



**Politecnico
di Torino**

ScuDo

Scuola di Dottorato - Doctoral School
WHAT YOU ARE, TAKES YOU FAR

Doctoral Dissertation
Doctoral Program in Electrical, Electronics and Communications Engineering
(37th cycle)

Visual Perception for Service Robotics: from Frame-based to Event-based sensors

By

Chiara Boretti

Supervisor(s):

Prof. Gianluca Setti, Supervisor

Prof. Marcello Chiaberge, Co-Supervisor

Politecnico di Torino

2026

Declaration

I hereby declare that, the contents and organization of this dissertation constitute my own original work and does not compromise in any way the rights of third parties, including those relating to the security of personal data.

Chiara Boretti
2026

* This dissertation is presented in partial fulfillment of the requirements for **Ph.D. degree** in the Graduate School of Politecnico di Torino (ScuDo).

Abstract

Autonomous service robotics represents a rapidly advancing field of research. The primary goal of a service robot is to assist humans and enhance the quality of their everyday lives. These robots are specifically designed to perform complex, dangerous, or physically demanding tasks, enabling greater safety and convenience for users in a fully autonomous manner. A critical aspect of a robotic platform's autonomy is its ability to perceive and interpret the environment effectively. Various sensors can be used to gather information about the surroundings, with vision-based sensors being among the most widely utilized. Visual data acquisition and processing play an essential role in enabling robots to understand and interact with the environment successfully, allowing tasks such as navigation and human-robot interaction. This data can be captured using two primary types of vision systems: traditional frame-based cameras and neuromorphic-inspired vision sensors, known as event-based cameras. Frame-based cameras process sequential images to extract visual cues. Alternatively, event-based vision sensors offer an innovative approach where each pixel operates independently, detecting and transmitting events triggered by brightness changes. This new working principle offers several advantages, including low latency, low power consumption, high dynamic range, and reduced redundancy in information. Neuromorphic vision has emerged as a promising foundation for developing more efficient and lightweight navigation strategies. Inspired by the way biological systems process visual stimuli, these approaches enable faster and more energy-efficient decision-making.

This dissertation explores the usage of biologically inspired strategies and of cutting-edge sensing technologies to develop efficient and versatile visual perception systems for autonomous robotics. The research is divided into two main parts: one focusing on the use of conventional frame-based cameras and the other on exploring the potential usage of event-based sensors. In detail, frame-based cameras are utilized to derive biologically inspired visual cues that facilitate autonomous navigation in

GPS-denied, unexplored environments while minimizing the sensor payload. Central to this approach is the concept of time-to-transit, which represents the time an autonomous platform requires to reach a specific feature detected in the image and present in its surroundings. Extensive experimentation has been conducted to improve the estimation of this quantity by leveraging various deep learning models, effectively addressing the limitations inherent in traditional computer vision techniques. Both simulation and real-world testing validate the feasibility of using time-to-transit as a signal for simple yet effective control strategies. The second part shifts the focus to novel event-based cameras, exploring their potential applications in service robotics. This includes contributing to the enrichment of the number of existing event datasets for tasks such as person detection and tracking, through the development of a new person-based dataset called PEDRo. Additionally, a new pre-processing pipeline, called Memory of Events through Spatial Attention (MESA), is introduced to organize events while preserving the temporal characteristics of event streams. The proposed methodology aims to expand the utility of event-based sensors within contemporary deep learning frameworks without the need to employ more complex neural networks such as recurrent architectures. Results on object classification and detection demonstrate that this technique improves what is achievable using state-of-the-art event representations alone, all while maintaining computational efficiency.

The findings discussed in this dissertation highlight the potential of biologically inspired vision strategies and innovative sensor technologies to advance the efficiency of service robotics platforms.