Descriptor-In-Pixel
Point-Feature Tracking for Pixel Processor Arrays

Laurie Bose, Piotr Dudek

vision chip

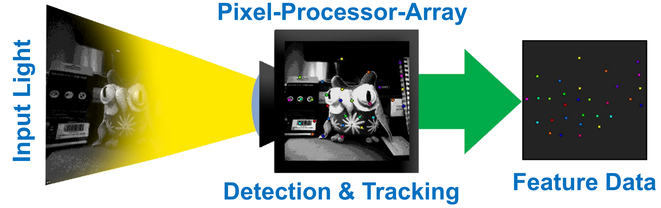
MANCHESTER 1824
The University of Manchester

CVPR Nashville JUNE 11-15, 2025

What's New

A novel Point-Feature Detection & Tracking approach

- All computation performed in sensor by "Pixel-Processors"
- No image transfer, runs at 3000 FPS
- Each pixel stores a descriptor, and computes its "Response"
- Detection & Tracking are unified via Descriptor Response

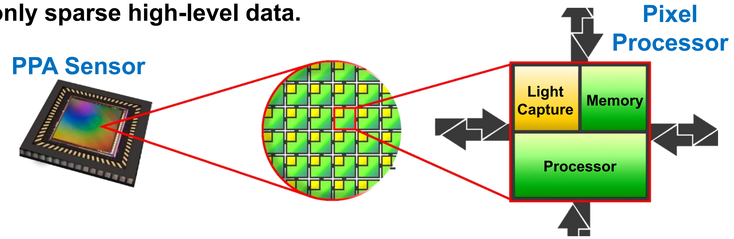


Motivation

Computer vision today transfers whole images from sensor to processor. This is slow, power hungry, and most pixels do not even contain useful information. How can we avoid this?

In-Pixel Processing

The SCAMP-7 Pixel-Processor-Array (PPA), is a sensor where each pixel is a simple programmable processor. These thousands of "Pixel-Processors" operate together as a SIMD computer. This enables "In-Pixel" computer vision to reduce sensor output to only sparse high-level data.



How It Works (Briefly)

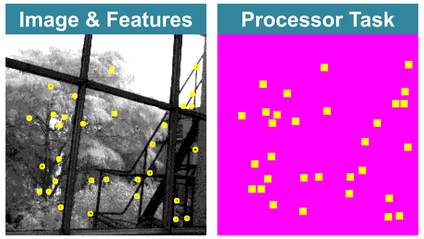
Our approach involves placing descriptors, encoding specific image structures & patterns, into every Pixel-Processor.

Processors near a feature

- Used to track that feature
- Store the descriptor of that feature

All other Processors

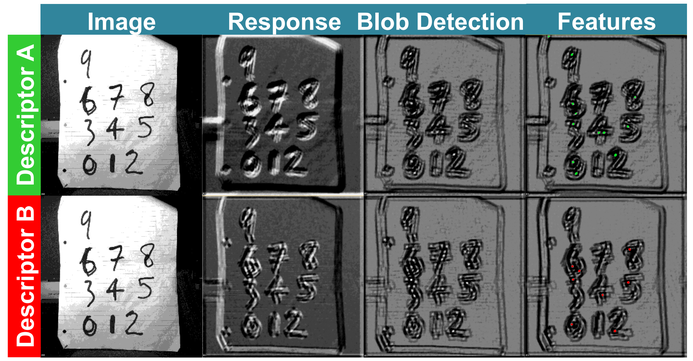
- Detect new features
- All store same "search" descriptor



Each pixel-processor measures how well its stored descriptor matches the local image. This produces a "descriptor response" map, which is used to both detect and track point-features.

Descriptor Response

Descriptor response highlights visual structures that match the descriptors stored within each pixel-processor. Point-features appear as well localized maxima within this response.

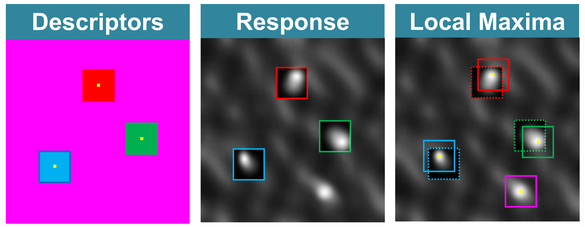


For simplicity, the examples above use the same descriptor across all pixel-processors. Each of the descriptors produces a different response, revealing a different set of point-features.

Unified Tracking & Detection

Features follow the local maxima of their descriptor's response. At high frame rates the locations of these maxima barely change between frames. As a result, each feature can be tracked from just a small patch of its descriptor's response, centred around its previous location.

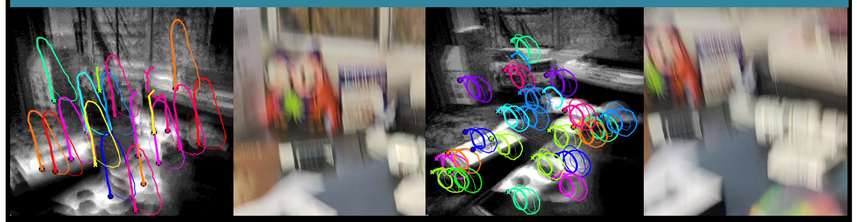
Thus, each feature's descriptor is spread into a local patch of processors. Into all other processors a randomized "search" descriptor is loaded



The responses of all these stored descriptors are computed in parallel, revealing the locations of features both old and new. This unifies detection & tracking into a single response driven process.

Performance

SCAMP-7 Features vs Smart Phone



Output Mode	FPS	Bits Per Frame
Feature Locations	~3000	16 Per Feature
Features Locations + 1Bit Image	~1000	16 Per Feature + 65,536
Features Locations + 4Bit Image	~300	16 Per Feature + 262,144

Our approach produces temporally stable features, that can be tracked under rapid motion at thousands of FPS. Implemented on SCAMP-7, an academic prototype with only 20bits of memory per Pixel-Processor.