Comparative Analysis of Feature Extraction Techniques with Kinect V1 for Low-Cost Autonomous Robotics

Filipi Enzo Siqueira Kikuchi

2025

Abstract

This work aims to analyze and compare different feature extraction algorithms applied to RGB-D images captured by a Kinect V1 sensor, in order to identify the most suitable methods for low-cost mobile autonomous robot applications. By evaluating techniques such as ORB, SURF, and BRISK, we aim to understand each method's performance in terms of robustness, repeatability, and computational efficiency. This investigation supports the selection of viable algorithms for visual SLAM systems operating under budget and processing constraints.

1 Introduction

Autonomous navigation in mobile robots requires the real-time construction and interpretation of maps of the surrounding environment. Among the most common approaches to achieve this goal are Simultaneous Localization and Mapping (SLAM) systems, especially vision-based SLAM (Visual SLAM), which rely on image sensors to estimate the robot's trajectory while building a representation of the environment.

With advances in computer vision, the detection and tracking of keypoints (features) have become essential components in the functioning of Visual SLAM systems. These features are used to estimate the camera's motion and reconstruct the 3D space around the robot. The choice of the feature extraction algorithm directly affects the system's accuracy, robustness, and computational cost.

This work investigates different feature extraction methods applied to RGB-D data captured by a Kinect V1 sensor, aiming to identify which performs best for low-cost mobile robotics scenarios. Choosing efficient algorithms can benefit academic projects, educational platforms, or commercial low-investment solutions, broadening access to autonomous navigation technologies.

2 Justification for Using Kinect V1 and Feature Extraction Comparison

The Kinect V1 sensor is widely adopted in robotics projects for offering both RGB images and depth data, making it a low-cost alternative to professional 3D mapping sensors such as LiDARs. Although originally developed for entertainment applications, the Kinect's real-time depth capabilities make it well-suited for navigation experiments and small-scale 3D reconstructions.

The goal of this project is to use the Kinect as an experimental platform to compare different feature extraction algorithms — such as ORB, SURF, and BRISK — applied to RGB and RGB-D images, evaluating their efficiency and robustness. These algorithms are extensively used in SLAM systems and computer vision applications, each with distinct trade-offs in execution time, invariance properties, and noise tolerance.

By understanding the behavior of these algorithms in a controlled environment with a low-cost sensor, we aim to contribute to the identification of a general-purpose approach for feature extraction in robotic platforms constrained by budget and processing power. This contribution is particularly relevant to educational contexts, emerging research labs, and accessible automation initiatives.

3 Literature Review

We present below a comparative analysis of key research papers that support the understanding of modern Visual SLAM approaches and their applications in autonomous robotics. This review provides a foundation to support the methodological decisions taken in this project.

Title	What It Is	Goal	How the Goal Is Achieved
3D Local Map Construction Using Monocular Vision (2010)	Study using monocular vision with delayed initialization, SURF, and Kalman filter for localization.	Develop an efficient monocular SLAM technique to generate local 3D maps.	Uses SURF feature detection, 3D line-based depth estimation, and Extended Kalman Filter, avoiding complex parameter tuning.
ORB-SLAM (2015)	A real-time monocular SLAM system based on ORB features, robust to viewpoint changes.	Propose a robust and accurate monocular SLAM system for varied environments.	Uses ORB features throughout the pipeline (tracking, mapping, loop closure), with graph-based structure and automatic initialization.
Direct Sparse Odometry (2016)	Monocular odometry system based on direct photometric error and sparse optimization.	Develop an accurate and real-time visual odometry method.	Joint optimization of camera parameters, depth, and pose using gradient-rich sparse points.
ExplORB-SLAM (2022)	Active extension of ORB-SLAM2 using pose graph structure for exploration.	Optimize SLAM performance using utility-based navigation (D-optimality).	Detects frontiers, predicts graph expansion through "hallucination", and selects motion based on expected utility.
Tightly-Coupled LiDAR-Visual SLAM (2023)	SLAM system integrating LiDAR and monocular camera data using geometric features.	Create a robust and accurate SLAM system with low-cost sensors.	Fuses data in spherical coordinates; visual subsystem refines depth, LiDAR adjusts feature direction; fallback mechanism ensures robustness.

The progression of these works clearly highlights the trend:

- From purely geometric and monocular solutions to multimodal sensor fusion approaches;

- From point-only features to the inclusion of line and semantic features;
- From passive observation to active utility-guided navigation;
- And from high-cost sensors to accessible alternatives — such as Kinect-based RGB-D cameras.

From this table, we can observe how Visual SLAM has evolved over time, with the emergence of new techniques and approaches aimed at improving the accuracy, robustness, and efficiency of these systems.

Due to the continuous and long-term operation of agents that rely on SLAM-based navigation in various environments, concerns about redundancy in sensor readings become a key focus in papers proposing robust solutions. In this context, more recent works tend to adopt a hybrid approach, combining geometric analysis, active perception strategies, and semantic understanding, in addition to the fusion of data from multiple sensors such as LiDAR and monocular cameras. This fusion aims to enhance robustness by achieving higher reliability in perception, relocalization, and loop closure.

Among the evaluation metrics used in SLAM system analysis, techniques involving photometric and geometric error optimization—such as Root Mean Square Error (RMSE)—are the most common. These are often accompanied by real-time performance evaluations and assessments of system resilience in dynamic environments.

References