

Visual Navigation Using Sparse Optical Flow and Time-to-Transit

*Original*

Visual Navigation Using Sparse Optical Flow and Time-to-Transit / Boretti, Chiara; Bich, Philippe; Zhang, Yanyu; Baillieul, John. - ELETTRONICO. - (2022), pp. 9397-9403. (Intervento presentato al convegno 2022 International Conference on Robotics and Automation (ICRA) tenutosi a Philadelphia, PA, USA nel 23-27 May 2022) [10.1109/ICRA46639.2022.9812032].

*Availability:*

This version is available at: 11583/2970216 since: 2022-07-21T12:39:21Z

*Publisher:*

IEEE

*Published*

DOI:10.1109/ICRA46639.2022.9812032

*Terms of use:*

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

*Publisher copyright*

IEEE postprint/Author's Accepted Manuscript

©2022 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collecting works, for resale or lists, or reuse of any copyrighted component of this work in other works.

(Article begins on next page)

# Visual Navigation Using Sparse Optical Flow and *Time-to-Transit*

Chiara Boretti, Philippe Bich, Yanyu Zhang, and John Baillieul

**Abstract**—Drawing inspiration from biology, we describe the way in which visual sensing with a monocular camera can provide a reliable signal for navigation of mobile robots. The work takes inspiration from the classic paper [3] which described a behavioral strategy pursued by diving sea birds based on a visual cue called *time-to-contact*. A closely related concept of *time-to-transit*,  $\tau$ , is defined, and it is shown that steering laws based on monocular camera perceptions of  $\tau$  can reliably steer a mobile vehicle. The contribution of the paper is two-fold. It provides a simple theory of robust vision-based steering control. It goes on to show how the theory guides the implementation of robust visual navigation using ROS-Gazebo simulations as well as deployment and experiments with a camera-equipped Jackal robot. As will be noted, there is an extensive literature on how animals use optical flow to guide their movements. The novelty of the work below is the introduction of the concepts of *Eulerian optical flow* and *time-to-transit*,  $\tau$  and the demonstration that control laws based on the  $\tau$ -values associated with an aggregated set of features in the field of view can be used to reliably steer a laboratory robot.

**Keywords:** Time-to-transit, Eulerian optical flow, Lagrangian optical flow, vision-based navigation

## I. INTRODUCTION

The terminology *time-to-contact* and the notation  $\tau$  appear to have originated among researchers working in the Cornell University lab of James J. Gibson

---

C.B. and P.B. are with Politecnico di Torino, Department of Electronics and Telecommunications (DET). Y.Z. and J.B. are with the College of Engineering at Boston University. Support from various sources including the U.S. Office of Naval Research grants N00014-10-1-0952, N00014-17-1-2075, and N00014-19-1-2571 is gratefully acknowledged. The authors are also grateful for support from Politecnico di Torino, DET. The authors would like to thank Roy Xing for his precious help with the robot experiments at Boston University.

in the 1970's and 1980's ([1],[2],[3]), and it has been suggested that it may play a role in goal-directed navigation, [6]. Since then,  $\tau$  has been a significant focus of research in perceptual psychology (e.g. [4],[5]). The basic idea is simple. When an observer is approaching an object at a constant velocity, the image of the object on the observer's retina (or camera's image plane) is an expanding "bundle" of vectors, and if we take any particular image point  $r(t)$  (in image plane coordinates) and differentiate with respect to time, then as noted in [3],  $r/\dot{r}$  is equal to  $\tau$ —the *time-to-contact* or *time-to-collision*. In order to explore how  $\tau$  might be used as a navigation signal, we recently introduced a generalization called *time-to-transit* ([7],[8]). To understand how time-to-transit can be perceived and used to guide navigation, we consider the geometry of planar motion of a camera-equipped unicycle vehicle relative to a feature point in the environment. The kinematics of the vehicle are:

$$\begin{pmatrix} \dot{x} \\ \dot{y} \\ \dot{\theta} \end{pmatrix} = \begin{pmatrix} v \cos \theta \\ v \sin \theta \\ u \end{pmatrix}, \quad (1)$$

where  $v$  is the forward speed in the direction of the body-frame  $x$ -axis, and  $u$  is the turning rate.

*Definition 1:* Consider a feature point with coordinates  $(x_f, y_f)$  and a vehicle whose body-frame configuration is as depicted in Fig. 1. Given the current configuration  $(x(t), y(t), \theta(t))$  and unicycle kinematics (1), the *geometric time-to-transit* (*geometric*  $\tau$ ) is the time it would take the vehicle to cross a line intersecting the feature and perpendicular to the current heading under the assumption that speed  $v$  and heading  $\theta$  are held constant.  $\square$

Simple plane geometry shows that geometric  $\tau$ , defined in this way, is given by

$$\tau(t) = \frac{\cos \theta(x_f - x(t)) + \sin \theta(y_f - y(t))}{v}. \quad (2)$$

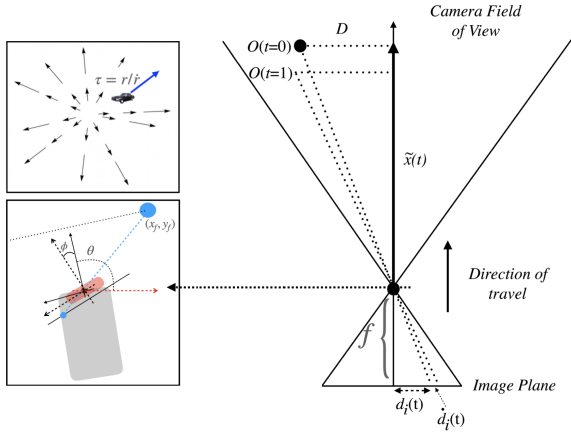


Fig. 1: When  $\phi = 0$ ,  $\tau$  is a proxy for distance or depth. If the forward-looking camera is directly abeam of the feature point to its right or left, the value of  $\tau$  is zero; it is the instant at which the transit occurs. If  $\phi \neq 0$ , then  $\tau_{per} \neq \tau_{geom}$ , but for  $|\phi|$  small and constant, the difference in these  $\tau$ -values is modest. In the top inset, constant speed forward motion of a forward-looking camera results in an expanding bundle of velocity vectors produced by the motions of feature images. The quantities  $r/\dot{r}$  in the inset and  $d_i/\dot{d}_i$  in the large figure, are the perceived *time-to-transit*.

Time-to-transit,  $\tau$ , is of interest as a navigation signal because under the assumption of constant speed  $v$  and heading  $\theta$ , this geometric value of  $\tau$  can be perceived on the image plane or retina as  $r/\dot{r}$ , which in the case of a forward-looking camera, is just  $\tau = d_i(t)/\dot{d}_i(t)$  as depicted in Fig. 1. Because equality only holds under the assumptions that  $\theta$  and  $v$  are constant and the camera is forward facing with the optical axis being aligned with the direction of motion, we call this value *perceived time-to-transit* and denote it by  $\tau_{per}$ . When these ideal assumptions (constant speed, heading, and camera angle) don't hold,  $\tau_{per}$  is distorted relative to  $\tau_{geom}$ . The extent of the distortion caused by the camera's misalignment with the vehicle heading can be illustrated by having the vehicle (1) follow a straight line path along the positive  $y$ -axis— $(x(t), y(t), \theta) = (0, t, \pi/2)$ . We assume that the feature point  $(x_f, y_f)$  (1) is located such that  $x_f > 0$ ,  $y_f > 0$ , and that  $v = 1$ . It can be shown that for this

motion<sup>1</sup>,

$$\tau_{per}(t) = y_f - t + \frac{\phi}{x_f}(y_f - t)^2 - \phi x_f + o(\phi). \quad (3)$$

When  $\phi = 0$ , this value yields the linear decrease  $\tau_{per}(t) = y_f - t$ , which is equal to  $\tau_{geom}$ . When  $\phi \neq 0$ , the expression for  $\tau_{per}(t)$  involves a second order term, whose effect, as illustrated in (3) is small if  $\phi$  has small magnitude. The value of  $\tau_{per}$  will be more severely distorted if the assumption of constant heading is relaxed, and this will be discussed in detail in the next section.

The remainder of the paper is organized as follows. In the next section, we briefly describe how a highly idealized model of optical flow and time-to-transit can provide a robust navigation signal. We briefly discuss factors that confound visual signals in implementations, and by considering how perceived values of  $\tau$  become distorted from the ideal, we are led to mitigation strategies that make  $\tau$ -based steering feasible. Implementations using resources from ROS, Gazebo, and OpenCV are discussed in Section II. Selected details of a theory of  $\tau$ -based motion control are presented in Section III, and experiments involving both simulations and tests with a Jackal robot are presented. The paper concludes with a discussion of future work involving experimental exploration of ideas motivated by the neurophysiology of animal movement.

## II. IMPLEMENTING A SENSE-ACT CYCLE

Our previous work on  $\tau$ -based navigation introduced an idealized model called *Eulerian optical flow*, which by analogy with the Eulerian description of fluid flow, assumes that feature images stream continuously across all photo receptors such that values of  $\tau$  at each photo receptor are continuously available, [9]. A number of  $\tau$ -based steering laws in the Eulerian setting have been shown to produce robustly stable steering in a variety of settings. In order to obtain reliable visual information in actual implementations using *perceived time-to-transit* relative to features in the environment, a less idealized

<sup>1</sup>More details can be found in the extended version of this paper at <https://arxiv.org/abs/2111.09669>.

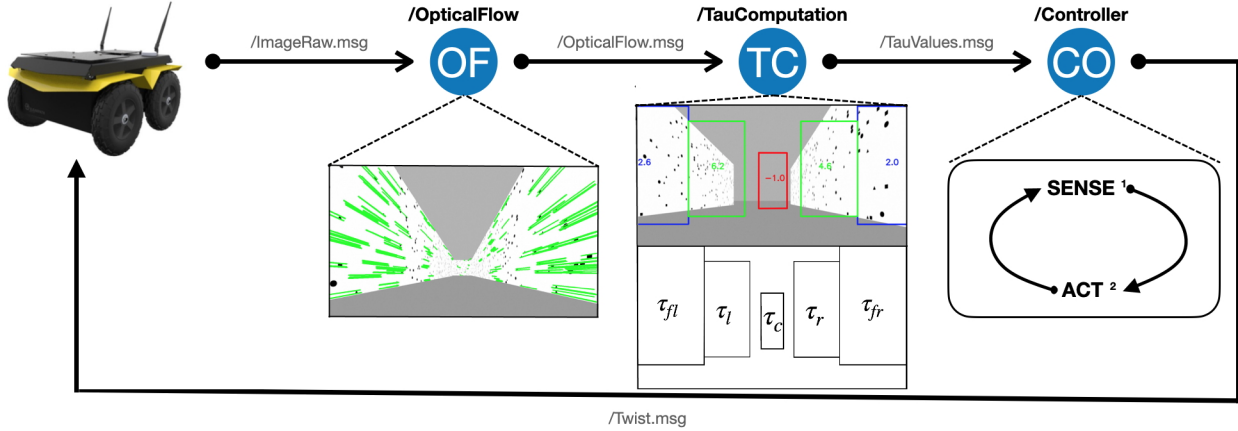


Fig. 2: Graphical representation of the ROS framework used in this paper. Navigation is accomplished by message passing among nodes in the system ROS graph.

*Lagrangian Optical Flow* (LOF) model is used, ([9]). Since one of the goals of this paper is to develop a navigation strategy capable of running in real-time on a robotic platform, we have created software that implements our navigation laws using ROS/Gazebo together with several well-known Optical Flow estimation techniques.

The Lucas-Kanade method ([10]) allows the computation of a sparse OF field (this means that OF vectors are estimated just for selected features in the image) by optimizing an energy-like function at points in the image that have been determined by a feature detector. We choose this technique because it is fast and computationally efficient. Since, by its nature, it works well for small movements but it fails when large pixel motion occurs, in our implementation it is used in its pyramidal form ([11]). In Fig. 2 an outline of the ROS framework developed in this work is proposed with snapshots of the functionalities implemented in each node.

The Optical Flow node is responsible for the OF estimation. It acquires a sequence of images from the camera mounted on the robot and it extracts the relevant features to finally compute the optical flow vectors. The minimum quality and the maximum number of features to retain are parameters provided to the detector.

The goal of the *tau computation* node in Fig. 2 is to analyze the array of keypoints with their velocities

packed in the OF message, to compute  $\tau$  values and to create the input signals for the controller. We transform the data coming from the Optical Flow node in order to make them coherent with a reference frame that has its origin at the intersection of the camera axis and the image plane. In this position, the origin of the coordinate frame is approximately at the Focus of Expansion (FOE), and since the precise estimation of the FOE is computationally expensive and sensitive to noise, and since our approximation has worked well in practice, we have used the camera-axis frame to determine  $\tau$ .

Once all the data are expressed in the right reference frame, time-to-transit is computed as  $\tau_i = \sqrt{x[i]^2 + y[i]^2} / \sqrt{v_x[i]^2 + v_y[i]^2}$ , where  $x$ ,  $y$ ,  $v_x$  and  $v_y$  are the arrays packed in the OF message. We define a fixed-size set of inputs that are given to the controller with the goal of providing a good representation of the environment without over burdening the controller with computational effort.

To obtain more robust  $\tau_{per}$  data we choose to divide the input image into five Regions of Interest (ROIs), as depicted in Fig. 2, which will output a single value each. The number and the dimension of these regions can be adapted to the specific environment in which we want to use the robot and the motion primitives that will be presented can be used regardless the number of regions.

Since the distribution of  $\tau$  values is strongly related

to the environment in which the robot is moving we can use it to select the right control law with the correct gains to be used. When moving in a straight corridor the tau values are continuous in time and they generally come from every region except the central ROI. The presence of a ninety degree turn, say at the end of a corridor, can be detected by a discontinuity in  $\tau$  values, a suddenly unbalanced OF field, and small  $\tau$  values in the central part of the FOV.

As noted in the introduction, camera rotation caused by a time-varying heading of the vehicle can cause large distortion of the estimate of  $\tau$ . To make  $\tau_{per} \approx \tau_{geom}$  a sense-act segmentation is implemented, which consists in segmenting the motion into alternating straight (sensing) and curved (acting) path segments. Care must of course be taken to ensure that neither portion is sustained for too long. An instantaneous sensing is possible only in an ideal scenario. In a real environment the sense phase must be long enough to obtain a good estimation of  $\tau$  values, but sensing for too much time means not being able to detect instantaneous variations of the OF field potentially losing important information coming from the environment. Moreover, the sense-act interleaving must guarantee the stability of the control law applied in the acting phase, and following a mathematical argument along the lines of [9] (the arXiv version), it can be shown that if both the cycle length and act phase are of sufficiently short duration, the robot will stably and reliably move through the corridor—always tending to center itself. This is checked with Matlab simulations assuming *Eulerian* sensing and we can conclude that the stability is guaranteed for a wide large of sense-act duration values. To understand how much this strategy can improve time-to-transit estimation a simple simulation in Gazebo using *Lagrangian* sensing has been set up. A feature is positioned in the left wall of a straight corridor and during three tests a robot collect information moving straight, then turning away from the feature and finally turning towards it. Fig. 3 shows a comparison between  $\tau_{geo}$  and  $\tau_{per}$ ; in (a) no sense-act interleaving is used, while in (b) it is shown the significant improvement in perception of  $\tau$  when a *sense-act* cycle is used (sense duration is  $0.4s$ , act

duration is  $0.25s$  and linear velocity  $v = 0.5m/s$ ).

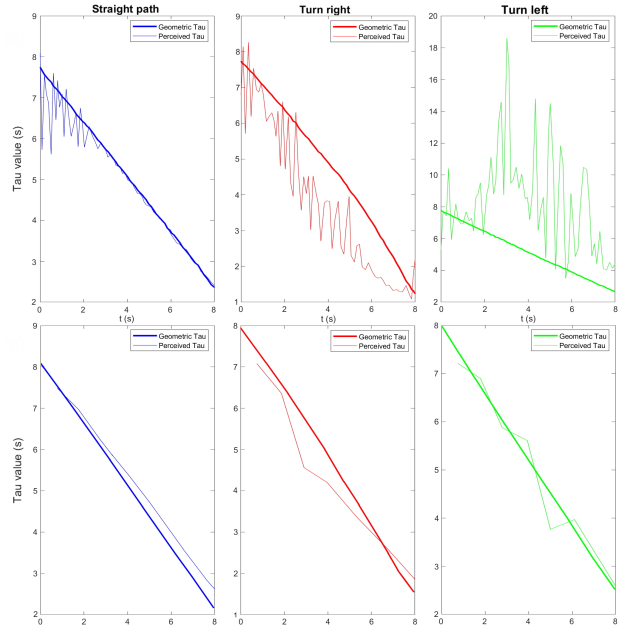


Fig. 3: Direct comparison between *geometric* (thick line) and *perceived* (fine line) *time-to-transit* values during three tests, at the top without and at the bottom with the implementation of the sense-act interleaving.

### III. STEERING CONTROL BASED ON *time-to-transit*

When flying through a narrow passage, bees position themselves in the center of it in order to experience the same image velocity in both eyes. This characteristic of bee’s flight has been highlighted by Srinivasan *et al.* in [12] and it represents the idea behind the *Tau Balancing* control law described in [13]. In this section we briefly describe how we adapt the motion primitives to the five ROIs. In particular, remembering (1), in this work we choose to use:

$$u(t) = k_f(\tau_{fl} - \tau_{fr}) + k_m(\tau_l - \tau_r) \quad (4)$$

where  $\tau_{fl}$ ,  $\tau_{fr}$ ,  $\tau_r$  and  $\tau_l$  are the time-to-transit value averages of the different ROIs of the image (the central ROI not included) as seen in the central panel of Fig. 2. Stability is guaranteed for every choice of  $k_f$  and  $k_m$ , but this parameter must be tuned to obtain best performances.

When features points are available only in the right or in the left part of the image (as in Fig. 4), tau balancing cannot be used. In [8], we introduced a *tau-difference maximizing* strategy for navigation using cues from a single wall. Unfortunately, this strategy has proven to be too sensitive to noise, and we will instead examine the control law

$$u(t) = \pm k(\tau_x - c) \quad (5)$$

where  $\tau_x \in \{\tau_{fl}, \tau_l, \tau_r, \tau_{fr}\}$  and  $c$  is a constant ([14]). To demonstrate the stability of (5), we analyze Eulerian models along the lines of [9], [13]. It is possible to isolate from (1) the following subsystem:

$$\begin{bmatrix} \dot{x} \\ \dot{\theta} \end{bmatrix} = \begin{bmatrix} \cos\theta \\ \pm k(\tau_x - c) \end{bmatrix}$$

and compute the rest point that, in this case, is  $(x, \theta) = \left( \frac{\pm(c-fR-f)}{f}, \frac{\pi}{2} \right)$  depending on the chosen  $\tau_x$ . The constant  $c$  is selected a priori depending on the general geometry of the environment, but a dynamic selection of  $c$  is possible. Using Eulerian sensing, a straightforward mathematical proof shows that the linearized controlled system is asymptotically stable. It can further be shown that if  $k < \frac{4}{fc^2}$  the eigenvalues of the linearized Eulerian dynamics become complex numbers, and this introduces oscillations in the system.

#### IV. PHYSICS-BASED SIMULATION AND IMPLEMENTATION ON ACTUAL ROBOTS

In this section we present some results from Gazebo simulations and from the tests performed on a real robotic platform<sup>2</sup>.

**A. Simulation Results:** To verify the behaviour of the system presented in Fig. 2 we run multiple Gazebo simulations in realistic and artificial environments and for each of them different geometric structures are considered (straight corridors, 90 degrees turns, corridors with multiple turns, single walls). Artificial environments have been used to verify the effectiveness of the software components in scenarios with a precise feature density. However, since the robot has to be able to navigate in the

<sup>2</sup>The code we used for the experiments is available on GitHub at [https://github.com/johnbailleul/Vision\\_based\\_Navigation\\_TTT](https://github.com/johnbailleul/Vision_based_Navigation_TTT).

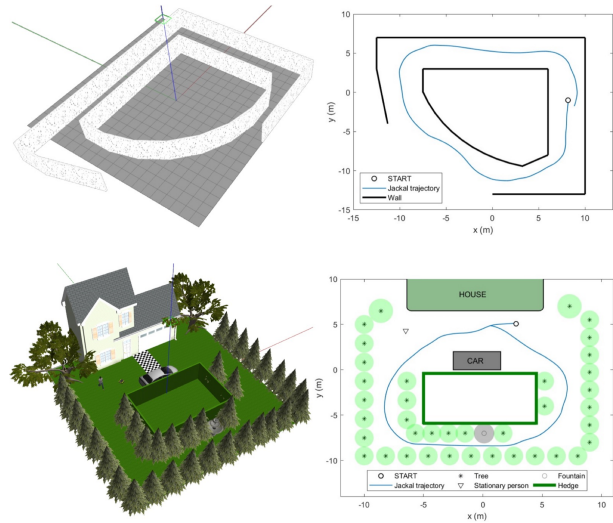


Fig. 4: At the top an artificial and at the bottom a realistic environment together with the trajectories followed by the robot during Gazebo simulations.

real world, realistic environments are used to test the real potential of the controlled system. Fig. 4 provides an example of both types of environments together with the possible challenges that the robot has to overcome when it navigates using our control strategy. The trajectories followed by the robot are also reported and they show that the platform is able to safely navigate in these environments by switching among the control laws.

**B. Experiments with a Jackal Robot UGV:** The promising results obtained with Gazebo simulations lead us to test our tau-based navigation strategy on real robots. The robotic platform we use is the Jackal UGV, an unmanned ground vehicle developed by Clearpath Robotics whose skid-to-turn steering is somewhat more challenging than the unicycle idealization (1). The extent of the difference between model (1) and the actual robot has proven to be fairly negligible, and the control laws described above have worked well. The Jackal robots have been equipped with stereo cameras (the MYNTEYE S1030 in Torino and the Stereolabs ZED at BU). We used the right monocular cameras as the only sensors of our navigation experiments. The most important characteristic is the horizontal field of view that has to be wide enough to recognize, with the



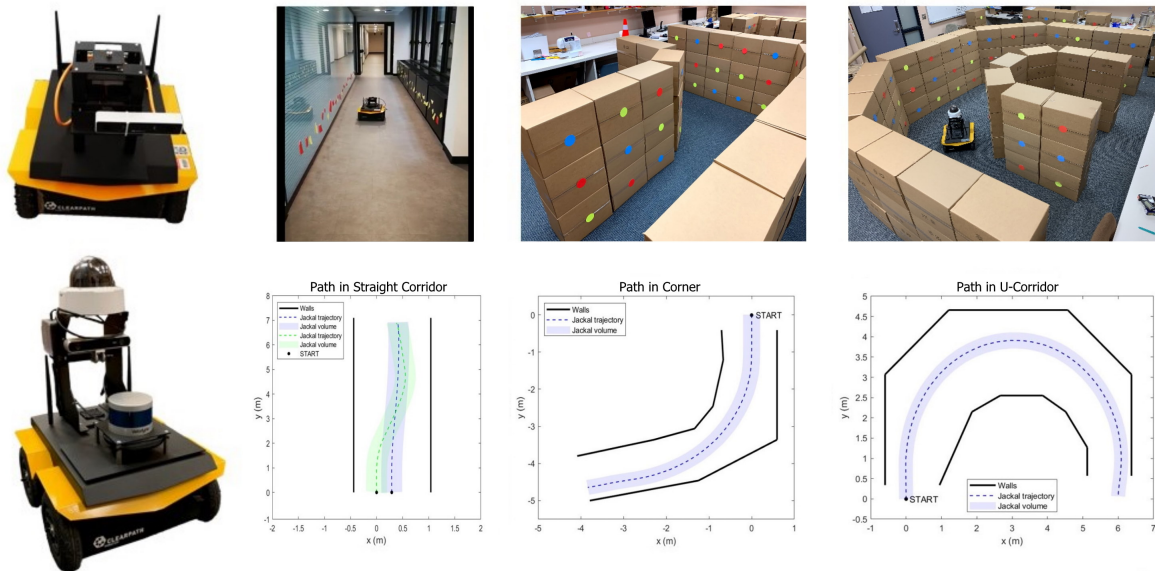


Fig. 5: On the left the Jackal UGV equipped with the MYNTEYE S1030 stereo camera (top image) and the Stereolabs ZED stereo camera (bottom image). On the right, some results from the tests run both at Boston University and at Politecnico di Torino. A variety of environments were constructed using moving boxes.

proper timing, specific characteristics of the visual cues, e.g. discontinuities of the time-to-transit signal in environments containing turns. The two robots are shown in the left panel in Fig. 5. We test the effectiveness of the entire algorithm by observing the behavior of the Jackal robot moving through different environments which have been set up at Politecnico di Torino and at the Boston University Robotics Lab. In each setting, we combine odometry and the IMU data to compute the position of the robot with respect to a coordinate frame that has its origin where the Jackal is powered on for the first time. In the upper images of Fig. 5 the different environment setups are shown, the features are mainly represented by the edges and the corners of the post-it notes but, since the tests have been conducted in real environments, there are also features belonging to other objects. The graphs in the lower panels of Fig. 5 are based on odometry data that show that the *tau-balancing* control law, as predicted by the theory, steers the robot toward the center of the corridor and even successfully negotiates corners.

## V. CONCLUSION AND FUTURE WORK

Selected results on reliable navigation of a mobile robot using the perceptual cue *time-to-transit* have been reported. The work validates ideas with origins in the Cornell University Perception Lab of James J. Gibson, and at the same time it sets the stage for a broader investigation of vision based navigation using perceptual cues associated with optical flow. The results on the use of a camera that is steered independently of the robot so as eliminate the confounding effects of movement along a curved path will be reported. While a simple strategy of steering the camera so as to cancel the rotational component of movement can be shown to be effective, our current work is also aimed at understanding how camera movement might be used to support perceptual processes that involve numerous animal and human brain regions that are known to support spatial cognition, [15],[16],[17]. In particular, our current research is being informed by investigations by Boston University colleagues (M. Betke and M. Hasselmo) who are working to correlate neural activity in running laboratory animals with 3D tracking of head and body poses recorded on thermal videos.

## REFERENCES

- [1] J.J. Gibson, 1950. *The perception of the visual world.*, Houghton Mifflin, First Edition Jan. 1, 1950.
- [2] Lee, D.N., 1976. A theory of visual control of braking based on information about time-to-collision. *Perception*, 5(4), pp.437-459.<https://doi.org/10.1068/p050437>
- [3] Lee, D.N. and Reddish, P.E., 1981. "Plummeting gannets: A paradigm of ecological optics." *Nature*, 293(5830), pp.293-294. <https://doi.org/10.1038/293293a0>
- [4] Baurès R, Fourteau M, Thébault S, Gazard C, Pasquiol L, Meneghini G, Perrin J, Rosito M, Durand JB, Roux FE. Time-to-contact perception in the brain. *J Neurosci Res*. 2021 Feb;99(2):455-466. doi: 10.1002/jnr.24740. Epub 2020 Oct 18. PMID: 33070400.
- [5] Tresilian, J.R. Perceptual and cognitive processes in time-to-contact estimation: Analysis of prediction-motion and relative judgment tasks. *Perception & Psychophysics* 57, 231–245 (1995). <https://doi.org/10.3758/BF03206510>
- [6] B.R. Fajen, "Steering toward a goal by equalizing taus". *Journal of Experimental Psychology: Human Perception and Performance*. 2001 Aug. 27(4), 953 - 968.
- [7] K. Sebesta and J. Baillieul. "Animal-Inspired Agile Flight Using Optical Flow Sensing". In: *Proceeding of the 51<sup>st</sup> IEEE Conference on Decision and Control (CDC)*, 2012, pp. 3721-3727.
- [8] Z. Kong, K. Özcimder, N.W. Fuller, A. Greco, D. Theriault, Z. Wu, T. Kunz, M. Betke, and J. Baillieul. "Optical flow sensing and the inverse perception problem for flying bats," In: *Proceeding of the 52<sup>nd</sup> IEEE Conference on Decision and Control (CDC)*, 2013, pp. 1608-1615.
- [9] J. Baillieul and F. Kang, "Visual Navigation with a 2-pixel Camera—Possibilities and Limitations," In *Proceedings of the 21st IFAC World Congress* in Berlin, Germany, July 12-17, 2020. Also available from <http://arxiv.org/abs/2103.00285>.
- [10] B. D. Lucas and T. Kanade. "An iterative image registration technique with an application to stereo vision". In: *IJCAI'81: Proceedings of the 7th International Joint Conference on Artificial intelligence*, 2 (1981), pp. 674–679.
- [11] J.Y. Bouguet. "Pyramidal Implementation of the Lucas Kanade Feature Tracker Description of the algorithm". Intel Corporation Microprocessor Research Labs, (2000). Available from [http://robots.stanford.edu/cs223b04/algos/affine\\_tracking.pdf](http://robots.stanford.edu/cs223b04/algos/affine_tracking.pdf).
- [12] M. V. Srinivasan, R. J. D. Moore, S. Thurrowgood, D. Socol and D. Bland. "From biology to engineering : insect vision and applications to robotics". In: *Frontiers in Sensing: From Biology to Engineering*, Springer, 2012.
- [13] J. Baillieul, 2019, December. "Perceptual Control with Large Feature and Actuator Networks." In *Proceedings of the 2019 IEEE 58th Conference on Decision and Control (CDC)* (pp. 3819-3826), DOI: 10.1109/CDC40024.2019.9029615.
- [14] C. Boretti and P. Bich, "Dictionary of motion primitives for vision-based navigation using Optical Flow", M.S. Thesis, Politecnico di Torino and Boston University, 2021. Available at <http://www.baillieul.org/Robotics/ThesisChiaraPhilippe.pdf>.
- [15] L.L. Chen, L.H. Lin, C.A. Barnes, B.L. McNaughton. "Head-direction cells in the rat posterior cortex" *Exp Brain Res*, 101, 24–34 (1994). <https://doi.org/10.1007/BF00243213>
- [16] J. Cho, P.E. Sharp, "Head direction, place, and movement correlates for cells in the rat retrosplenial cortex," *Behav. Neurosci.* 115, 3–25 (2001).
- [17] A.S. Alexander, L.C. Carsten, J.R. Hinman, F. Raudiesg, W. Chapman, and M.E. Hasselmo, 2020. "Egocentric boundary vector tuning of the retrosplenial cortex," *Science advances*. 2020 Feb 1;6(8):eaaz2322. DOI:10.1126/sciadv.aaz2322