cādence®

# SLAM and DSP Implementation

Amol Borkar, Senior Product Marketing Manager, Tensilica Vision and AI IP Group, Cadence

With the introduction of simultaneous localization and mapping technology, or SLAM, there comes a need for more sophisticated DSPs to handle the required computations. To address this need, Cadence has introduced the Tensilica® Vision Q7 DSP to handle the requirements of SLAM, including high performance, and low power, and an ease of development that engineers can leverage to design new and exciting applications with this technology.

## Contents

Introduction ....................................1
SLAM .............................................2
SLAM applications............................2
Market trends for SLAM ...................3
SLAM requirements..........................3
Tensilica Vision Q7 DSP.....................4
Ease of development and tools..........5
Performance comparison..................5
Conclusion .......................................6
References........................................6

## Introduction

When mobile phones were first introduced, they were intended for phone calls and texts only. As each generation of phones was brought into the mobile market, these phones became increasingly complex, more functions and capabilities were added, eventually leading to the smartphones of today. These smartphones not only allow you to make phone calls but also enable you to browse the internet in the palm of your hand. Additionally, they can also track your location and give directions—even evaluate your driving habits, among other things.

With the introduction of GPS technology in the ’90s, tracking movement became a relatively easy task. This technology opened the door to several navigation-, route-planning-, and surveying-type applications. GPS does have its limitations, however. It is accurate only to within a few meters, thereby restricting its use to applications in which tracking “micro-movements” is not necessary. And, in certain areas where access to GPS satellites is limited (cities with tall buildings, mountains, etc.), you don’t have access to the data that GPS supplies, nullifying its use. As newer and more demanding applications are developed, tracking these micro-movements is now becoming necessary; therefore, we must look beyond what GPS offers.

Fortunately, simultaneous localization and mapping (SLAM) is able to orient you to within inches and doesn’t require satellite connectivity. SLAM is the computational problem of constructing a map in an unknown environment while simultaneously keeping track of the device’s position (location and orientation) within it. SLAM comprises tracking six degrees of freedom (6DoF), which is composed of three degrees for position (up/down, back/forward, and right/left), and three for orientation (yaw, pitch, and roll) (Figure 1) to understand your position in an environment.

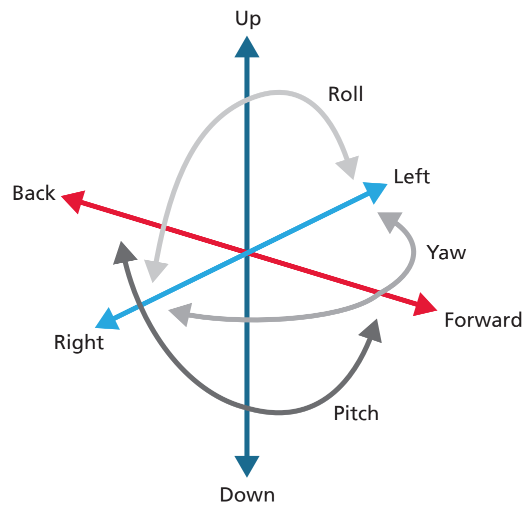


*Figure 1: 6DoF*

6DoF is an improvement over 3DoF which may only track yaw, pitch, and roll. So, for example, in an AR/VR environment, with a 3DoF system, you can view the environment only from a fixed location; a 6DoF system also tracks where you are in that environment.

Using SLAM technology, a mapping application can be used to identify where you are facing in an environment—for example, a street—and it can tell you whether to turn right or left. Using a simple GPS calculation only tells you that you are at an intersection; it won't know which way you are facing until you have already walked in the wrong direction.

## SLAM

SLAM is quickly becoming an important advancement in embedded vision. This capability provides the device with the ability of location awareness. So, for example, modern robots must be able to not only track where the sensor is pointing (orientation), but also how the sensor moves throughout space (speed, direction, and altitude).

Computations for SLAM were typically done with a camera sensor as the only form of input. This was known as Visual SLAM (VSLAM). But in the past few years, with the suite of additional sensors becoming available, SLAM has evolved to fusing additional sensor inputs.

Most SLAM systems work by tracking a set of points through successive camera frames and other sensor data to triangulate the camera's 3D position, while simultaneously using this information to approximate camera (or another sensor) orientation. Basically, the goal of these systems is to map their surroundings in relation to their own location for the purposes of navigation. As long as there is a sufficient number of points being tracked through each frame, both the orientation of the sensor(s) and the structure of the surrounding physical environment can be rapidly understood.

So, for example, in the case of smartphone-based implementations, SLAM may use a combination of cameras, GPS information, and inertial measurement units (IMUs, which provide data from internal accelerometers and gyroscopes that help to estimate the sensor's orientation) as inputs. SLAM is then used to determine which direction the phone is facing and how the phone is moving through space. When GPS data is available, it can be used to fortify the position estimate.

### SLAM applications

SLAM is a key component to location awareness in uncountable markets. SLAM-based applications are used in:

- **Mobile phones:** In augmented reality (AR) applications, SLAM is used to determine camera position and orientation to accurately render virtual objects on the screen. Indoor navigation also relies on SLAM to track movement and position inside buildings and malls where GPS is not available.
- **Automobiles and vehicles:** Self-driving vehicles often use SLAM along with GPS to better understand their position on the road and to navigate to their destination.

- **Drones and UAVs:** SLAM is used to estimate the movement of unmanned aerial vehicles (UAVs) for surveying and to build maps.
- **Robots in warehouses or fulfillment centers:** Robots use SLAM for navigation within a warehouse to determine and retrieve inventory.
- **Robot vacuum cleaners:** SLAM is used by the robots to efficiently navigate within the living space to avoid cleaning the same area twice.
- **AR/VR headsets:** User's head position and movement direction are determined using SLAM to accurately render virtual objects as well as make games and other interactions more immersive.

## Market trends for SLAM

As shown in Figure 2, the SLAM technology market size is set to exceed $2 billion by 2024 [1]. Major drivers for this market growth are the advancements in SLAM algorithms and the growth of SLAM in AR/VR. The rising technological developments and growing awareness regarding the benefits offered by the SLAM technology are primarily driving the market demand. Growing interest in the technology, particularly from industries including augmented virtual reality and autonomous vehicles, has resulted in the expansion of SLAM technology across the globe.

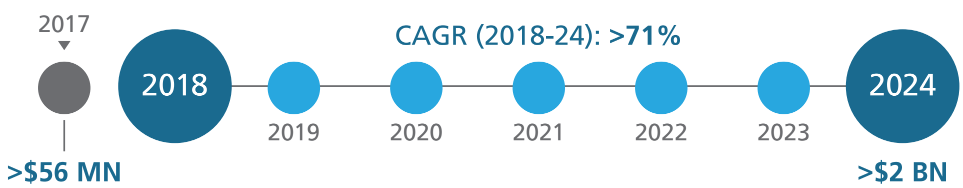


*Figure 2: The SLAM technology market is set to exceed $2 billion by 2024*

Moreover, SLAM technology in mobile robotics for both indoor and outdoor environment applications opens a pool of opportunity for the larger adoption of the technology across various end-user industries. Over the past five years, leading technology companies have made significant investments in SLAM technology to integrate into various business expansion strategies such as new product developments and mergers and acquisitions.

## SLAM requirements

Figure 3 shows a generalized flow of SLAM.

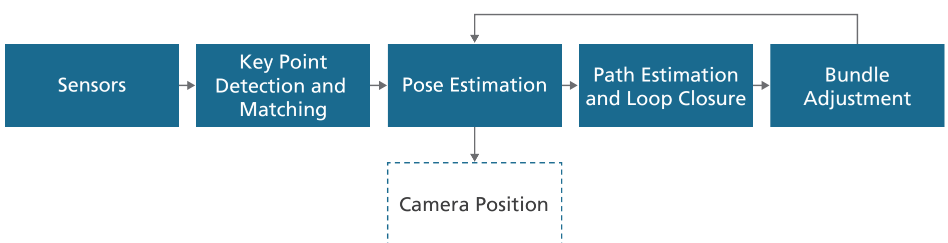


*Figure 3: SLAM process flow*

Each of the blocks is based on a classical computer vision (CV) approach. However, they rely heavily on a variety of linear algebra and matrix operations, so they are computationally heavy, and can be implemented on a CPU or GPU.

Using a CPU is great for general-purpose usage and prototyping, but it has limited performance capabilities. One limitation is a small number of SIMD lanes for parallel processing. Secondly, it is not power efficient, so it's not the best option to scale, and, in some cases, may not be able to deliver SLAM performance in real time.

Using a GPU is the next level up, in terms of computational ability. It has a variety of modalities for parallel processing, which can help achieve great performance and to meet real-time requirements. But again, GPUs are also power hungry and generate a lot of heat, which can be an issue for devices that have limited battery life, such as, a mobile phone. In other markets, SoC vendors cannot justify adding the real estate needed for a GPU in their floorplan just to do processing in this way.

This is where a specialized DSP comes in. DSPs are highly programmable and require a small area, making them scalable for mass deployment in devices of various markets.

## Tensilica Vision Q7 DSP

The Cadence® Tensilica Q7 DSP is designed from the ground up to meet the needs of applications using SLAM, enabling high-performance SLAM on the edge and in other devices. The Vision Q7 DSP is the sixth generation of vision and AI DSPs from the Tensilica family. Cadence has optimized instructions for faster performance on matrix operations, feature extraction, and convolutions to give the best performance yet on vision DSPs, providing the perfect balance of high performance and low power that is essential to SLAM applications at the edge. It can deliver up to 2X greater performance for vision and AI in the same area compared to its predecessor, the Tensilica Vision Q6 DSP.

Figure 4 shows the architecture and key features of this DSP.

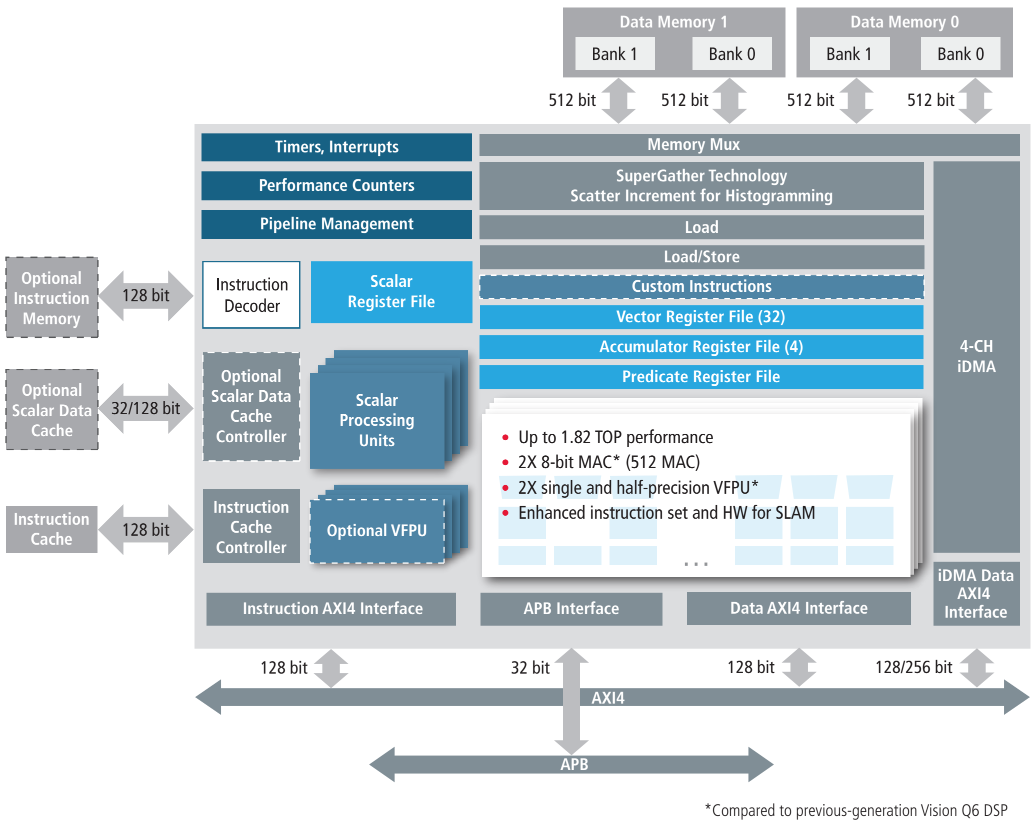


*Figure 4: Tensilica Vision Q7 DSP architecture*

The Tensilica Vision Q7 DSP offers the following high-level features:

- 512 MAC (8-bit) processing
- 64-way SIMD VLIW processor
- 1024-bit memory interface with dual load and store
- 2X vector floating point unit (vFPU) processing compared to previous DSPs
- Integrated 3D DMA with four channels
- Optional packages to accelerate SLAM performance
- Delivering up to 1.82 tera-operations per second (TOPS)
- 1.5GHz peak frequency

Additionally, the Vision Q7 DSP is designed for ISO 26262 certification, making it a great candidate for automotive applications.

## Ease of development and tools

In addition to being fully supported in the Tensilica Xtensa® Xplorer development environment, the Vision Q7 DSP also leverages the mature and highly optimized Cadence Xtensa Imaging Library. Inspired by OpenCV (the C++ computer vision library), Cadence has ported many of the OpenCV functions, maintaining similar function names and API, so transitioning from OpenCV is straightforward.

The Vision Q7 DSP is supported by the Tensilica Neural Network compiler. The Tensilica Neural Network compiler maps neural networks into executable and highly optimized high-performance code for the Vision Q7 DSP, leveraging a comprehensive set of optimized neural network library functions.

The Vision Q7 DSP also supports Halide, a computer programming language designed for writing digital image processing code that takes advantage of memory locality, vectorized computation, and multi-core CPUs and GPUs. This allows the programmer to experiment with different schedules and find the most efficient one. The Vision Q7 DSP allows the developer to write high-performance image and array processing code by taking advantage of several constructs inherent on the DSP.

## Performance comparison

Cadence has performed an in-house implementation of VSLAM using a single camera input and profiled the various blocks of the SLAM pipeline on both the Vision Q7 DSP and its predecessor, the Vision Q6 DSP (see Figure 5).

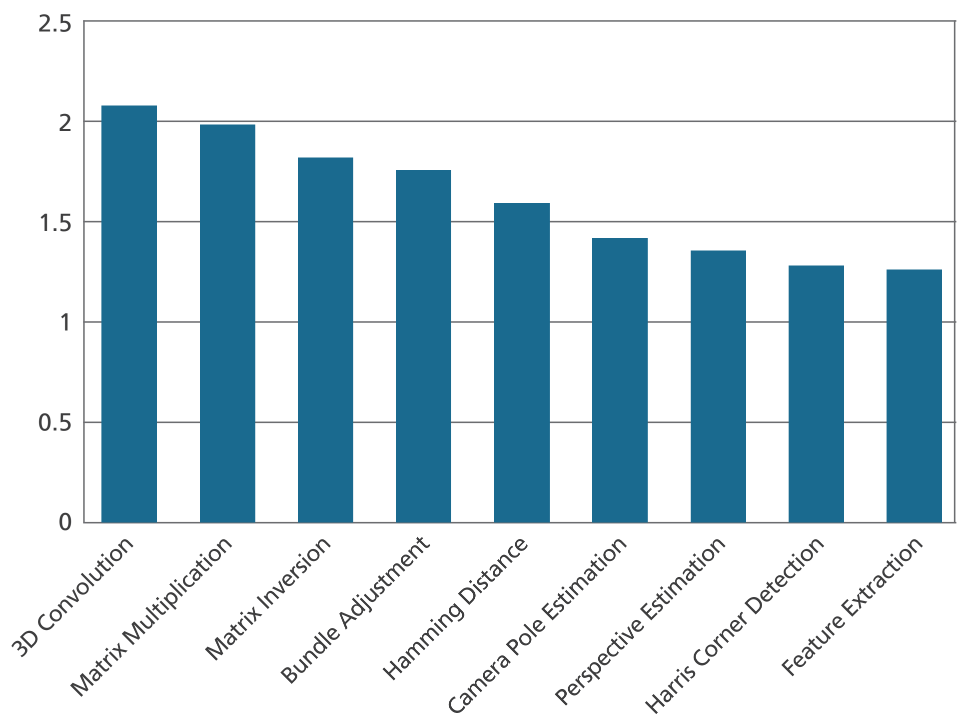


*Figure 5: The Vision Q7 DSP speed over the Vision Q6 DSP: Up to 2X improvement on various blocks of SLAM*

The Vision Q7 DSP shows close to 2X performance gain over the Vision Q6 DSP in various blocks of the SLAM pipeline. Improved instructions, optimized packages, and more MACs result in higher frequency for estimating camera position, and, furthermore, a better experience when the Vision Q7 DSP is used to accelerate SLAM-based applications. While providing this performance gain, the Vision Q7 DSP also requires a smaller area than the Vision Q6 DSP and consumes less power, making it a great platform for next-generation devices.

## Conclusion

In this paper, we introduced the concept of SLAM and walked through the implementation of our Vision DSPs. We have also shown a comparison between the Vision Q7 DSP and its predecessor, the Vision Q6 DSP, and the improvements in performance in the various blocks.

This paper focuses on purely computer vision approaches to implement a SLAM workflow. Recent advances have been made by integrating various convolutional neural network (CNN) layers to enhance the feature extraction block. The Cadence Tensilica Q7 DSP supports many layers required by the latest neural networks, making this type of fusion between vision and AI possible on the same DSP, and creating a harmonious marriage of vision processing and AI for next generation of SLAM-based applications.

## References

1. A. Bhutani and P. Wadhwani, "SLAM Technology Market size worth over $2bn by 2024," Global Market Insights, 1 October 2018. [Online]. Available: https://www.gminsights.com/pressrelease/slam-technology-market. [Accessed 1 May 2019].