

Seeing Is Geometry Plus Filtering

Review of *Foundations of Computer Vision: Computational Geometry, Visual Image Structures and Object Shape Detection* (Springer, Intelligent Systems Reference Library, 435 pp.)

TLDR / 80-20

A machine does not see an image. It reconstructs a scene from one.

The stated aim is to reconstruct and interpret natural scenes from captured images — which makes vision an inference problem, not a recording problem.

The Keys

- **The goal is reconstruction, not capture.** Cameras record intensities. Vision recovers the structure that produced them, which is strictly harder and sometimes impossible.
- **Images are treated as geometry.** The book's distinctive move is a computational-geometry view: images as vector spaces with structures to be detected rather than pixels to be processed.
- **A pixel has a location, not just a value.** Working with pixel centres on a grid makes neighbourhood and adjacency the primitive operations.
- **Linear filtering is the workhorse.** A filtered pixel is a weighted linear combination of its neighbours, and the weights live in a kernel the same size as the neighbourhood.
- **Filtering is convolution.** A linear spatial filter convolves a kernel across the image — the same operation Fourier optics uses to describe a lens.
- **Edges are where the content is.** Filtering exists to suppress noise, sharpen features, and detect edges and corners, because structure lives at discontinuities.
- **Shape detection is the destination.** Colour spaces, filtering and texture are preliminaries; the object of interest is the shape an image implies.
- **It explicitly emulates human vision.** The stated design reference is human visual perception, which is a modelling choice with consequences.

AI's Take

- **This is the applied end of the same mathematics.** Convolution kernels in software are the discrete version of the impulse response optics describes in glass.
- **The geometry framing is what distinguishes it.** Most vision texts are statistical or learning-based; treating images as geometric objects is an older and more structural approach.
- **It predates the deep-learning consensus.** That makes it dated for practitioners and more useful for understanding, since the mechanisms are visible rather than learned.
- **The hard problem is named but not solved.** Reconstructing a three-dimensional scene from a two-dimensional projection is underdetermined, and no filter fixes that.
- **Read it beside Fourier optics.** The two describe the same operation from opposite sides — one in physics, one in code.