continuity equation, Eq. (3.2). This is the backbone theory of the more well-known Lucas-Kanade optic flow algorithm [27].

Hessian constraint

Similarly to the previous optic flow constraint equation, but using a different quantity $q = \nabla_{xy} \cdot f$ under the same assumption that $\nabla_{xy}\vec{v} = \vec{0}$ directly yields the Hessian constraint equation

$$\nabla \partial_t g(x, y, t) + H_g \vec{v} = 0 \tag{3.6}$$

where H_g is the Hessian matrix of the scalar valued function g. The Hessian constraint equation was developed by the work of Uras *et al.* [44].

Phase constancy constraint

Instead of using the brightness attribute in an image, Fleet and Jepson suggested using image phase attribute instead [16]. This changes the quantity $q = \phi$ and using the same assumption as previous constraint that $\nabla_{xy}\vec{v} = \vec{0}$ results in

$$(\dot{x}, \dot{y}, 1) \cdot \nabla_{xyt}\phi(x, y, t) = 0. \tag{3.7}$$

What makes this approach practical and feasible is that Gabor filters may be used to compute the image phase ϕ thus creating an alternative estimation of optic flow.

While there are more than 200 listed approaches that advance the estimation of optic flow and improves existing methods, most of the traditional estimators can be traced back to the continuity equation with the exception of end-to-end estimators such as deep neural network approaches. This creates challenges when attempting to gain an overview of the landscape of proposed optic flow estimators. However, optic flow estimation benchmarks and leaderboards exist and may provide a general overview of state-of-the-art methods (see Sect. 3.3.2 for data sets providers and leaderboard maintainers).

3.3.1 Machine learning approach

Since the introduction of **convolution neural networks** (CNN) in the 1980s, various fields have accelerated in computer vision research, and not the least, optic flow estimation research in the recent decade. CNN was proven to be very suitable for dealing with image interpretation such as shapes detection, object recognition, and motion perception. Before the inception of CNN in computer vision, using only deep neural networks was mostly ineffective in terms of computational resources and under-performed due to the number complexity of weight grew with the number of layers and neurons. With CNNs, space pixel neighborhoods can be processed and considered as a simultaneous input to the network layer. Furthermore, CNN is shift-invariant meaning that introduced translational or rotational displacement of the kernel can help the filter to generalize learned patterns.

In recent years, one can see that the machine learning approach has dominated in the benchmarking of optic flow and many cases outperforming the more traditional optic flow estimation algorithms in both accuracy, precision, and run-time. This can be seen in the popular datasets benchmark leaderboards as mentioned and discussed in Paper B and the further section below (Sect. 3.3.2).

Considering the development of application specialized hardware such as GPUs as which excels in parallel data computing and processing. Despite being named and associated with graphics computing, the application may be extended further to general-purpose computing on graphics processing units and even real-time computing (see Sect. 3.2.1).

Due to the very nature of deep neural networks, how it solves the estimation problem is per design an end-to-end system, i.e. black box. For this very reason, deep learning is not a core topic of this thesis, but an acknowledged field in computer vision.

3.3.2 Benchmarking performance and evaluation

It is common in the computer vision community to compare their proposed algorithms and estimators. In the case of optic flow, one usually adopts the common metrics of *average end-point error* (AEE) and *average angular error* (AAE). The former error is calculated using euclidean distance between the estimated flow and the ground truth

$$E_{AEE} = \frac{1}{N}\sum_{x,y} E_{endpoint}(x,y) = \frac{1}{N}\sum_{x,y} |\vec{Q}_{estimated}(x,y) - \vec{Q}_{groundtruth}(x,y)| \tag{3.8}$$

while the latter is calculated angle of the two flow vectors

$$E_{AAE} = \frac{1}{N}\sum_{x,y} E_{angle}(x,y) = \frac{1}{N}\sum_{x,y} \arccos \frac{\vec{Q}_{estimated}(x,y) \cdot \vec{Q}_{groundtruth}(x,y)}{|\vec{Q}_{estimated}(x,y)||\vec{Q}_{groundtruth}(x,y)|} \tag{3.9}$$

where N is the number of sample points i.e. usually the image resolution. These metrics were first used in the context of performance evaluation of optic flow techniques by Barron *et al.* [5] and has nevertheless become *de facto* standard to use when evaluating optic flow estimators. Since then there has been several research groups and communities who provide different types of data sets for open benchmarking such as KITTI [18, 29, 30], Middlebury [3], Sintel [10], and the robust vision challenge [1].

As mentioned in Paper B, it is a challenge to perform a fair run-time evaluation due to the complexity of executing binaries and their interaction with hardware. Thus, it is a challenge to perform a fair and comparable run-time evaluation in research. Traditionally, computational complexity is used to classify different software methods but doing so for software for run-time execution or real-time computing, in general, would be proven challenging and arguably meaningless.

Nevertheless, from observing the general trends in the optic flow estimation leaderboards, it may be noticed that estimators are improving both accuracy and run-time, and even the trade-off between the two. Moreover, as mentioned above, machine learning approaches are dominating among the top-performing methods, further not only limited to the optic flow category.

Chapter 4

Discussion

This thesis focuses mainly on retinal optic flow in perception and considers it in a navigational context for robotic applications. The emphasis is especially on visual perception: how it may be estimated, visualized, quantified, and used for navigation. In this chapter, the aspects of impact and challenges regarding retinal optic flow will be presented but also if the technology is matured enough to embrace modern software in road vehicles.

4.1 During active fixation and circular steering motion

A proof of concept of modeling retinal optic flow field by combining gazing behavior and optic flow was demonstrated in Paper A. The flow field is one of the cornerstones which lays the foundation for the prominent works of Wann, Lands, Swapp, and Wilkie and many other similar strategies. The model is shown through using test subjects in an experiment, designed similarly as in other related work [43, 23].

During the analysis in Paper A, it became apparent that the accuracy of reconstructing retinal optic flow would heavily depend on the optic flow estimator. The method of estimating the optic flow field had to be carefully selected and properly evaluated. The complex environment of independently moving objects, challenging high dynamic range in the scene, and rapid motion caused by head rotation, significantly impact the accuracy performance of optic flow estimation. As discussed and highlighted in Paper B, there is a clear relation between computational resource and performance accuracy, resulting in longer computational time giving better accuracy results. This, however, poses a challenge when perception is part of a time-sensitive critical system such as steering control. One possible solution to mitigate this is to use dynamic vision sensors for estimating the optic flow fields instead since it does not suffer from motion blur or challenging outdoor scenes with high dynamic range light.

On the gaze analysis, naive and simple classification attempts were made to identify active fixations. The classification proved to be challenging without going into in-depth statistical approaches. Retinal optic flow estimation would greatly improve if the identification of active fixation was properly implemented. In our reconstruction of retinal optic flow, active fixation was naively assumed during non-blinks and low confidence level gaze data. Existing eye gazing classification methods have been demonstrated in literature [34] which could be utilized in this type of research. As Pekkanen *et al.* pointed out in their work, passive and active fixations are generally hard to distinguish from each other. This could potentially present difficulties for the method of reconstruction assumes the presence of angular velocity of the eye. One way to mitigate this is to redirect our attention back to what the agent is looking at and perform analysis on the likelihood of passive or active fixation.

Although Wann and Swapp showed the *nulling*-strategy is mathematically plausible to be exploited, there are caveats to their locomotor control strategy proposal. One is that the agent is assumed to have an active fixation that is *parallel* to the ground-plane, i.e. gazing to the horizon. While this has been observed occurring in Paper A, active fixations are occasionally made on visual details, like road markings or objects thus slightly deviating from the mathematical construction. By slightly shifting the gaze, for example downwards, adds a non-trivial contribution to retinal optic flow thus altering the Eq. (2.3). This contribution can most likely be neglected due to its magnitude assuming in the proximity of the horizon, but further extensions to the mathematical description can be made to confirm this claim.

Another caveat is that the circular steering motion assumption is imposed on the locomotion. While this might be suited for high-velocity locomotion and road vehicles, this is not the case during relatively slow walking gaits. Matthis and Fajen investigated the interaction between gazing and foot placement during navigation in natural and complex terrain [28]. They concluded that humans look downwards, tuning their gazing strategies to the environment they navigate in while maintaining a constant temporal look-ahead window. In contrast to this, and nulling the flow curve strategy requires noticeable motion to generate flow, and circular motion which is not the case of careful and precise walking gaits. Further investigation of how retinal optic flow is exploited in the case of walking gaits with active foot placement.

4.2 Non-steering-related fixation

So far, the retinal optic flow has mostly been considered during steering tasks in this thesis. However, as discovered in Paper A and further reported and concluded by Lehtonen *et al.* [25], all gazing do not always lead to steering. In Paper A where the subjects were tasked to keep the desired speed in addition to the steering task in the track. This resulted in the subjects occasionally shifting their gaze to the speedometer. This has a significant implication of retinal flow optic flow, specifically the *nulling flow curvature*-strategy proposed by Wann *et al.* which requires the agent to look towards the intended path.

This leads us to the question, *what purpose do the retinal optic flow serve when not explicitly used in navigation?*. One possible explanation is that extra-retinal information is continuously and subconsciously extracted to match expected the internalized model which is in line with the internal model hypothesis. This hypothesis argues that obtaining experience and learning results in predictions and anticipation. This is motivated by the work of Tuhkanen *et al.* [43] where their subjects developed an internal model of predicted waypoints during steering tasks resulting in them gazing into the empty space.

4.3 Steering control and modeling

The thesis argues that locomotor control and scene perception are tightly coupled, as shown in Paper A. From the study, one could clearly see that an individual control correction was a direct consequence of gaze. In addition, it was found that sequential gazing directly formed overlapping, i.e. superpositioned, corrections with accordance with the original theory [6]. Moreover, further evidence of the open-loop characteristics of steering corrections was found *after* saccades shortly after correction onset, even in connection with secondary tasks.

The early steering models of Swapp *et al.* and Warren *et al.* (see Sect.2.2.2) fail to capture one of the fundamental property of human control and movements, namely the distinctive motion pattern formed by the antagonistic muscle pairs, in static contexts referred to as *reaching*. Benderius and Markkula were the first to discover and report this kind of human behavior in human steering by studying steering data from naive subjects [6]. By proposing a super-position of bell-shaped based correction rates, the steering behaviors could be more accurately explained. A formal mathematical steering model to incorporate retinal optic flow perception and intermittent control has yet to be proposed.

4.4 Application in live robotic road vehicles

In combined findings of Paper B and Paper C, it infers positive results to facilitate and enable modern software development in robotic road vehicles when adopting containerization and microservice design paradigm. This trend has allowed for state-of-the-art research, development, and technology to be rolled out on the road as opposed to experimental methods in simulations or virtual environments.

In Paper B, it was shown that near or close to real-time computing with optic flow estimators is possible with the exception of some run-time outliers using a software scheduler with the emphasis on run-time consistency and fast computation. The estimators were further implemented with a containerized approach which demonstrates the modularity and flexibility of such a deployment strategy. Such development is preferable for continuous experimentation during a development phase and more importantly retrospective analysis for a fair comparison and traceability in deployed systems. So far, no dense optic flow estimator has been formally proven and demonstrated to exhibit real-time capabilities with sufficient accuracy.

In Paper C, it was concluded that modern software development using containerized development in road vehicles is possible despite having different various types of hardware architectures and platforms ranging from miniature vehicles, racing cars, heavy trucks, passenger cars, and marine vehicles. Various computational resources were abstracted and generalized, with an attempt to increase the separation between software and hardware.

By making careful and conscious design choices of hardware and software systems for robotic vehicles, it possible to create a real-world application using bionic components, for example, retinal optic flow perception for driver models. It remains to be seen if the computational performance of a bionic system can yield acceptable driving behavior compared to its classical control theory counterpart.

Chapter 5

Concluding remarks and future work

5.1 Conclusion

The main conclusion from this thesis that can be drawn from this thesis is that the retinal optic flow field can be computationally reconstructed using experimental data from human subjects. The model complies with previous research under the correct conditions imposed by optic flow estimation and gazing behavior. This was shown by analyzing data from two test subjects during repeated vehicular steering task.

In Paper A, it was concluded that the behavioral patterns as predicted by intermittent control is formed from interconnection with gazing and retinal optic flow. Through collecting vehicle and gaze data from test subjects during a steering task, the retinal optic flow was reconstructed and visualized and further analysis of the flow provided a perceptual flow path. This retinal optic flow path is a central concept and the main component of *nulling flow curvature*-strategy proposed by several researchers.

To support the reconstruction made in Paper A, an investigation of various optic flow estimators was conducted in Paper B. It was successfully demonstrated that containerization and microservice design, aid in the reproducibility of research findings in computational tasks. Furthermore, the methodology provides more valuable and comparable results by normalizing run-time metrics to one single host machine. During the generation of the results, the real-time capable system kernel was enabled on the system, providing key insight analysis regarding real-time computing and embedded system analysis. All of the results were generated in the example of optic flow estimation using various algorithms.

As one of the envisioned milestones is to deploy our work in a live robotic system, Paper B and Paper C combined concluded that adopting container technology and microservice design mitigates the challenges of transition from archaic system architecture to a modern one. The introduction of such workflow and design paradigm enables continuous experimentation, continuous deployment, and continuous integration which further support robotic research and development.

5.2 Future work

While the proof of concept in constructing the retinal optic flow field is based on sampled experimental data, there are still some aspects left to be investigated and further improved. Early works of estimating optic flow using dynamic cameras demonstrate promising accuracy, reliability, and nonetheless faster data acquisition. These aspects make the event-based camera superior when estimating optic flow, and further extended retinal optic flow estimation. Due to the lesser amount of data to process and fundamentally different data flow, it could enable retinal optic flow field estimation to be implemented in a live real-time robotic system.

As found in Paper A, all gazing does not always use for locomotor control thus further investigation needs to be done to identify and interpret intentional gazing, i.e. why we look on the road and sometimes not. Furthermore, it has been reported and observed both in this thesis and by other researchers that steering corrections without gazing directly on the road is an occurring phenomenon in human drivers. Future research remains to investigate how this is possible, and further if retinal optic flow plays a significant role in the peripheral perception in this particular scenario.

To complete the task of navigation, closing the loop of locomotor control, the missing pieces of control action and motion planning must be considered. Early analysis in Paper A shows that intermittent control for steering correction could explain observed human driving behavior. This could be exploited when designing a human-like navigating agent in robotics. When the agent would be implemented in an autonomous passenger road vehicle, a test similar to the Turing test determining whether a human passenger would be able to tell the difference between an automated driver from a human one could be of interest.

Another future topic is to investigate how one can generate artificial retinal optic flow further implying that gazing behavior also has to be artificially constructed as well. This would be a critical step towards a designing robotic navigation application that exploits the bio-inspired retinal optic flow. One possible road map to achieve this is to perform road segmentation and motion planning creating an ideal path. From this ideal path, a gaze point should be placed towards the path, and *nulling the flow curvature with active fixation*-strategy could be transferred to the robotic domain for steering control.

Finally, and as suggested by the outline of this thesis, the retinal optic flow remains to be implemented in a live robotic navigation system. Validation and comparison of such a system could bring a valuable analysis of the existing automated driving system. Further work will be to improve the current estimation of retinal optic flow and artificially create gazing behavior to support this kind of bionic subsystem.

Chapter 6

Summary of included papers

This thesis consists of three papers that investigate the human visual perception in navigation for robotic purposes. By integrating human gazing and optic flow, retinal optic flow can be obtained and numerically quantified and depicted which has been a challenge due to technological and scene data constraints. Paper A demonstrates a proof of concept of estimating a bio-inspired retinal optic flow using a conventional RGB camera and eye-tracking technology on human test subjects during road vehicle steering tasks. In Paper B, various optic flow estimators were evaluated and benchmarked using a real-time scheduler and containerization methodology. Paper C showcases the technical challenges, the best practices, and experiences of adopting microservice design paradigm with containerization for the software development and deployment on the vehicle fleet of ranging scales and various purposes in Chalmers' vehicle laboratory.

6.1 Paper A

The paper as mentioned above demonstrates the proof of concept of numerically estimating and reconstructing the retinal optic flow using data from passive vision sensors and live test subjects. This reconstruction attempts to quantify and mimic human perception of motion. Furthermore, retinal optic flow path where downward flow relative to the agent's gaze is detected and may be used as perception data for steering control tasks as theorized in *nulling flow curvature with active fixation*-strategy. The work further shows intermittent control as highly interconnected with active gazing and retinal optic flow, and an observed complex navigation behavior pattern emerges. The reconstruction of the retinal optic flow field as done by fusing optic flow field estimation and gazing behavior under the assumption that active fixation was in effect during non-saccadic gazing was in play.

6.2 Paper B

Paper B addresses an evident challenge of reproducing research findings in scientific papers, especially in claimed run-time in performance benchmarks without any normalization of the computational resources used or retrospective performance analysis. By adopting containerization as shown in Paper C, it was demonstrated in the example of optic flow estimation that a fair and reproducible comparison can be made, furthermore aiding the process of replicating the research results. The evaluations in the paper were based on computation using a real-time capable system. By doing so, the run-time results are normalized to one single system which may provide more insight into the comparison. Moreover, one can observe that the performance uplift both on error metrics or in run-time metrics (see Sect. 3.3.2) compared to the claimed ones which are expected due to technological advancement in both dependency libraries and computational resources. Using this methodology, it is possible to discuss the trade-off between computational performance and cost, and further compare this to other algorithms.

6.3 Paper C

The paper details the accumulated experience, best practices, and pitfalls of using containerized software with the microservice design paradigm for self-driving vehicles. This methodology of creating a microservice, to then package it into a self-contained software bundle, with all necessary dependency libraries, allows for clear binary traceability, software modularity, and easy deployment. Another motivating purpose was to quickly introduce new researchers and students to the hardware and software, allowing them to focus on their goals and problems. In this way, a decoupling of computational resources from the algorithms to a dynamic restructuring of data processing pipelines could be made. This allows for more flexible and interchangeable software blocks resulting in convenient ways to introduce continuous integration, continuous deployment, and continuous experimentation. As presented in the paper, one implementation of these design choices, OpenDLV, includes a standard message set, an *a priori* established communication protocol among microservices.

References

[1] `http://robustvision.net/`. [Online; accessed 23-January-2021].

[2] R. Andreev, O. Scherzer, and W. Zulehner. Simultaneous optical flow and source estimation: Space–time discretization and preconditioning. *Applied Numerical Mathematics* **96** (2015), 72–81.

[3] S. Baker et al. A database and evaluation methodology for optical flow. *International journal of computer vision* **92**.1 (2011), 1–31.

[4] P. Bardow, A. J. Davison, and S. Leutenegger. "Simultaneous optical flow and intensity estimation from an event camera". *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2016, pp. 884–892.

[5] J. L. Barron, D. J. Fleet, and S. S. Beauchemin. Performance of optical flow techniques. *International journal of computer vision* **12**.1 (1994), 43–77.

[6] O. Benderius and G. Markkula. "Evidence for a fundamental property of steering". *Proceedings of the human factors and ergonomics society annual meeting*. Vol. 58. 1. SAGE Publications Sage CA: Los Angeles, CA. 2014, pp. 884–888.

[8] J. D. A. Bermudez. "Real-time Visual Flow Algorithms for Robotic Applications". PhD thesis. The Australian National University (Australia), 2017.

[9] C. Blakemore and G. F. Cooper. Development of the brain depends on the visual environment. *Nature* **228**.5270 (1970), 477–478.

[10] D. J. Butler et al. "A naturalistic open source movie for optical flow evaluation". *European Conf. on Computer Vision (ECCV)*. Ed. by A. Fitzgibbon et al. (Eds.) Part IV, LNCS 7577. Springer-Verlag, Oct. 2012, pp. 611–625.

[11] D. Calow and M. Lappe. Efficient encoding of natural optic flow. *Network: Computation in Neural Systems* **19**.3 (2008), 183–212.

[12] S. A. Cooper et al. Akinetopsia: acute presentation and evidence for persisting defects in motion vision. *Journal of Neurology, Neurosurgery & Psychiatry* **83**.2 (2012), 229–230.

[13] G. A. Elliott and J. H. Anderson. Globally scheduled real-time multiprocessor systems with GPUs. *Real-Time Systems* **48**.1 (2012), 34–74.

[14] G. A. Elliott and J. H. Anderson. "Robust real-time multiprocessor interrupt handling motivated by GPUs". *2012 24th Euromicro Conference on Real-Time Systems*. IEEE. 2012, pp. 267–276.

[15] T. E. Feinberg and J. Mallatt. The evolutionary and genetic origins of consciousness in the Cambrian Period over 500 million years ago. *Frontiers in psychology* **4** (2013), 667.

[16] D. J. Fleet and A. D. Jepson. Computation of component image velocity from local phase information. *International journal of computer vision* **5**.1 (1990), 77–104.

[17] D. Gehrig et al. "End-to-end learning of representations for asynchronous event-based data". *Proceedings of the IEEE International Conference on Computer Vision*. 2019, pp. 5633–5643.

[18] A. Geiger, P. Lenz, and R. Urtasun. "Are we ready for Autonomous Driving? The KITTI Vision Benchmark Suite". *Conference on Computer Vision and Pattern Recognition (CVPR)*. 2012.

[19] J. J. Gibson. The perception of the visual world. (1950).

[20] R. Held et al. The newly sighted fail to match seen with felt. *Nature neuroscience* **14**.5 (2011), 551–553.

[21] J. M. Henderson. Human gaze control during real-world scene perception. *Trends in cognitive sciences* **7**.11 (2003), 498–504.

[22] J. Heutink et al. The effect of target speed on perception of visual motion direction in a patient with akinetopsia. *Cortex* **119** (2019), 511–518.

[23] O. Lappi et al. Humans use optokinetic eye Movements to track Waypoints for Steering. *Scientific reports* **10**.1 (2020), 1–14.

[24] O. W. Layton, E. Mingolla, and N. A. Browning. A motion pooling model of visually guided navigation explains human behavior in the presence of independently moving objects. *Journal of vision* **12**.1 (2012), 20–20.

[25] E. Lehtonen et al. Gaze doesn't always lead steering. *Accident Analysis & Prevention* **121** (2018), 268–278.

[26] H. C. Longuet-Higgins and K. Prazdny. The interpretation of a moving retinal image. *Proceedings of the Royal Society of London. Series B. Biological Sciences* **208**.1173 (1980), 385–397.

[27] B. D. Lucas, T. Kanade, et al. An iterative image registration technique with an application to stereo vision (1981).

[28] J. S. Matthis, J. L. Yates, and M. M. Hayhoe. Gaze and the control of foot placement when walking in natural terrain. *Current Biology* **28**.8 (2018), 1224–1233.

[29] M. Menze, C. Heipke, and A. Geiger. "Joint 3D Estimation of Vehicles and Scene Flow". *ISPRS Workshop on Image Sequence Analysis (ISA)*. 2015.

[30] M. Menze, C. Heipke, and A. Geiger. Object Scene Flow. *ISPRS Journal of Photogrammetry and Remote Sensing (JPRS)* (2018).

[33] F. Ovsiew. The Zeitraffer phenomenon, akinetopsia, and the visual perception of speed of motion: a case report. *Neurocase* **20**.3 (2014), 269–272.

[34] J. Pekkanen and O. Lappi. A new and general approach to signal denoising and eye movement classification based on segmented linear regression. *Scientific reports* **7**.1 (2017), 1–13.

[35] J. Pekkanen et al. A computational model for driver's cognitive state, visual perception and intermittent attention in a distracted car following task. *Royal Society open science* **5**.9 (2018), 180194.

[36] K. Prazdny. Egomotion and relative depth map from optical flow. *Biological cybernetics* **36**.2 (1980), 87–102.

[37] K. Prazdny. On the information in optical flows. *Computer Vision, Graphics, and Image Processing* **22**.2 (1983), 239–259.

[38] F. Raudies. Optic flow. *Scholarpedia* **8**.7 (2013), 30724.

[39] C. Rudin. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence* **1**.5 (2019), 206–215.

[40] P. Salvini, C. Laschi, and P. Dario. Design for acceptability: improving robots' coexistence in human society. *International journal of social robotics* **2**.4 (2010), 451–460.

[41] D. D. Salvucci and J. H. Goldberg. "Identifying fixations and saccades in eye-tracking protocols". *Proceedings of the 2000 symposium on Eye tracking research & applications*. 2000, pp. 71–78.

[42] C. L. Smith, N. Pivovarova, and T. S. Reese. Coordinated feeding behavior in Trichoplax, an animal without synapses. *PloS one* **10**.9 (2015), e0136098.

[43] S. Tuhkanen et al. Humans use predictive gaze strategies to target waypoints for steering. *Scientific reports* **9**.1 (2019), 1–18.

[44] S. Uras et al. A computational approach to motion perception. *Biological Cybernetics* **60**.2 (1988), 79–87.

[45] S. Vedula et al. Three-dimensional scene flow. *IEEE transactions on pattern analysis and machine intelligence* **27**.3 (2005), 475–480.

[46] J. Wann and M. Land. Steering with or without the flow: is the retrieval of heading necessary? *Trends in cognitive sciences* **4**.8 (2000), 319–324.

[47] J. P. Wann and D. K. Swapp. *Flow vectors on the image plane when fixating on a future circular path*. `https://static-content.springer.com/esm/art%3A10.1038%2F76602/MediaObjects/41593_2000_BFnn0700_647_MOESM1_ESM.pdf`. [Online; accessed 23-January-2021]. 2000.

[48] J. P. Wann and D. K. Swapp. Why you should look where you are going. *Nature neuroscience* **3**.7 (2000), 647–648.

[49] W. H. Warren and B. R. Fajen. "From optic flow to laws of control". *Optic flow and beyond*. Springer, 2004, pp. 307–337.

[50] W. H. Warren et al. Optic flow is used to control human walking. *Nature neuroscience* **4**.2 (2001), 213–216.

[51] R. P. Wildes et al. Recovering estimates of fluid flow from image sequence data. *Computer Vision and Image Understanding* **80**.2 (2000), 246–266.

[52] R. M. Wilkie and J. P. Wann. Eye-movements aid the control of locomotion. *Journal of vision* **3**.11 (2003), 3–3.

[53] R. M. Wilkie, J. P. Wann, and R. S. Allison. Active gaze, visual look-ahead, and locomotor control. *Journal of experimental psychology: Human perception and performance* **34**.5 (2008), 1150.