Perceiving Through Time: Spatio-Temporal Intelligence for Robust Robotic and Autonomous Systems

Robotic systems operating in dynamic, human-centric environments require reliable perception capabilities to ensure safe navigation, effective interaction, and long-term autonomy. Accurate spatial and temporal understanding is essential for anticipating human behaviour, responding to environmental changes, and maintaining robustness under real-world conditions. However, current perception systems often struggle with domain shift, limited generalization, and the gap between simulation and deployment. These challenges motivate the development of advanced spatio-temporal learning frameworks that can leverage temporal correlations and adapt to diverse, real-world scenarios.

This research aims to develop robust spatio-temporal perception frameworks to enhance video understanding for multi-object tracking and abnormal human behaviour forecasting in dynamic real-world environments. The framework addresses two primary tasks: multi-object tracking (MOT) and video anomaly detection. For MOT, it proposes a unified, open-set multi-modal tracking framework suitable for real-world deployment, enabling detection and tracking in crowded environments using visual cues or natural language descriptions, allowing flexible specification beyond predefined categories. For video anomaly detection, the approach follows a two-stage strategy. The first stage focuses on constructing datasets for abnormal human behaviour forecasting and video anomaly anticipation. The second stage integrates multiple modalities, including pose, depth, panoptic segmentation, optical flow, and language semantics, to capture detailed motion, spatial relationships, and contextual information. The framework would focus on real-time inference and investigate temporal search strategies using vision–language large models to identify anomalies in long video sequences.

This research advances anomaly detection and behaviour forecasting by modelling long-term temporal dependencies to enable coherent scene understanding, multi-agent tracking, accurate scene parsing, and safe, reliable human–robot interaction in dynamic environments. The framework has broad applications in robotics, intelligent mobility, autonomous driving, and smart health, addressing cognitive disorder monitoring and assisted living to promote safe, independent living while reducing healthcare costs. It also enhances public transportation by developing robust perception modules for autonomous vehicles, improving road safety, reducing reliance on individual cars, and supporting more efficient mobility.

Experimental validation will be carried out through real-world deployments and collaborative case studies, including home-care applications with Nice Hospital (CoBTeK Lab) and assisted-living robotic platforms in partnership with Toyota. These evaluations will demonstrate the framework's robustness, generalization, and practical impact. Overall, the research aims to bridge the gap between perception research and real-world deployment, strengthening Université Nice Côte d'Azur's leadership in smart cities, robotics, smart homes, and healthcare innovation.