Novel Hardware for Spatial AI

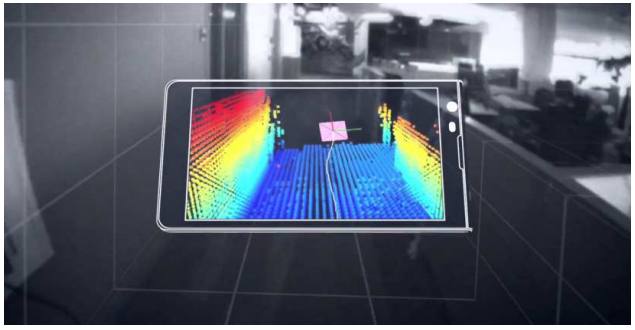
Andrew Davison

Dyson Robotics Laboratory
Department of Computing
Imperial College London
Twitter: @AjdDavison

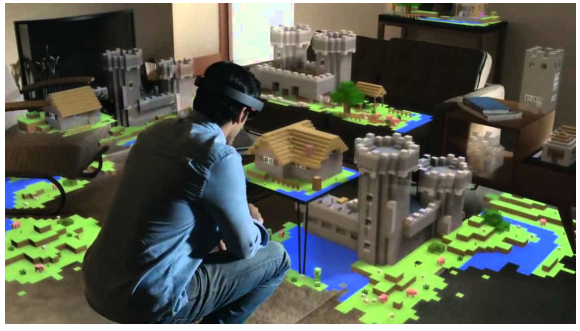
June 17, 2019

SLAM-Enabled Products and Systems

Dyson/iRobot/etc.



ARKit/ARCore/etc.

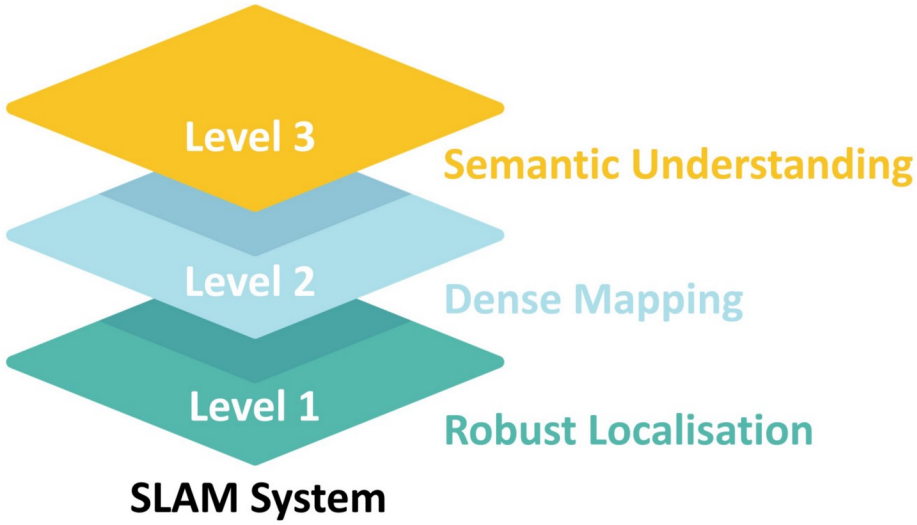


Oculus/HoloLens/etc.

DJI/Skydio/etc.

- Positioning and sparse/semi dense reconstruction now rather mature. . . and enabling real products.
- Dense and semantic mapping are being developed towards product extremely rapidly.

Cumulative Levels of Performance in SLAM



SLAM is evolving into Spatial AI

SLAMCORE

- Visual SLAM research has always been about robust real-time systems, with progress best demonstrated by demos, which have gradually become more capable.
- As SLAM performance improves, and the scope of what we consider SLAM broadens, it is our best model of the way to achieve Spatial AI.

Breakthrough Spatial AI Products of the Future!

- Example 1: A mass market household robot. Its tasks will include the ability to check whether furniture and objects have moved or changed; to clean surfaces and know when they are clean; to manipulate and tidy arbitrary objects; and to deal promptly and respectfully with humans by assisting them. Its Spatial AI system will be constrained by price, aesthetics, size, safety, and power usage, which must fit within the range of a consumer product.

A Breakthrough Spatial IA Product!

- Example 2: A future augmented reality system with the form factor of a standard pair of glasses. It will provide its wearer with a robust real-time overlay and spatial memory of all of the places, objects and people they have encountered, enabling things such as easily finding lost objects, and the placing of virtual notes or other annotations on any world entity. It will weigh 65g and have all-day battery life.

Spatial AI

- Spatial AI is the online problem where vision is to be used, usually alongside other sensors, as part of the Artificial Intelligence (AI) which permits an embodied device to interact usefully with its environment.
- 'FutureMapping: The Computational Structure of Spatial AI Systems', A. J. Davison, arXiv:1803.11288, 2018.

Hypotheses

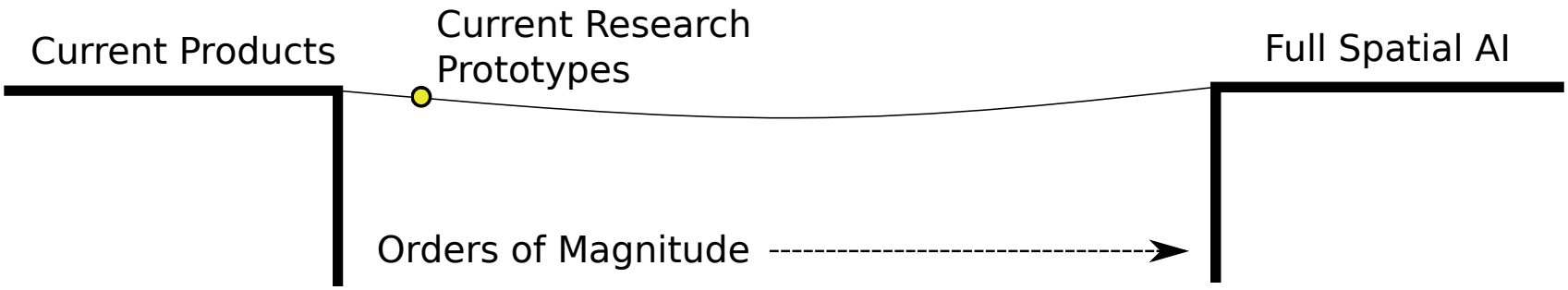
- When a device must operate for an extended period, carry out a wide variety of tasks, and communicate with other entities, its Spatial AI system should build a persistent and understandable scene representation which is close to metric 3D geometry, at least locally.
- The usefulness of a Spatial AI system for different applications is well represented by small number of performance measures.

Spatial AI: The Future of True, Embodied Intelligence

- In the recent Technical Report 'A Berkeley view of systems challenges for AI' (Stoica *et al.*, 2017), it is argued that devices which can act intelligently in their environments via continual learning must be capable of Simulated Reality (SR) which can 'faithfully simulate the real-world environment, as the environment changes continually and unexpectedly, and run faster than real time'.
- Spatial AI enables simulated reality (SR). Judea Pearl, recently discussing efficient situated learning and the need to reason about causation, argues that 'what humans possessed that other species lacked was a mental representation, a blue-print of their environment which they could manipulate at will to imagine alternative hypothetical environments for planning and learning'.

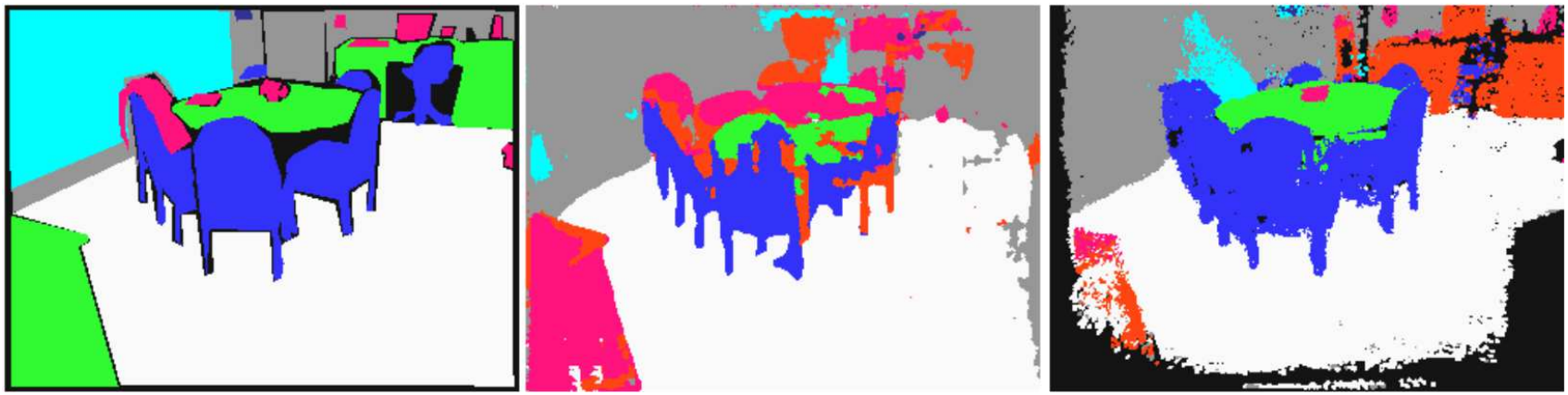
A Large Gap to Close

- AI (robotics) and IA (AR) need very much the same capabilities from a perception system.
- Precise, low-latency localisation; dense, predictive local scene modelling; semantic object instance mapping; long-term scene understanding in difficult, changing scenes; with very efficient operation on low cost hardware.



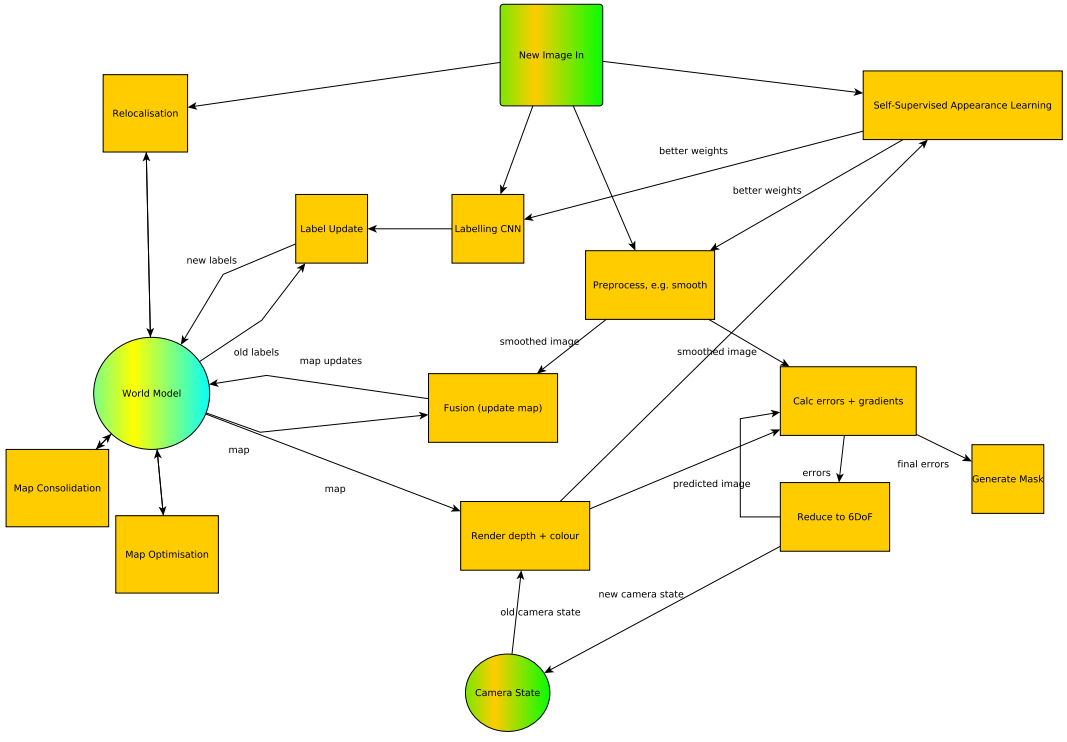
- Despite rapid progress, I still think we have a long way to go!

SemanticFusion: Prototype Dense, Semantic SLAM



- SemanticFusion: McCormac, Handa, Davison, Leutenegger, ICRA 2017.
- Dense SLAM (ElasticFusion) with per-frame CNN labelling of frames fused into 3D surfels. Smoothing of labels and geometry with MRF.

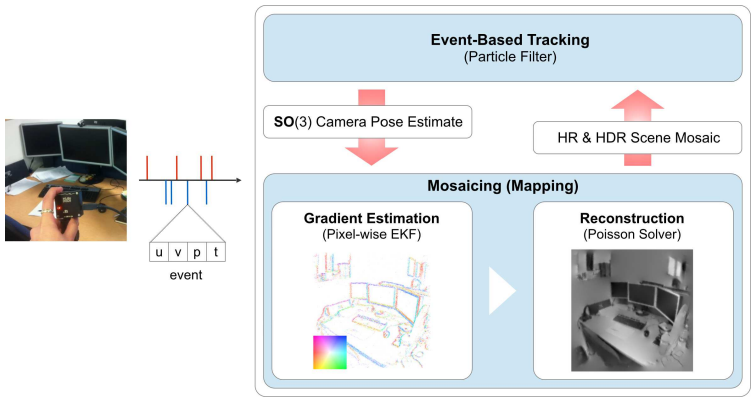
Semantic SLAM Computation Graph

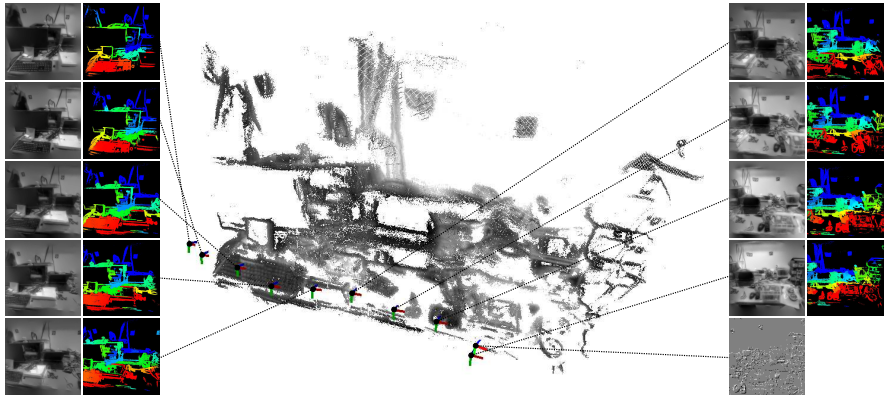


Complicated, loopy storage, computation and communication.

- **This** (and more) is what we need to optimise to get to true Spatial AI. Learned and designed modules.
- Lines of attack: 1. Hardware; 2. Representation.

Event Cameras: Towards SLAM Competences





- Event cameras have sensors which have per-pixel, asynchronous brightness change detectors. The output is a stream of events with image location, sign and timestamp.
- Alternating filters to estimate tracking and dense scene gradient (upgraded to intensity). Using DVS128.
- Kim, Handa, Ieng, Benosman, Davison, BMVC 2014 (Best Industry Paper).

Our Approach to Event-Based SLAM

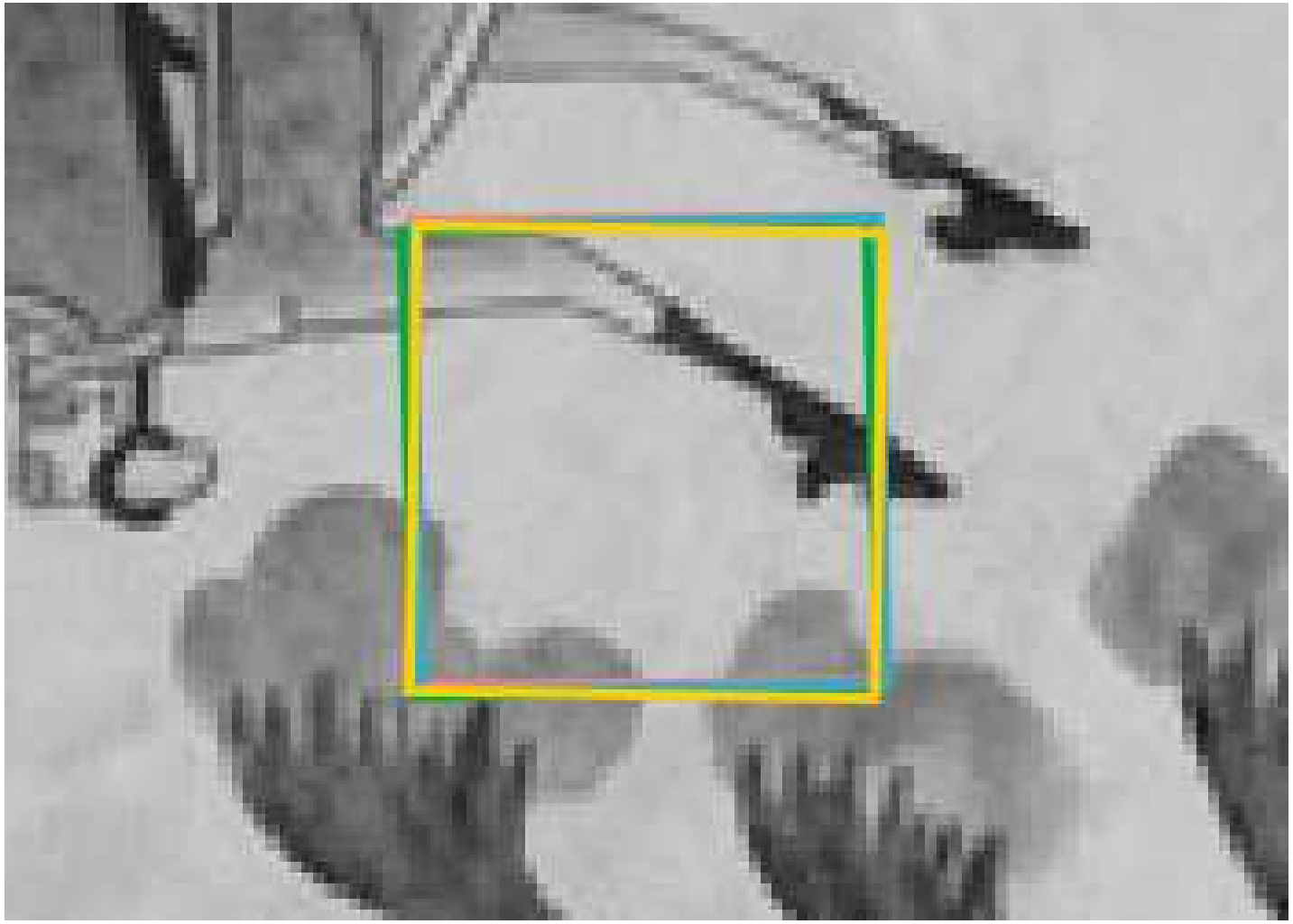
Generative

- Forward model of event generation; and inference where data is compared against a fully predictive model; comparable to ‘direct’ methods with standard cameras.
- This takes us on the route to ‘generally aware’ vision systems, where we pay attention to every piece of data

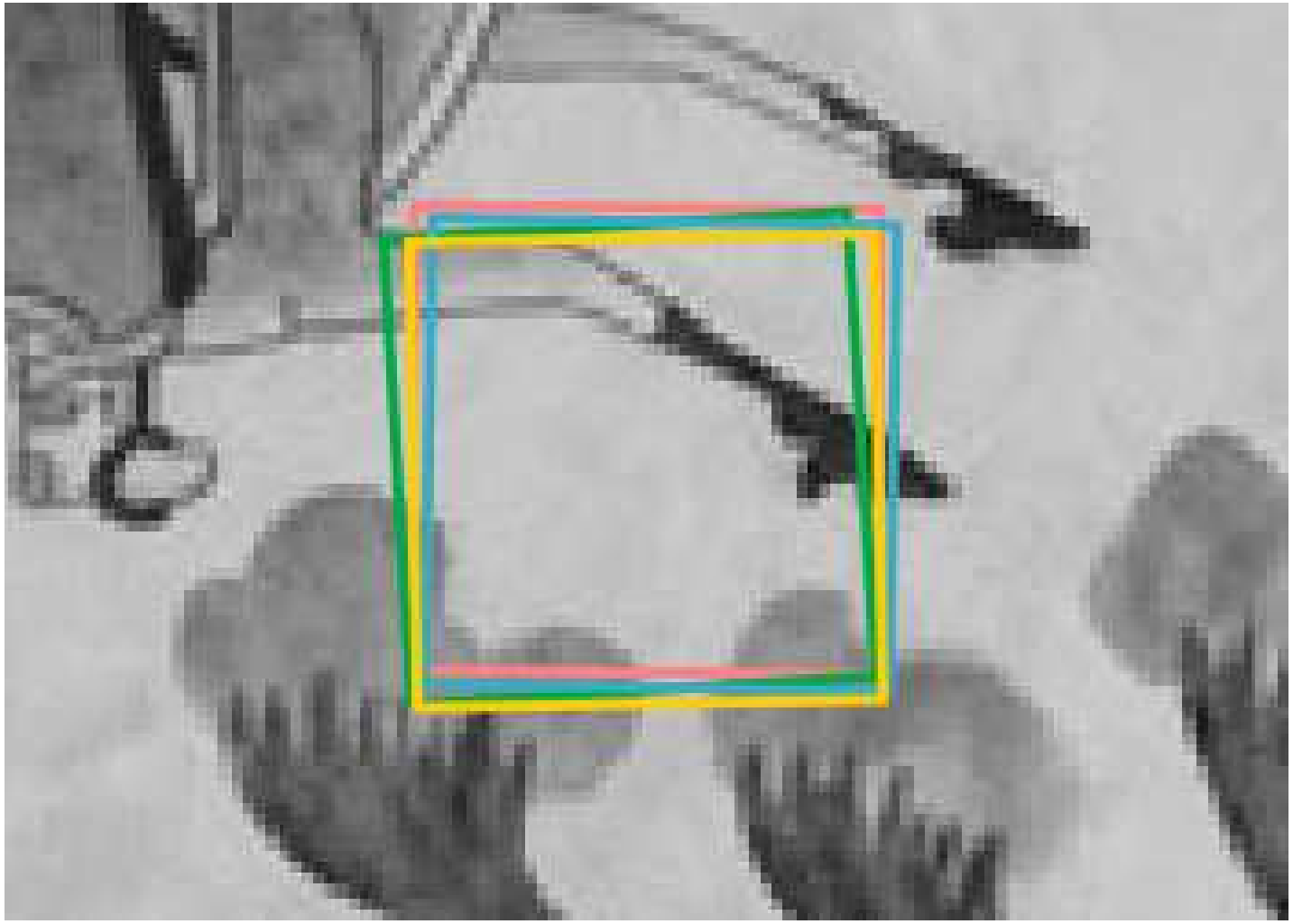
Event by Event Processing

- Purely event-based; minimise latency; use filtering methods.
- Dealing with very high event rates one by one is tough on a CPU, but this problem should go away with future integrated sensor/processor architectures.

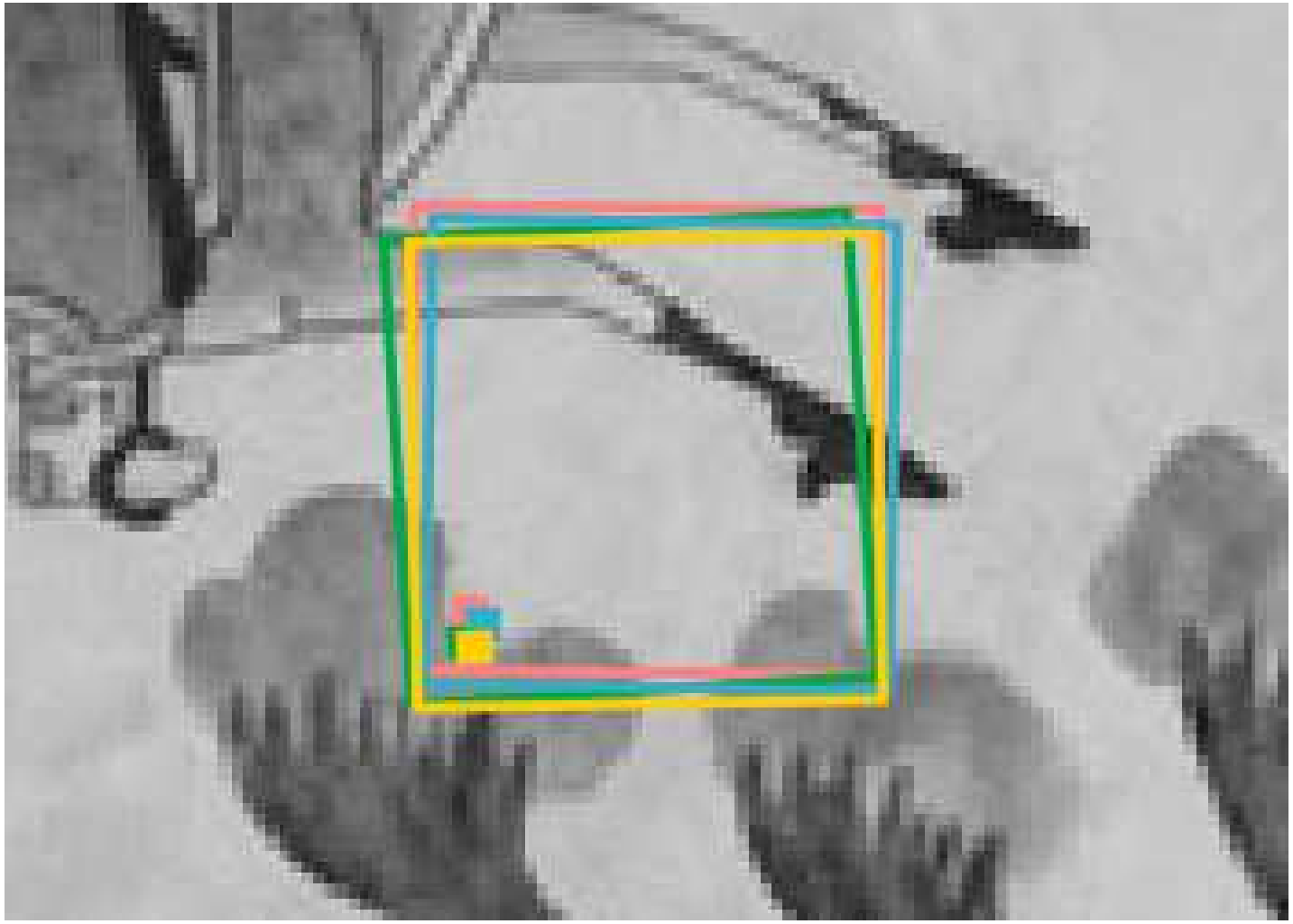
Tracking Filter Update from One Event



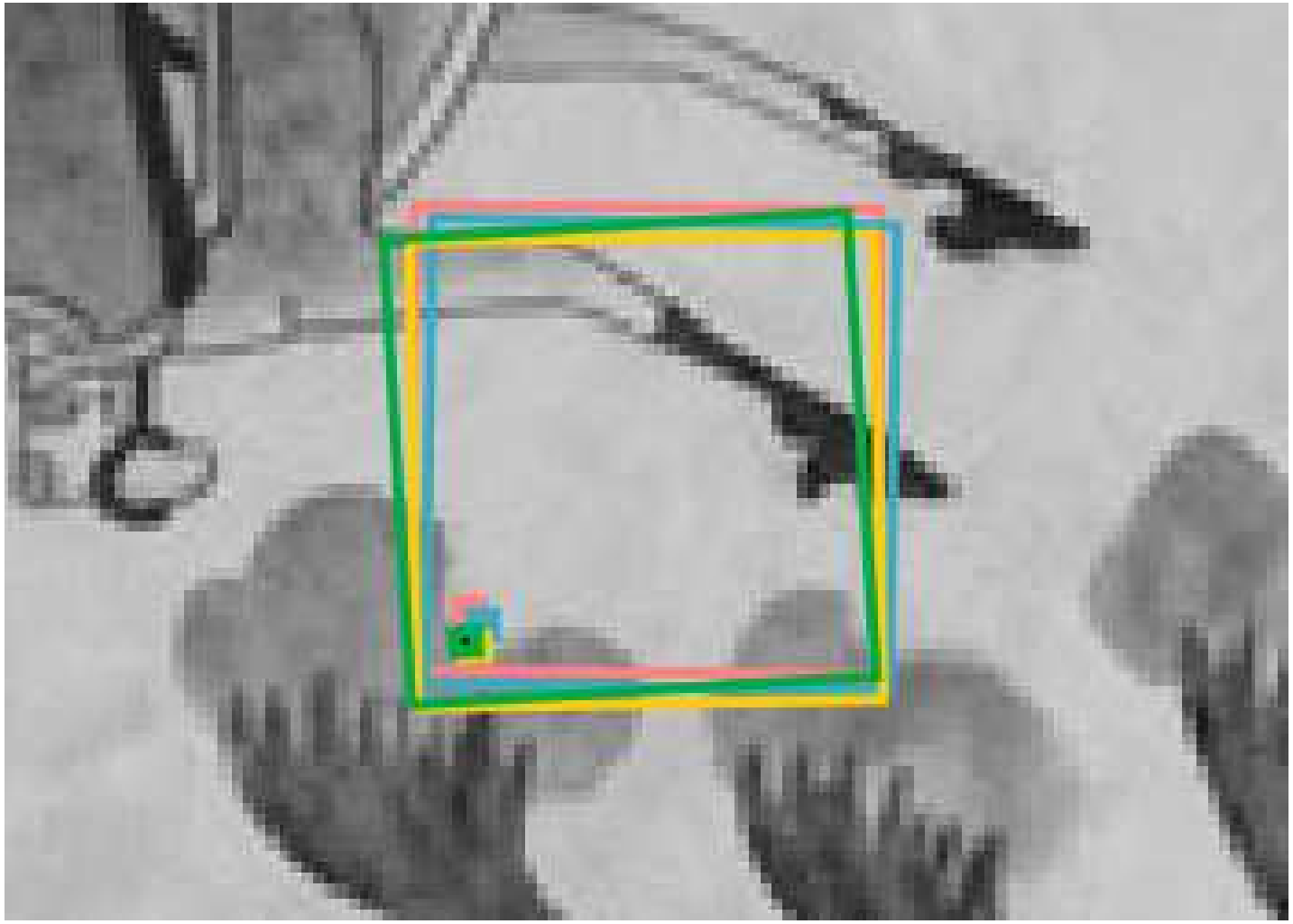
Tracking Filter Update from One Event



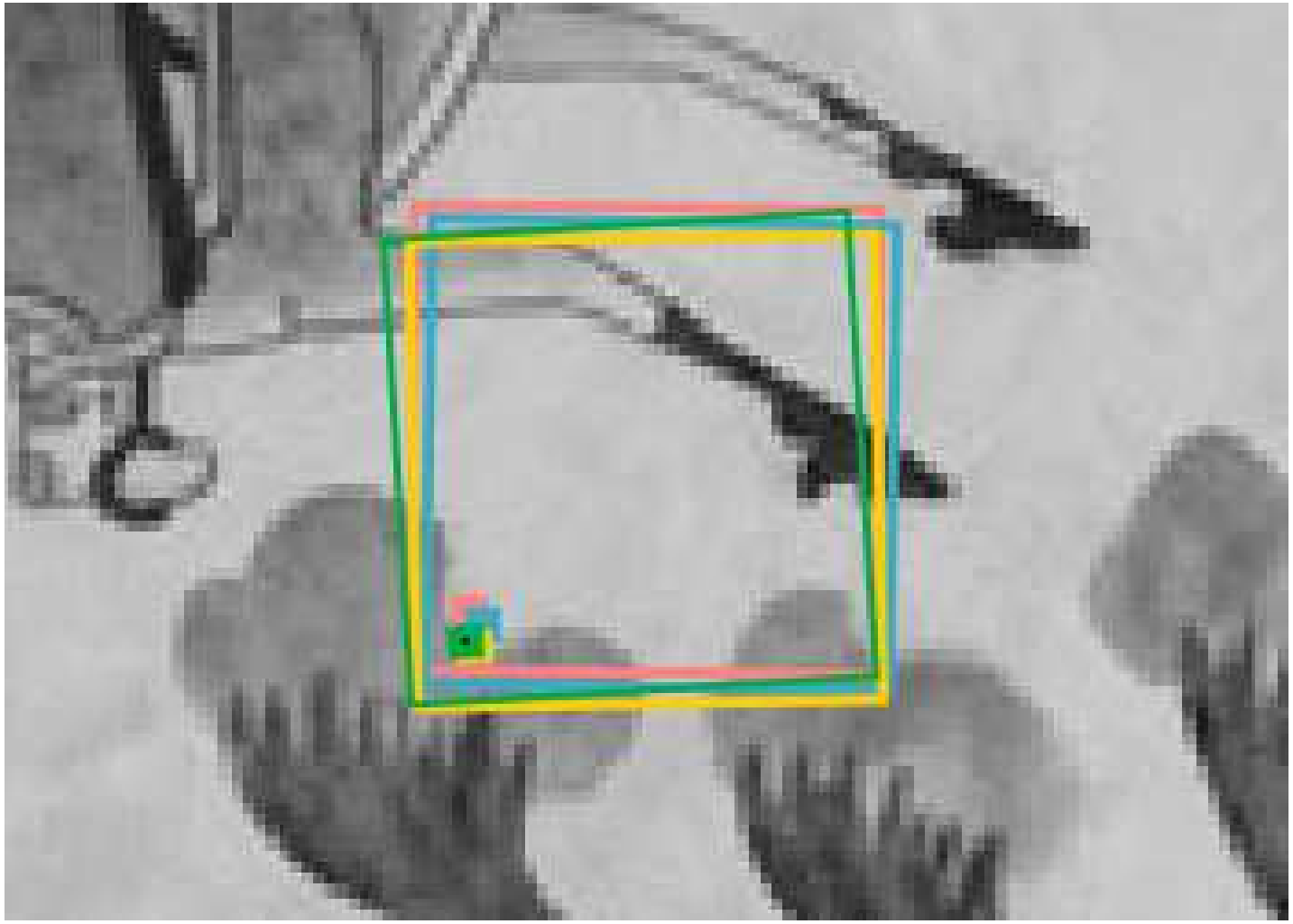
Tracking Filter Update from One Event



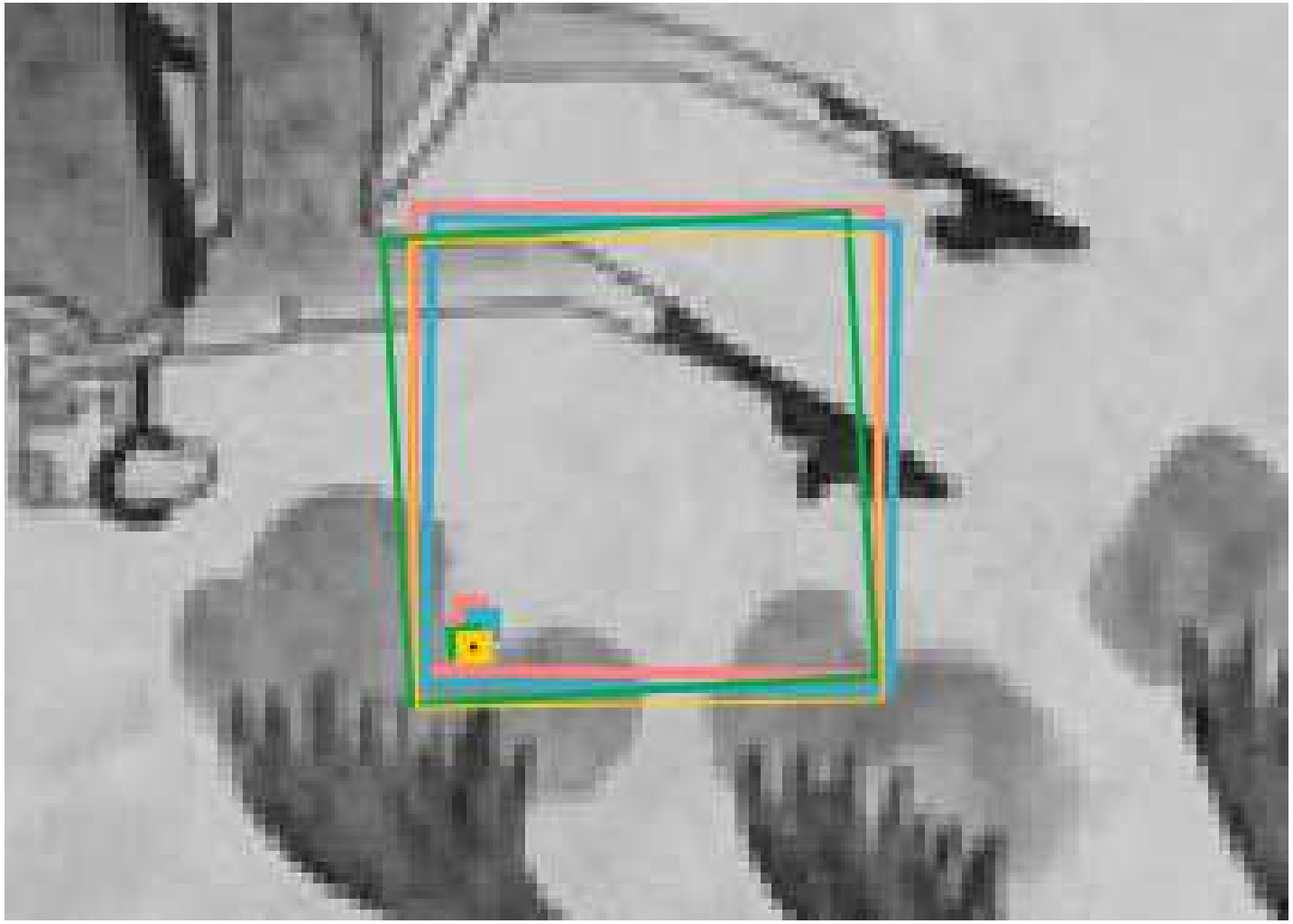
Tracking Filter Update from One Event



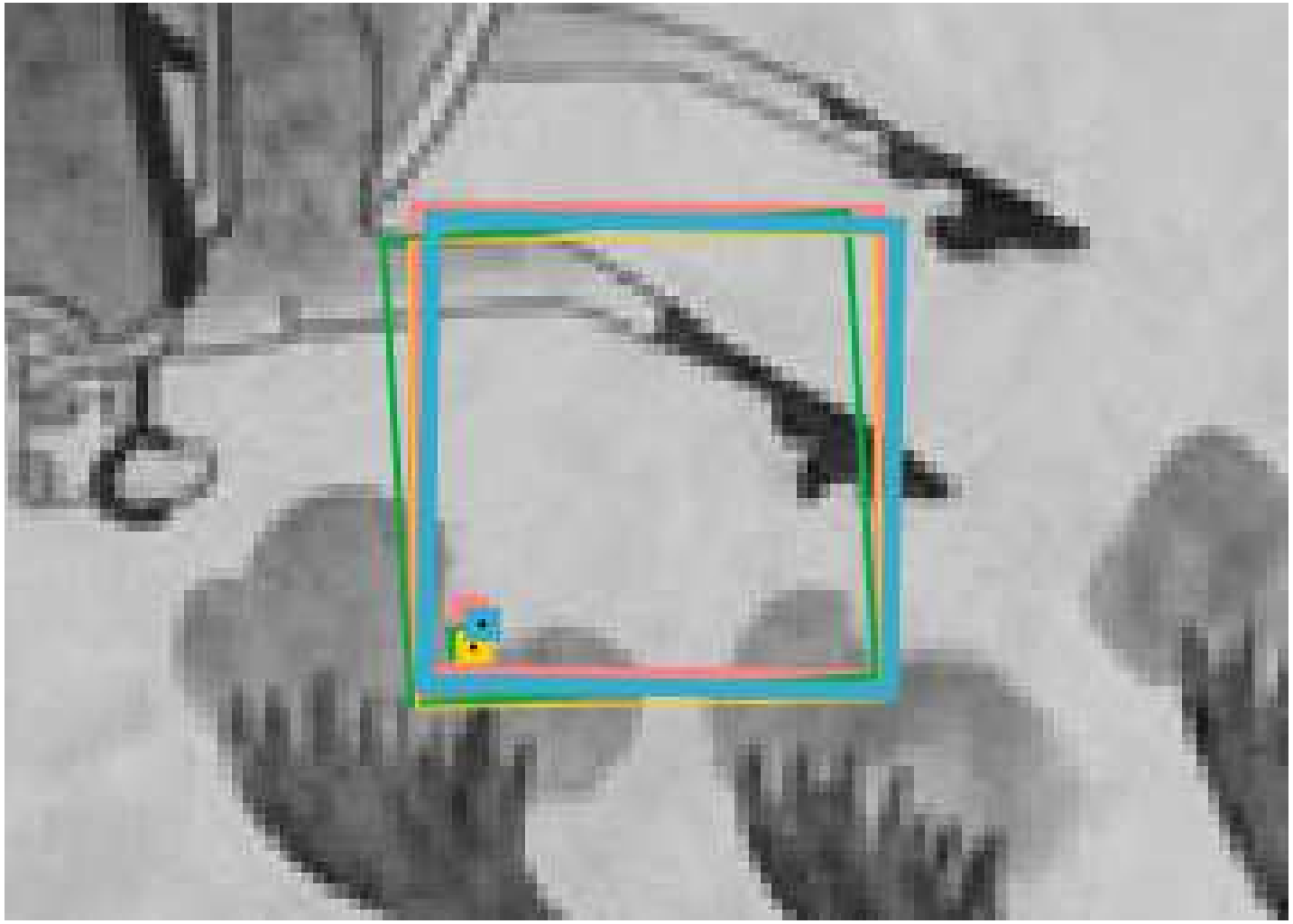
Tracking Filter Update from One Event



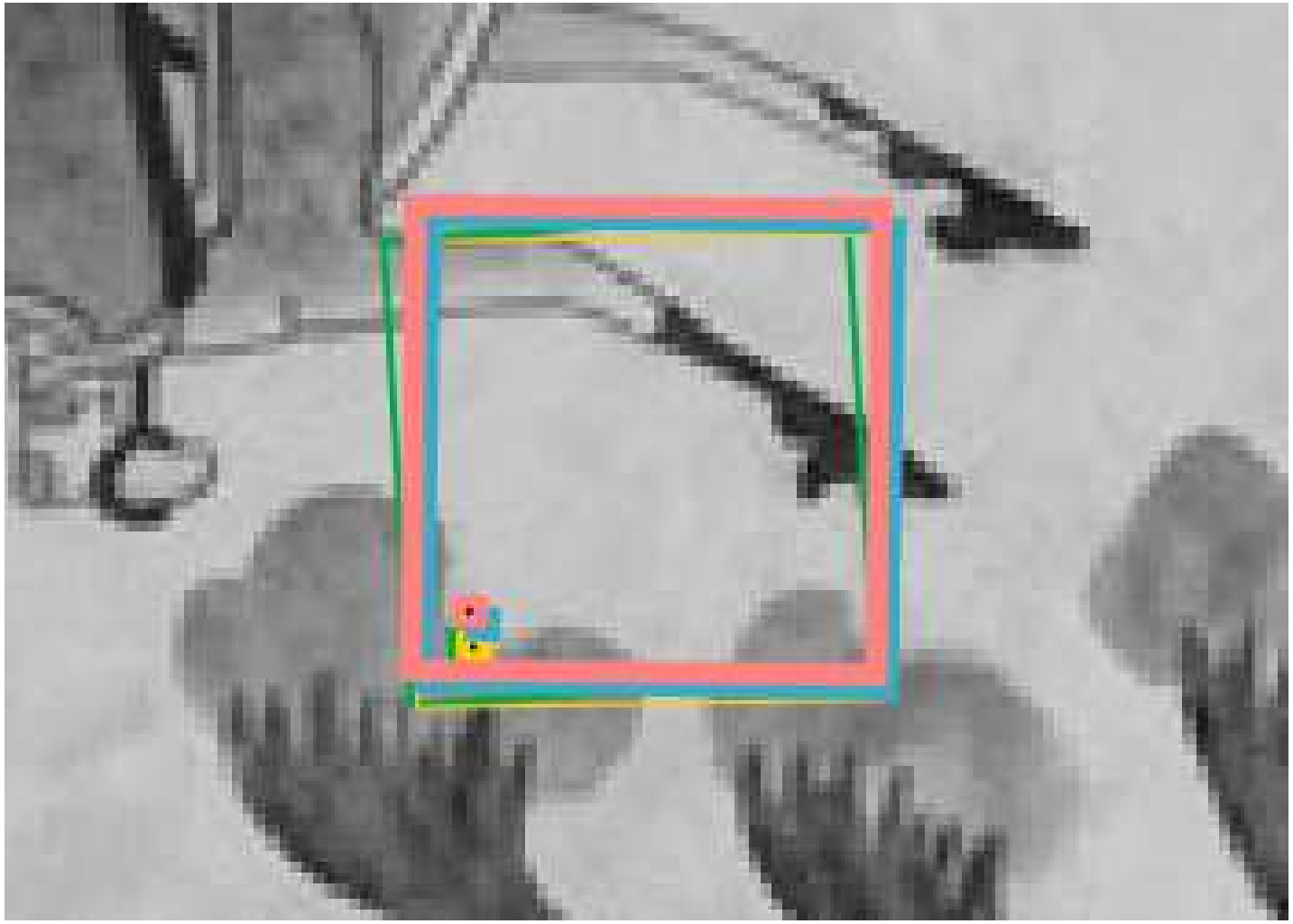
Tracking Filter Update from One Event



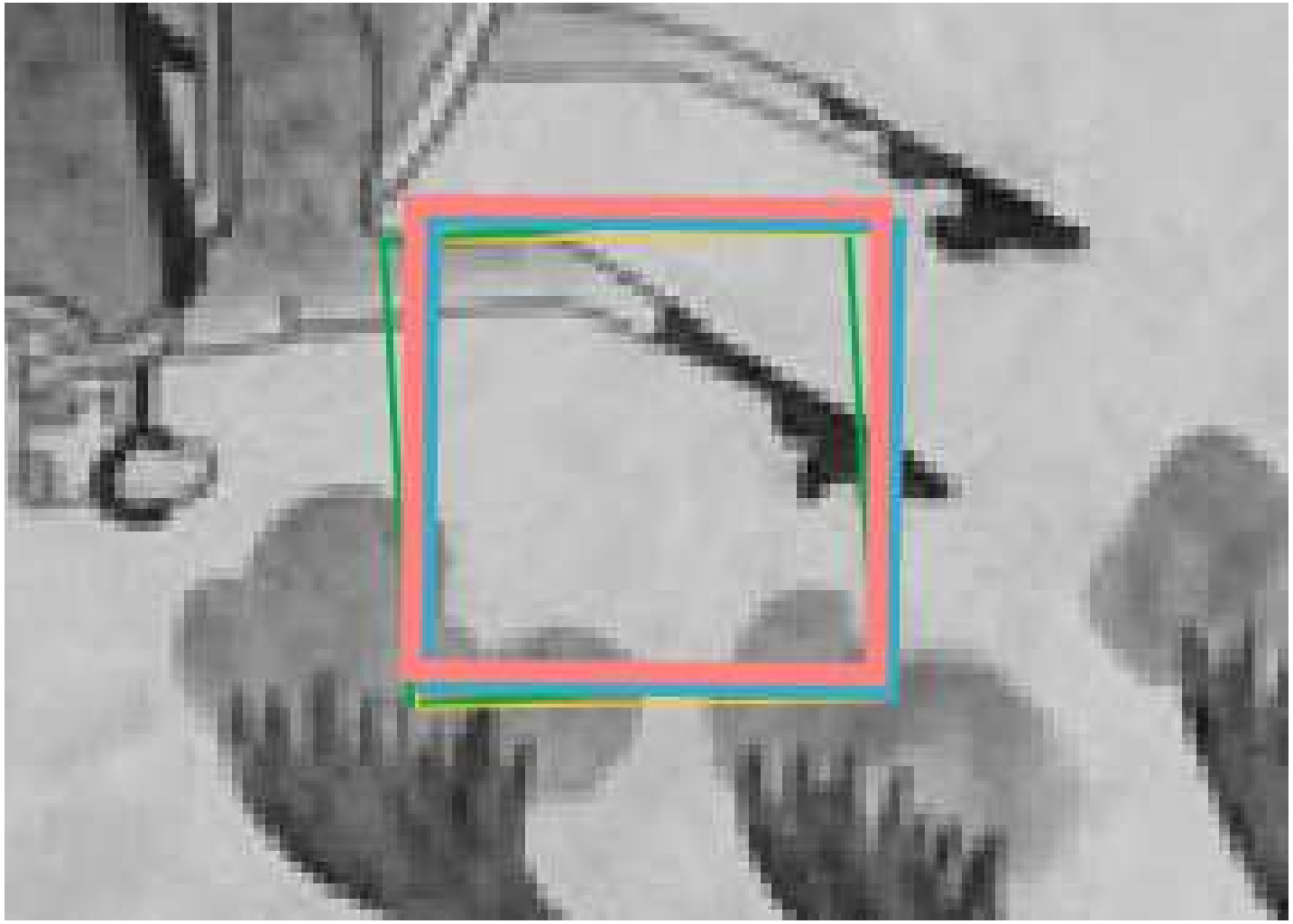
Tracking Filter Update from One Event



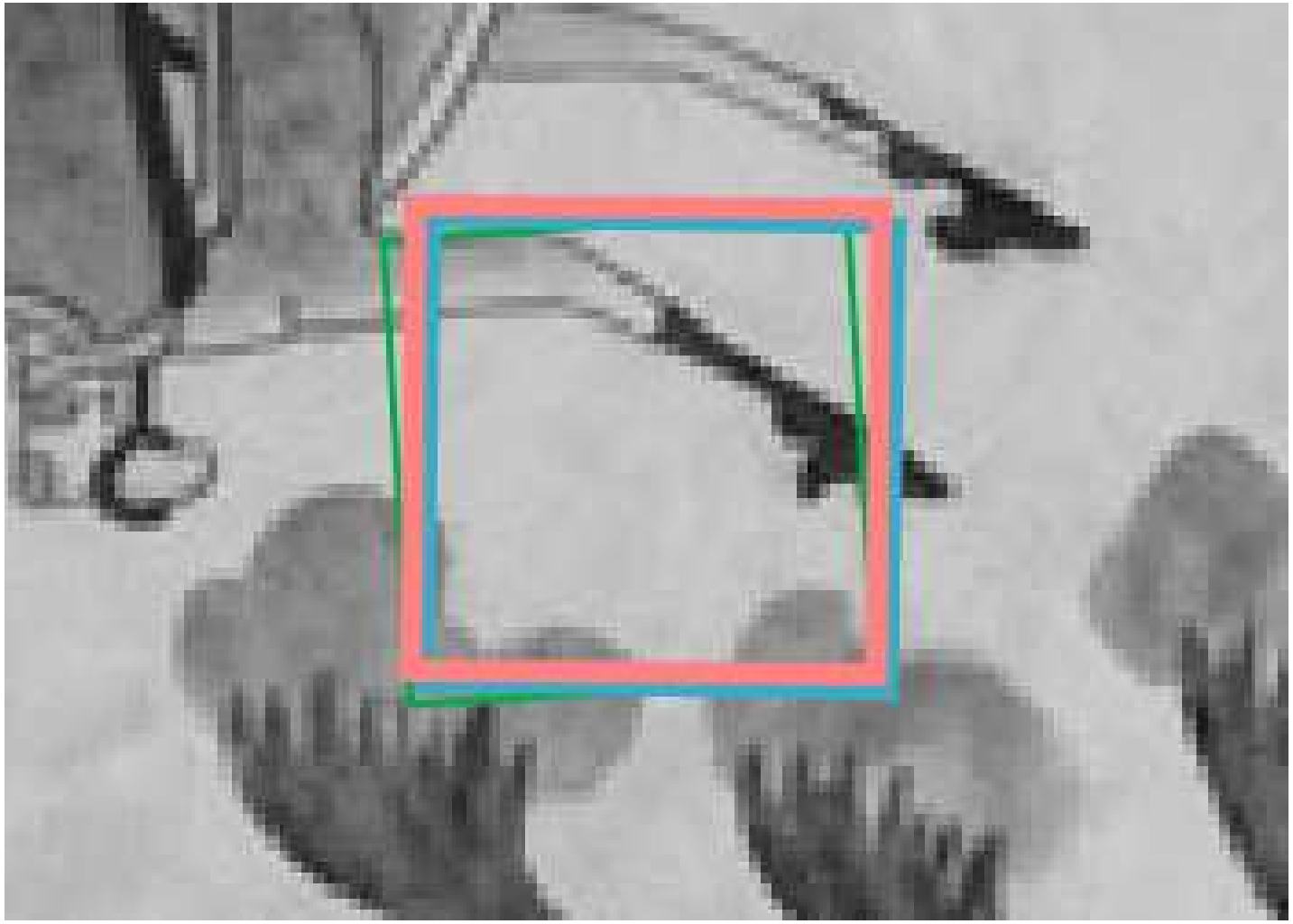
Tracking Filter Update from One Event



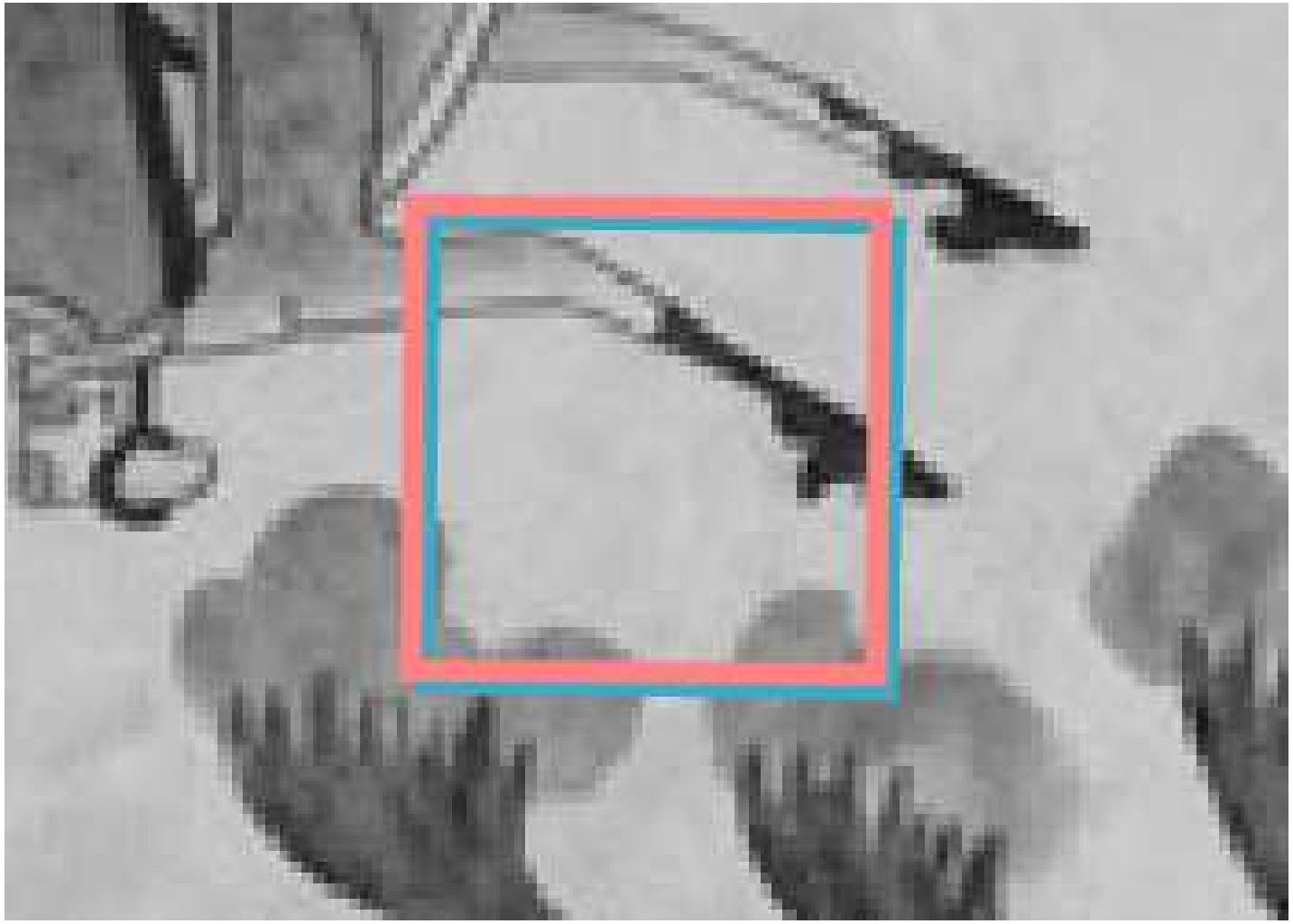
Tracking Filter Update from One Event



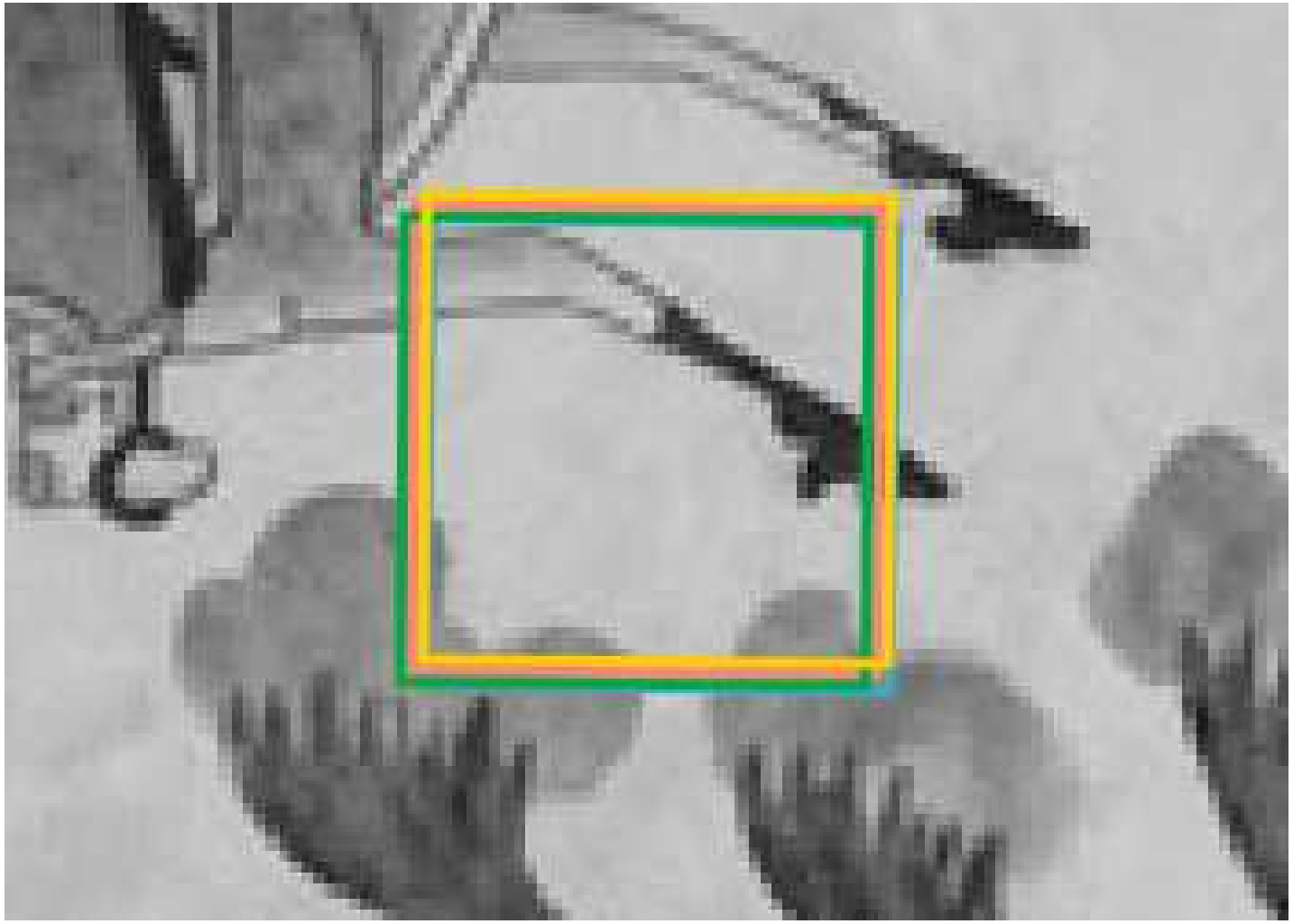
Tracking Filter Update from One Event



Tracking Filter Update from One Event

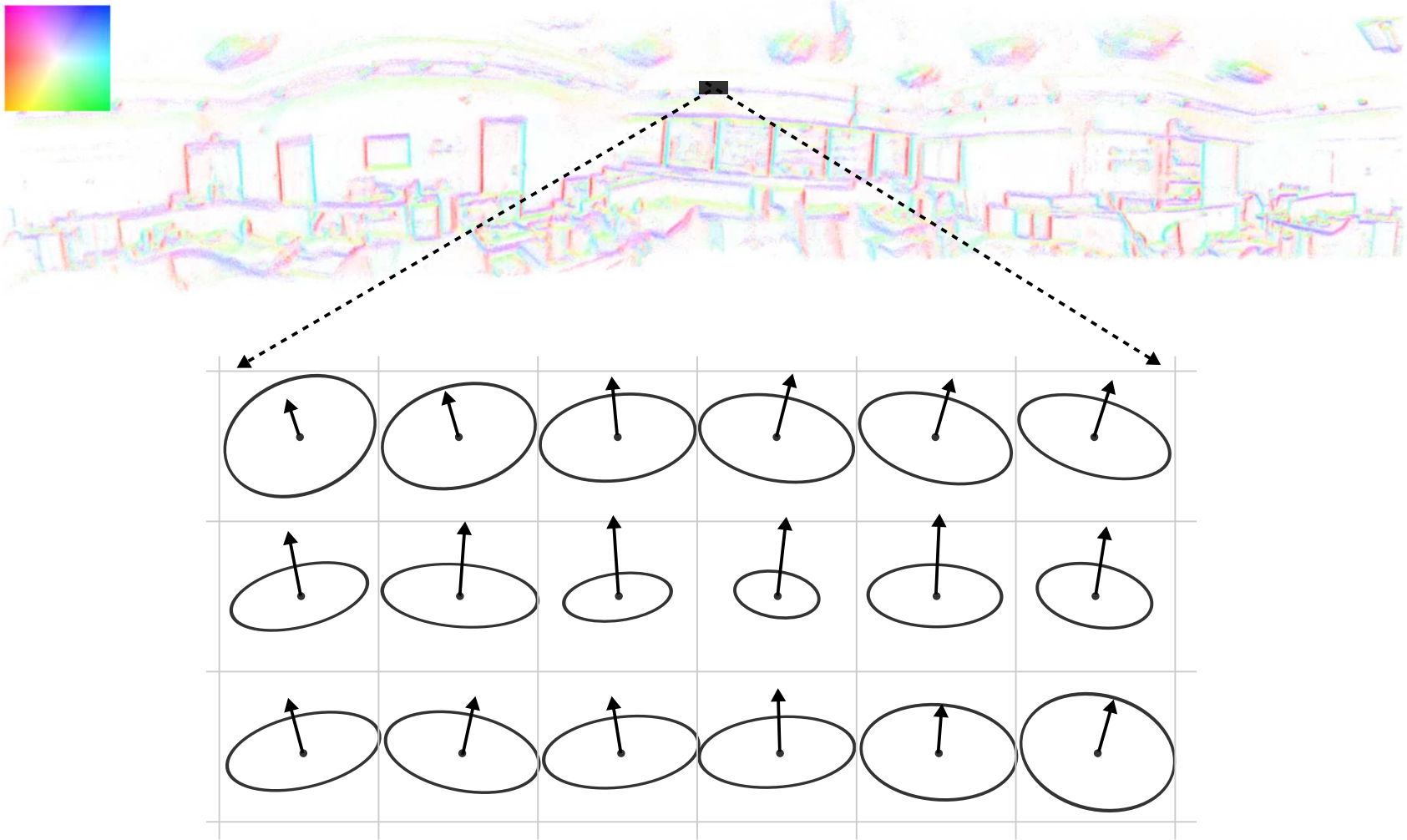


Tracking Filter Update from One Event

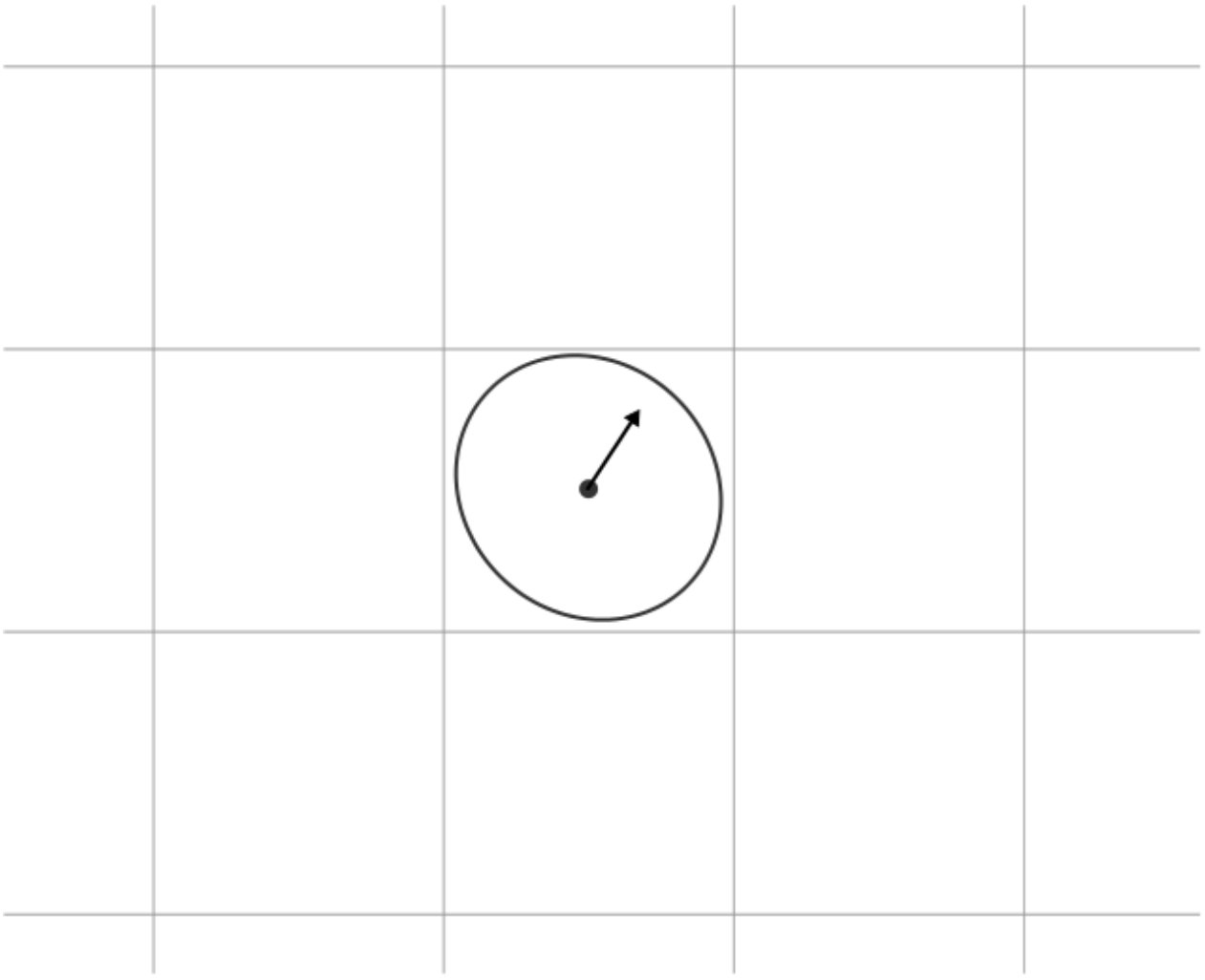


Gradient Estimation

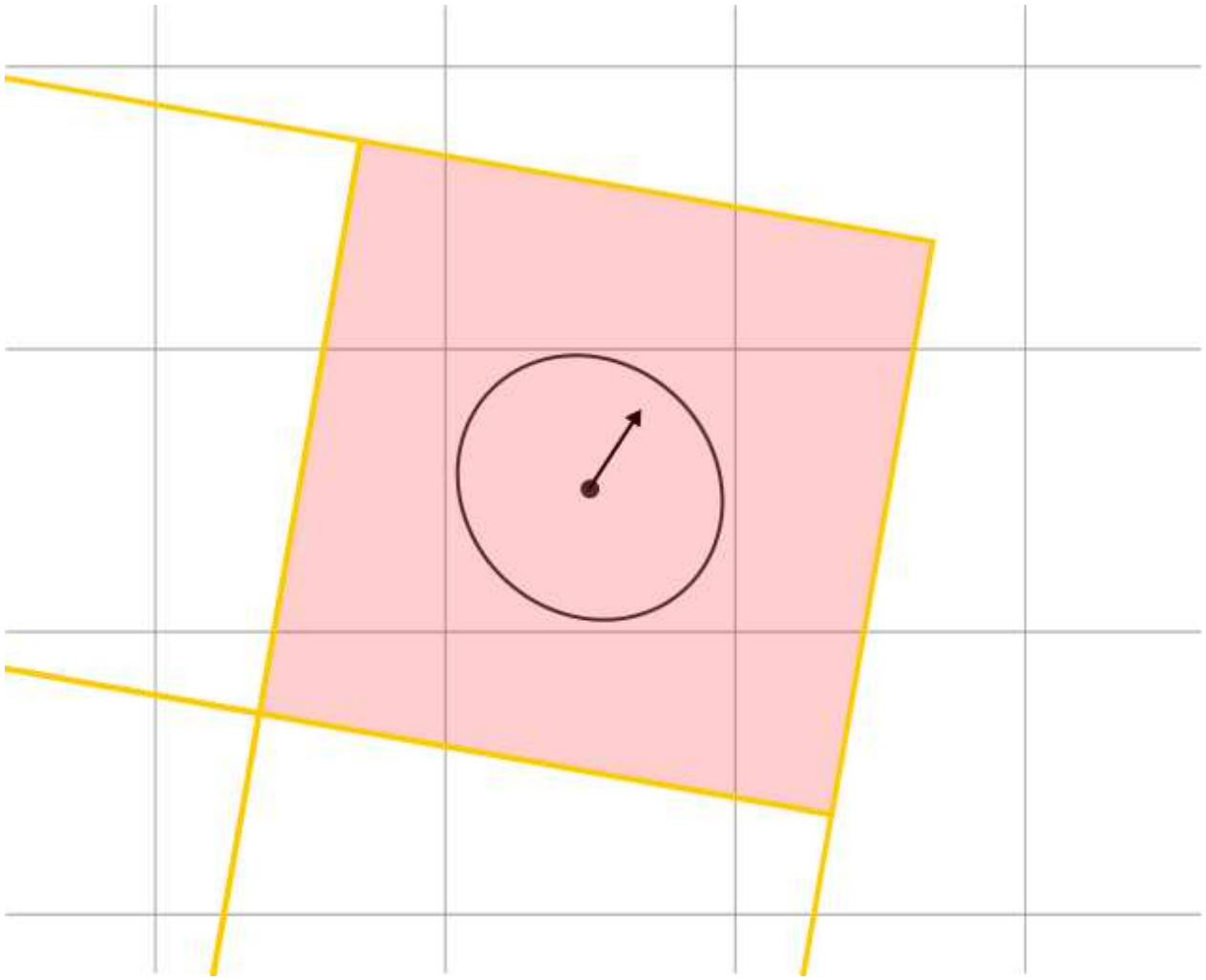
Pixel-wise EKF Gradient Estimation



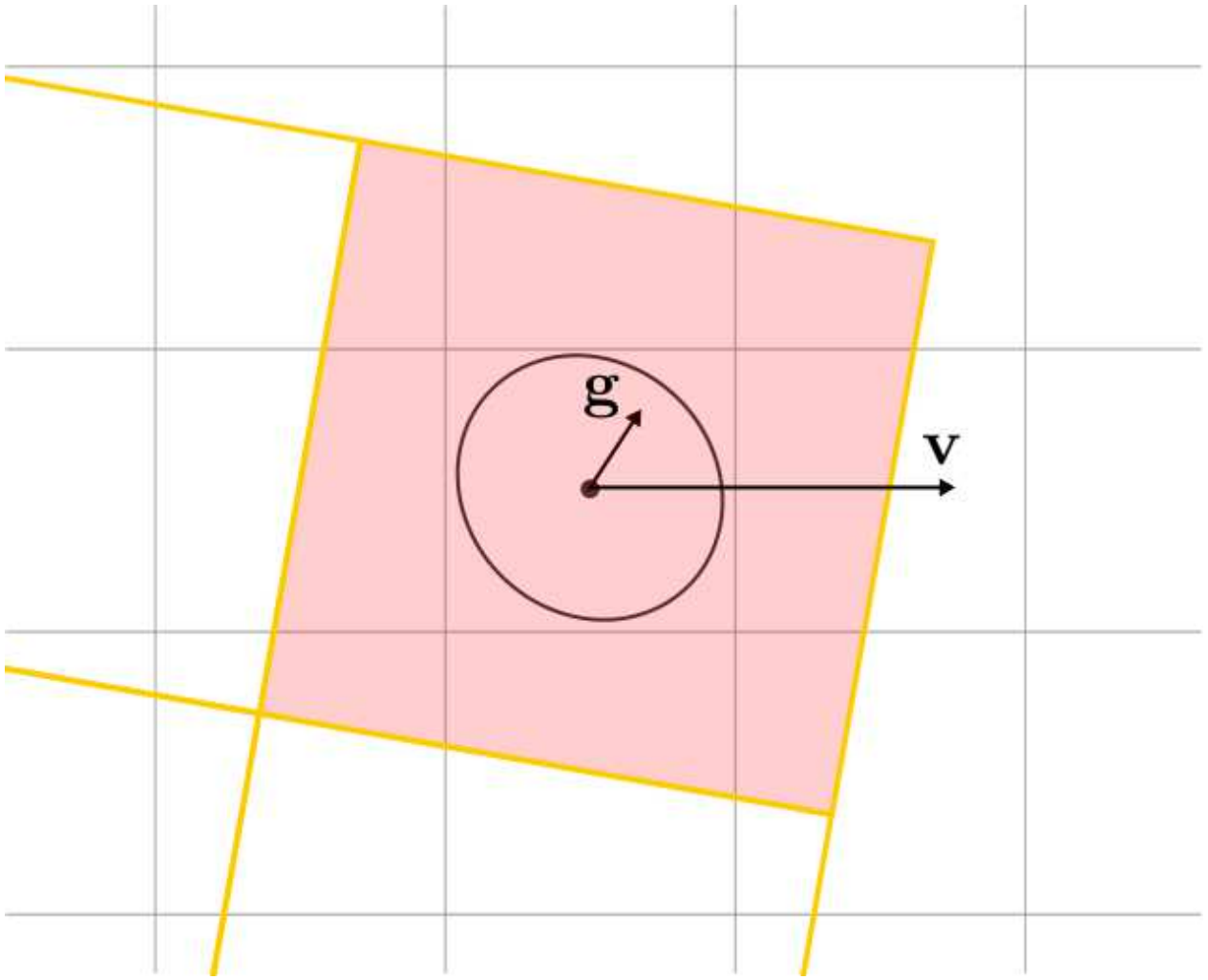
Pixel-wise EKF Gradient Estimation



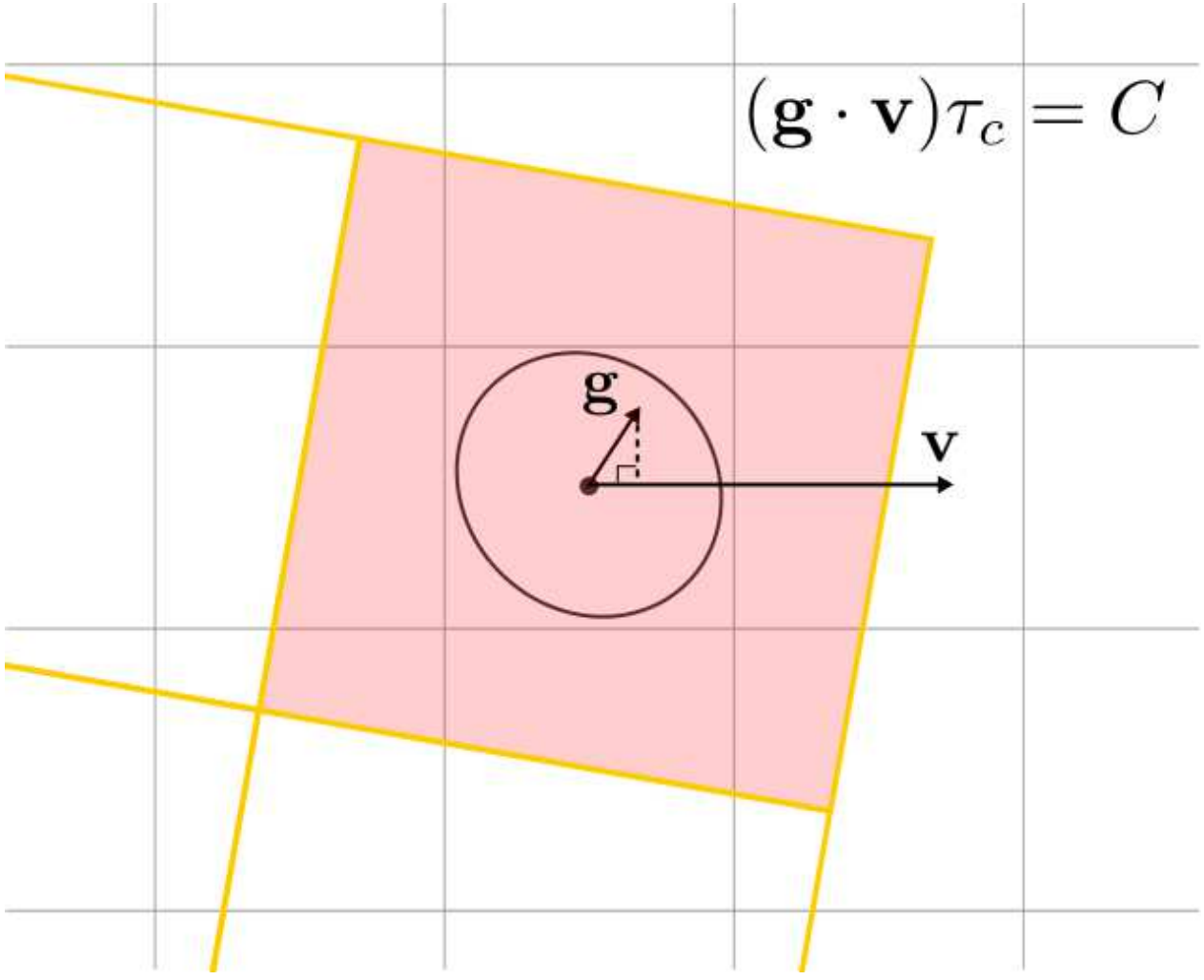
Pixel-wise EKF Gradient Estimation



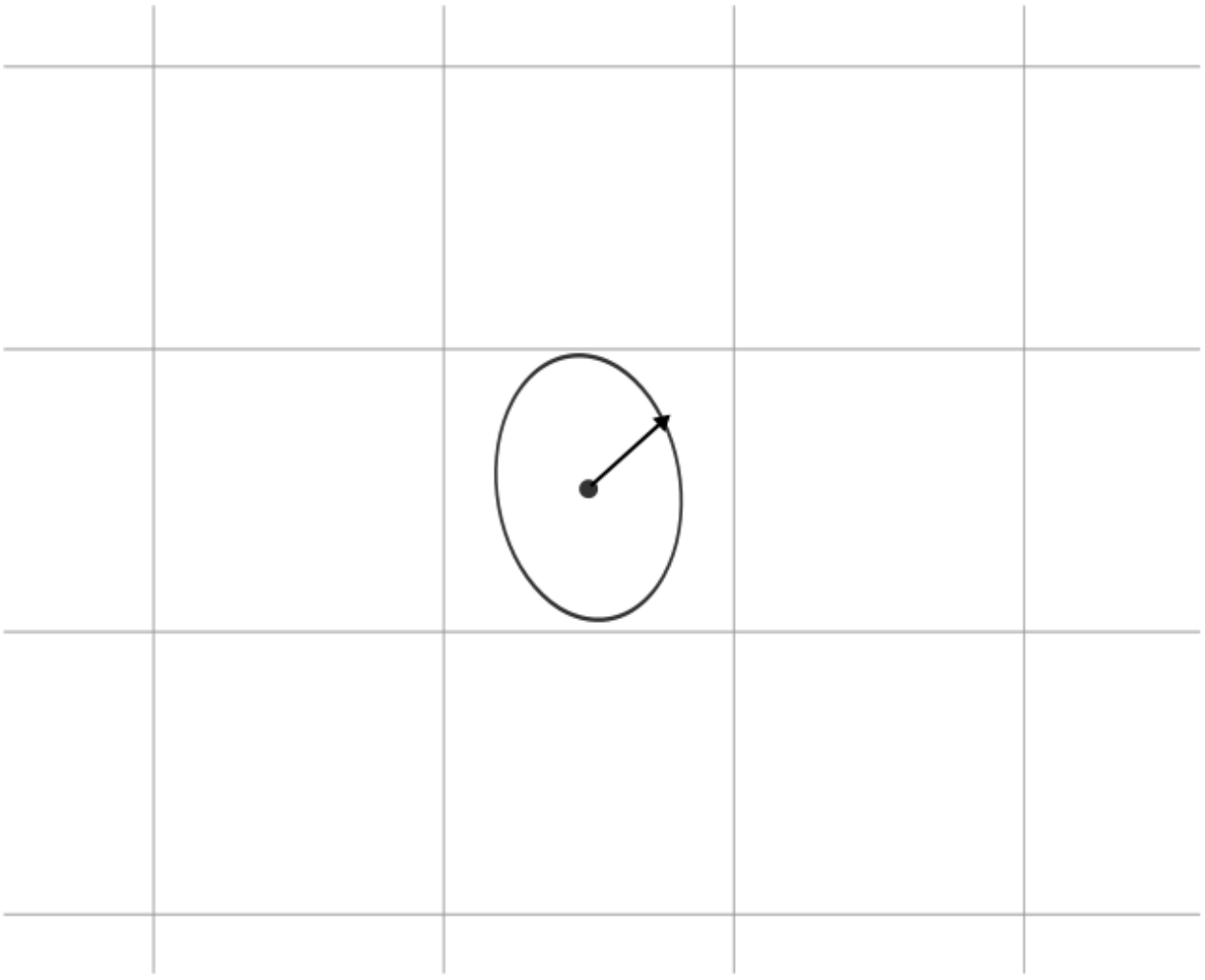
Pixel-wise EKF Gradient Estimation



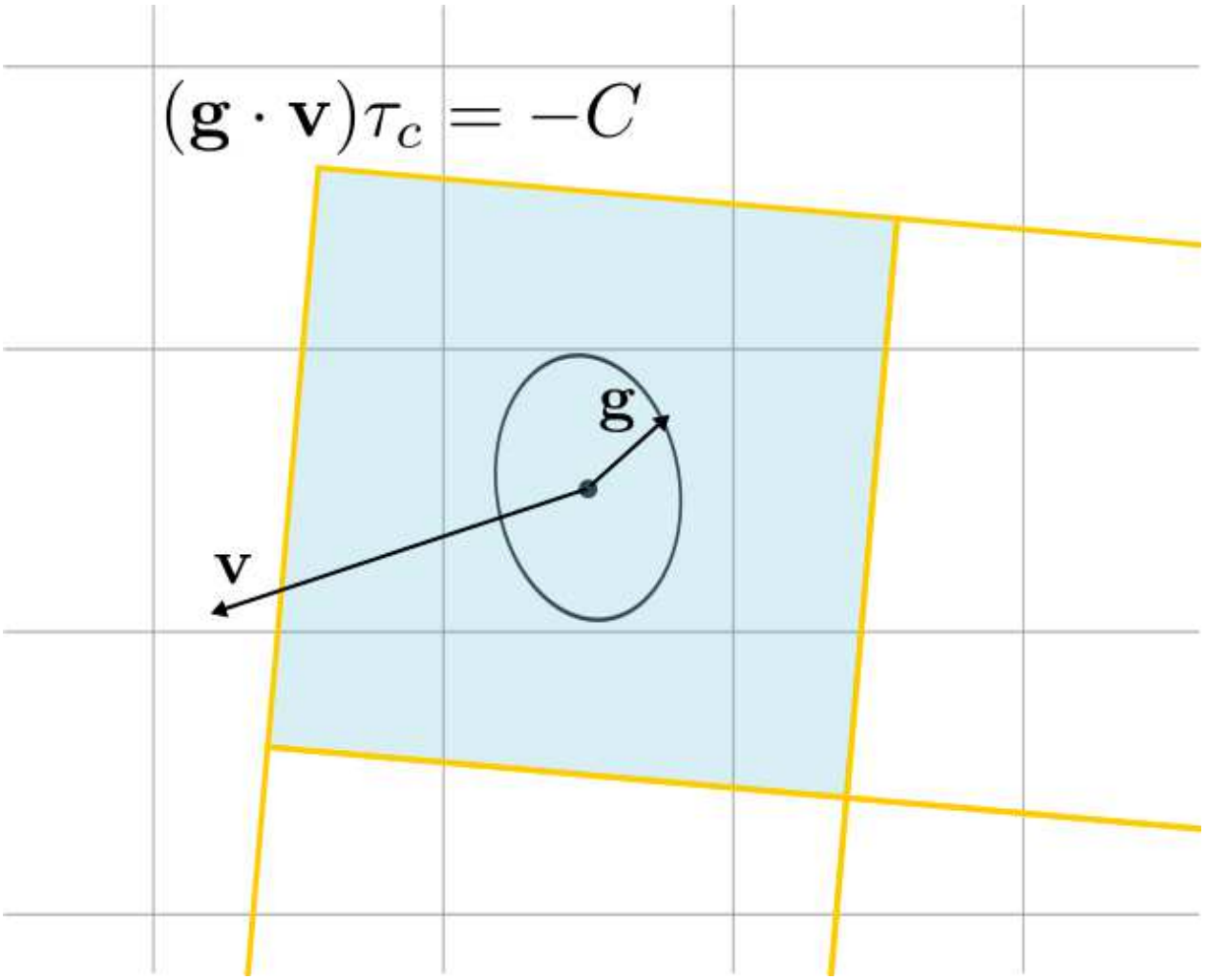
Pixel-wise EKF Gradient Estimation



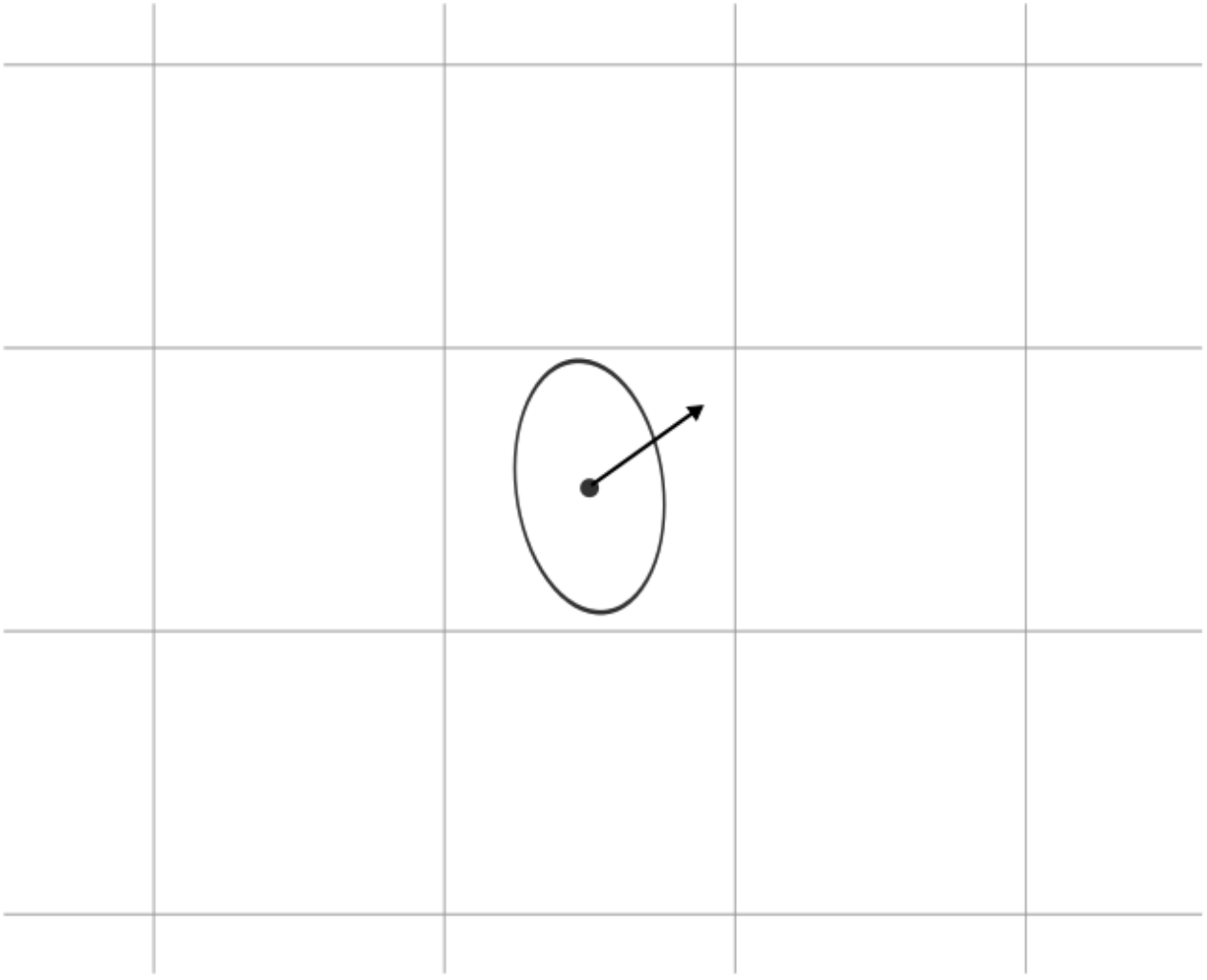
Pixel-wise EKF Gradient Estimation



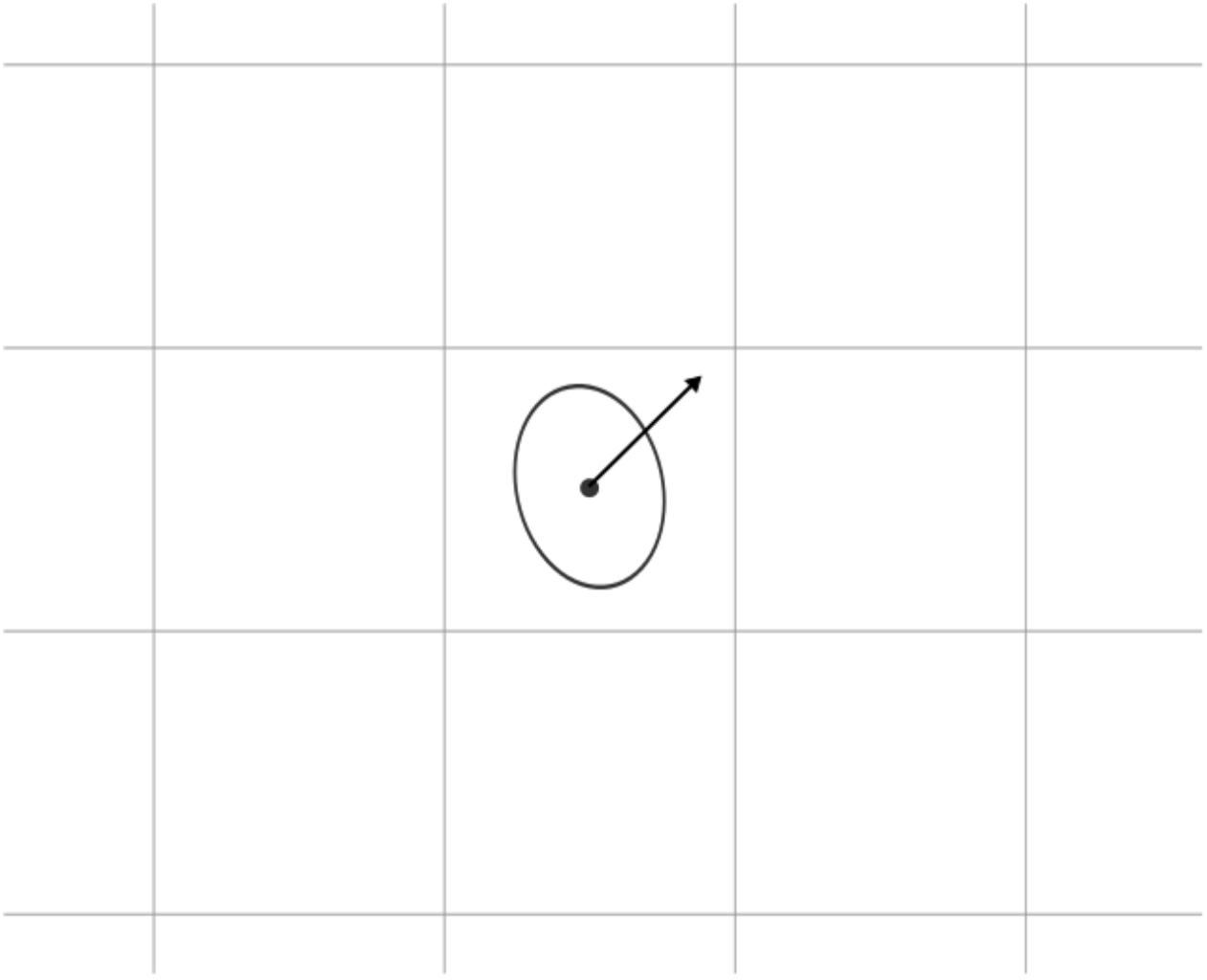
Pixel-wise EKF Gradient Estimation



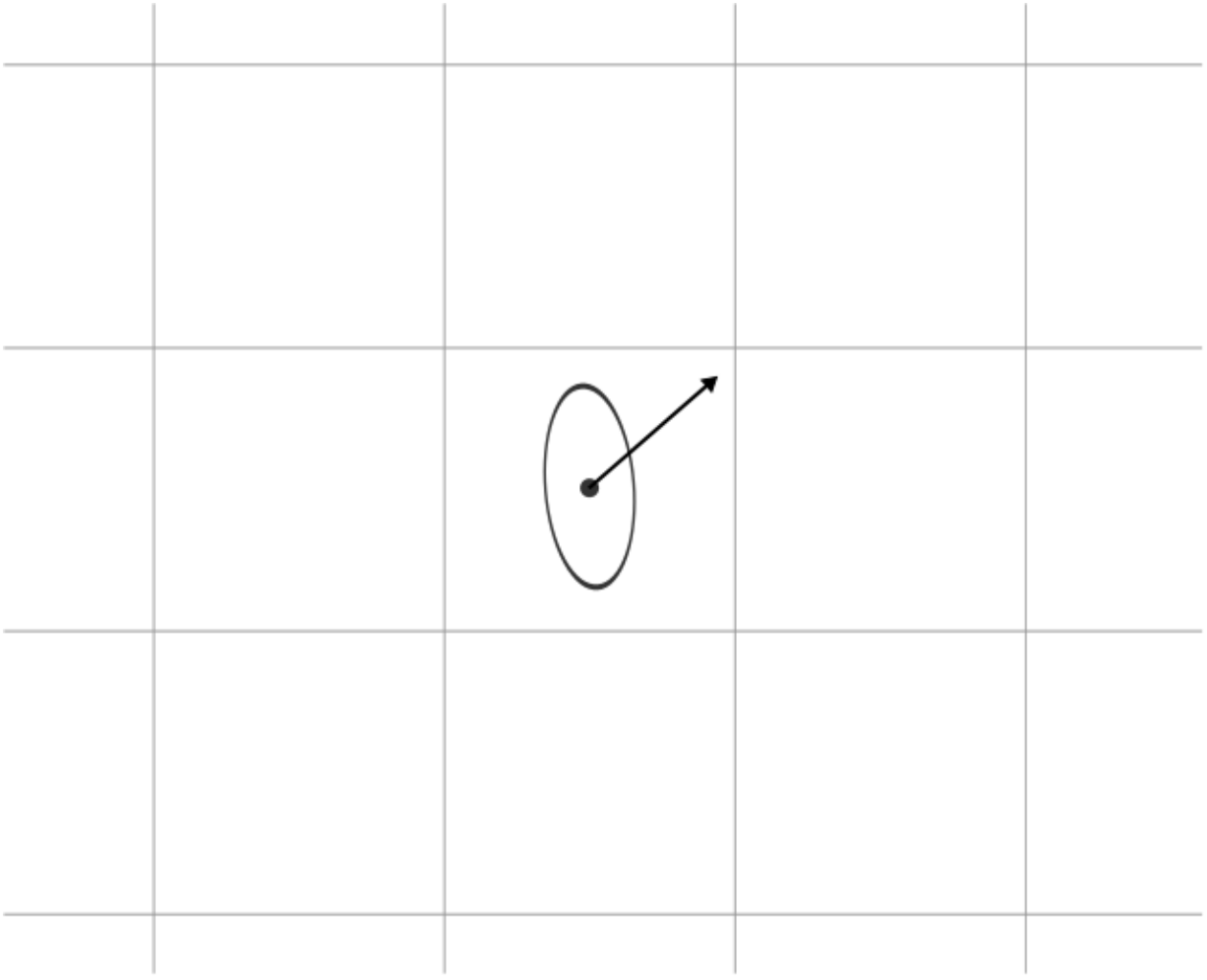
Pixel-wise EKF Gradient Estimation



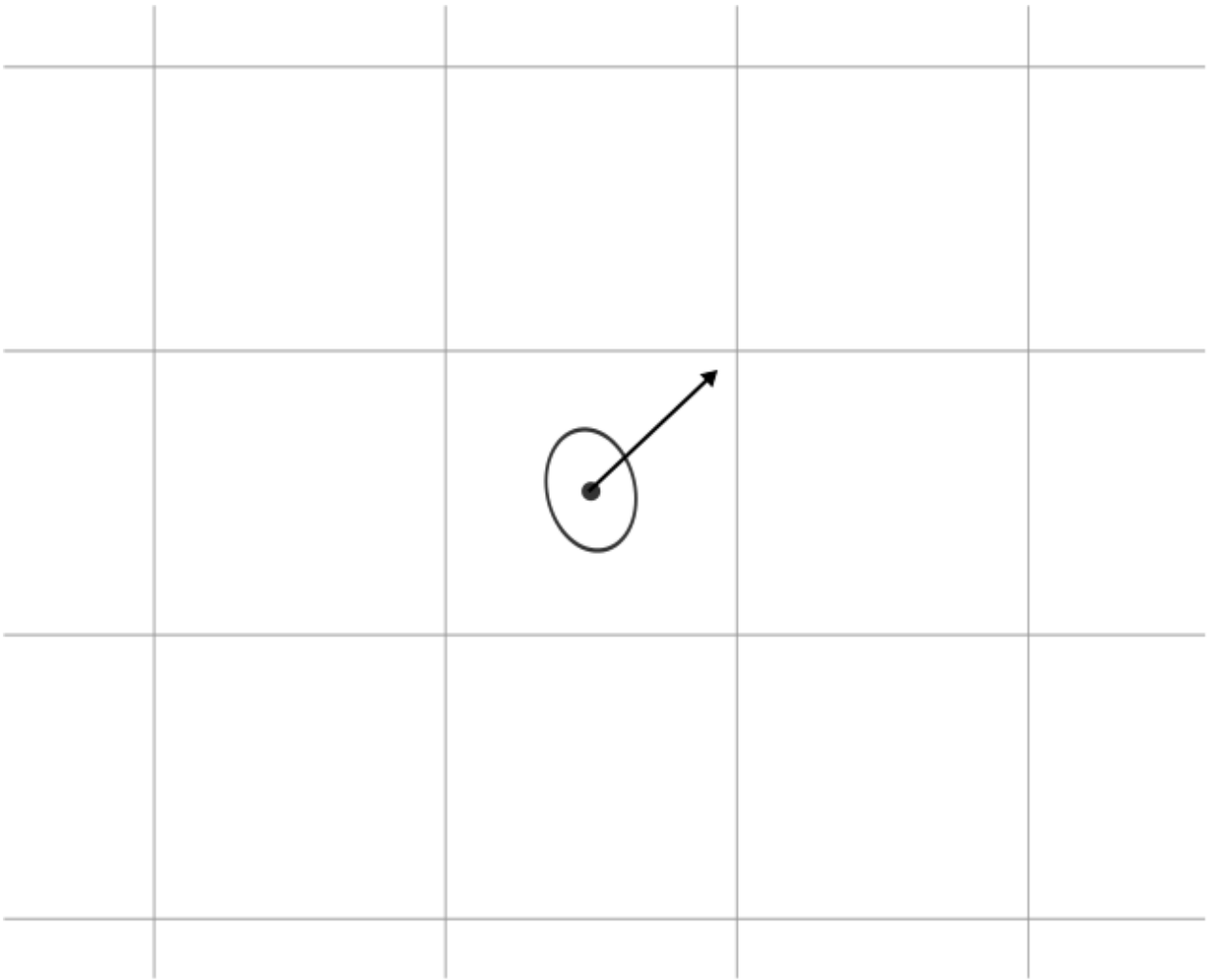
Pixel-wise EKF Gradient Estimation



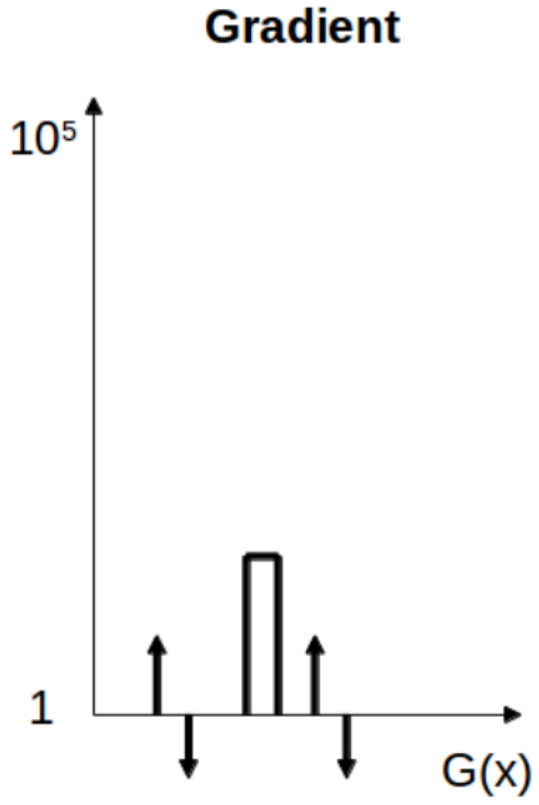
Pixel-wise EKF Gradient Estimation



Pixel-wise EKF Gradient Estimation

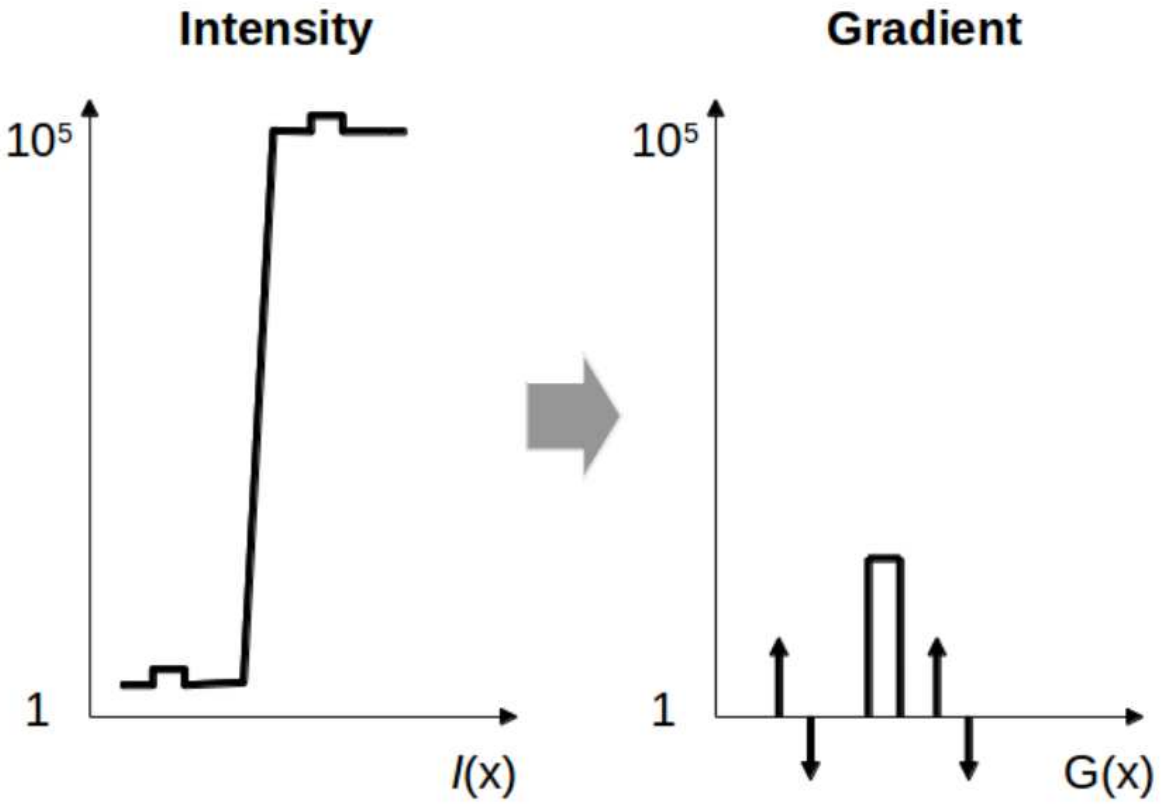


Reconstruction from Gradients in 1D



Agrawal, A. and Rasker, R., 2007

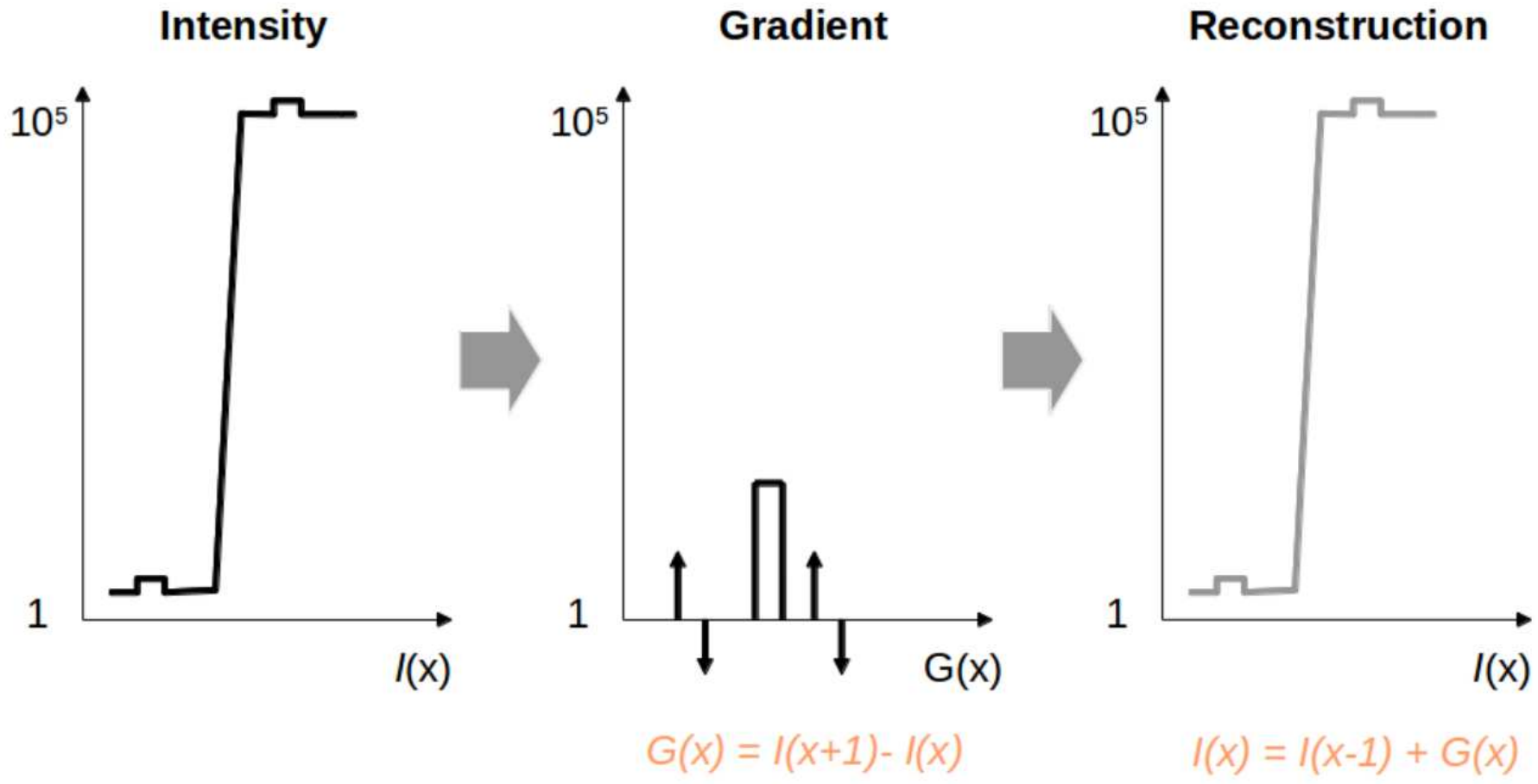
Reconstruction from Gradients in 1D



$G(x) = I(x+1) - I(x)$

Agrawal, A. and Rasker, R., 2007

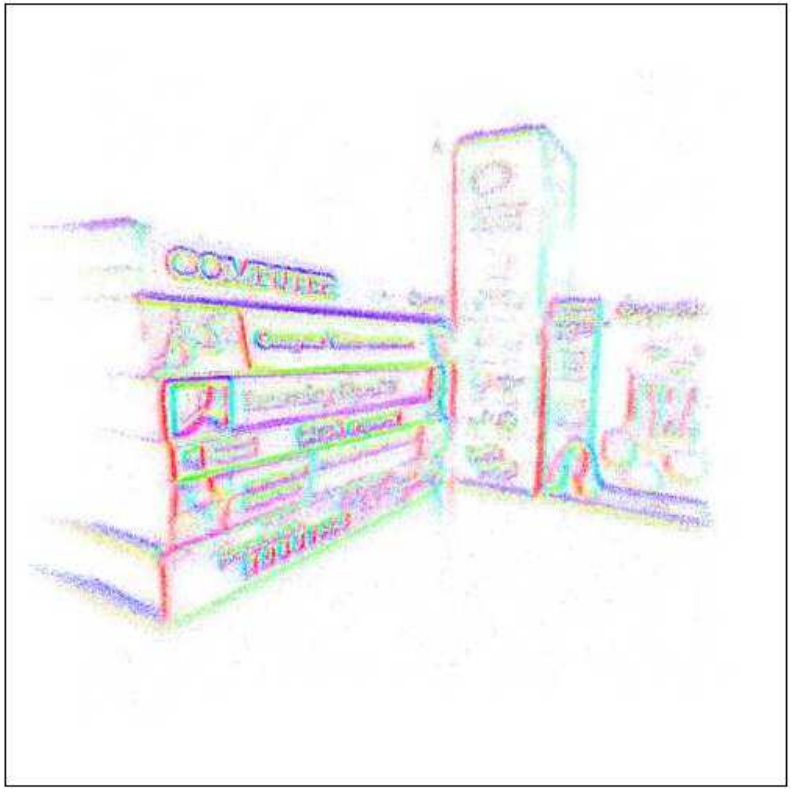
Reconstruction from Gradients in 1D



Agrawal, A. and Rasker, R., 2007

Intensity Reconstruction from Gradients in 2D

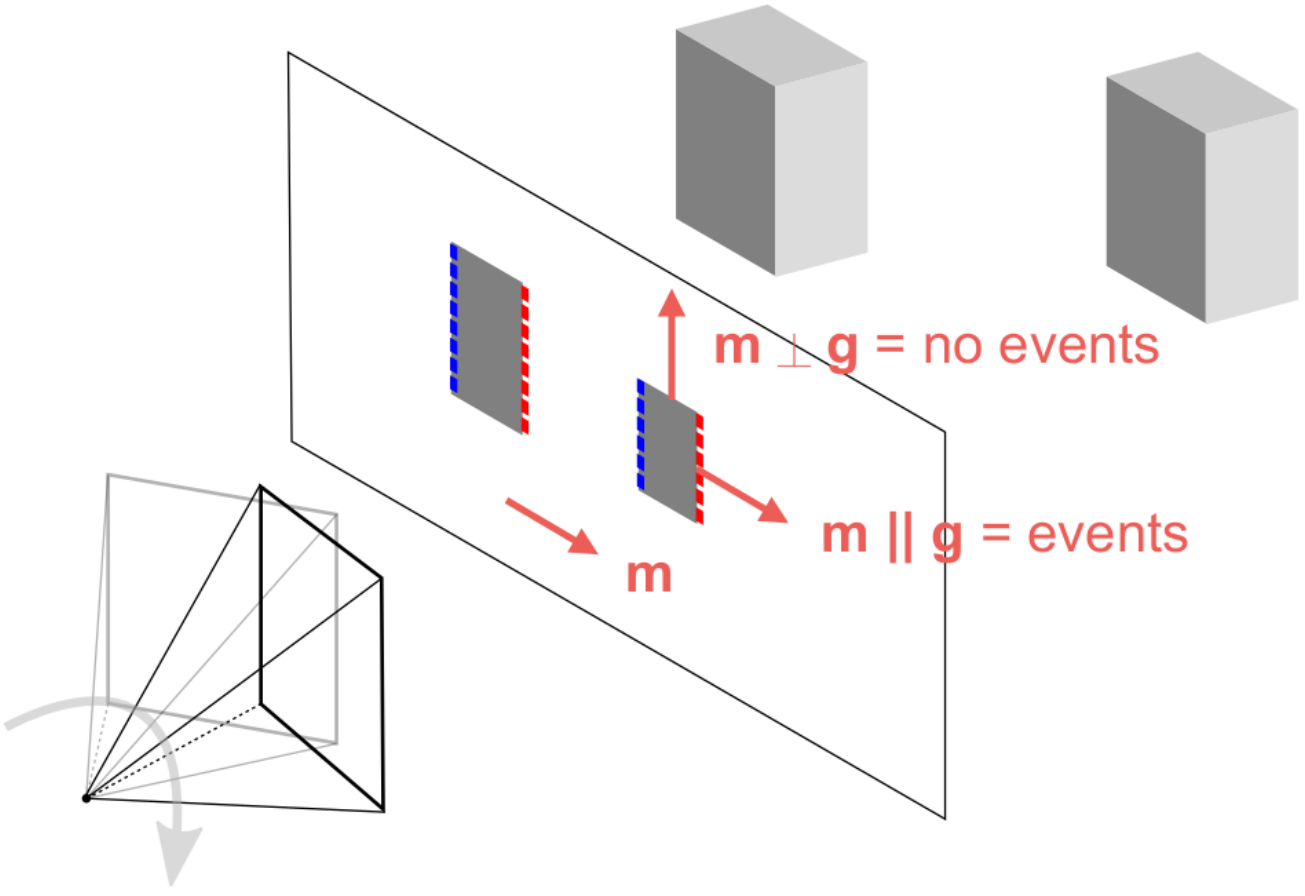
$$\min_{\mathbf{I}_l} \left\{ \int_{\Omega} ||\mathbf{g}(\mathbf{p}_k) - \nabla \mathbf{I}_l(\mathbf{p}_k)||^h_{\epsilon_d} + \lambda ||\nabla \mathbf{I}_l(\mathbf{p}_k)||^h_{\epsilon_r} \mathrm{d}\mathbf{p}_k \right\}$$



gradient estimation

intensity reconstruction

Events Induced by Pure Rotation Motion

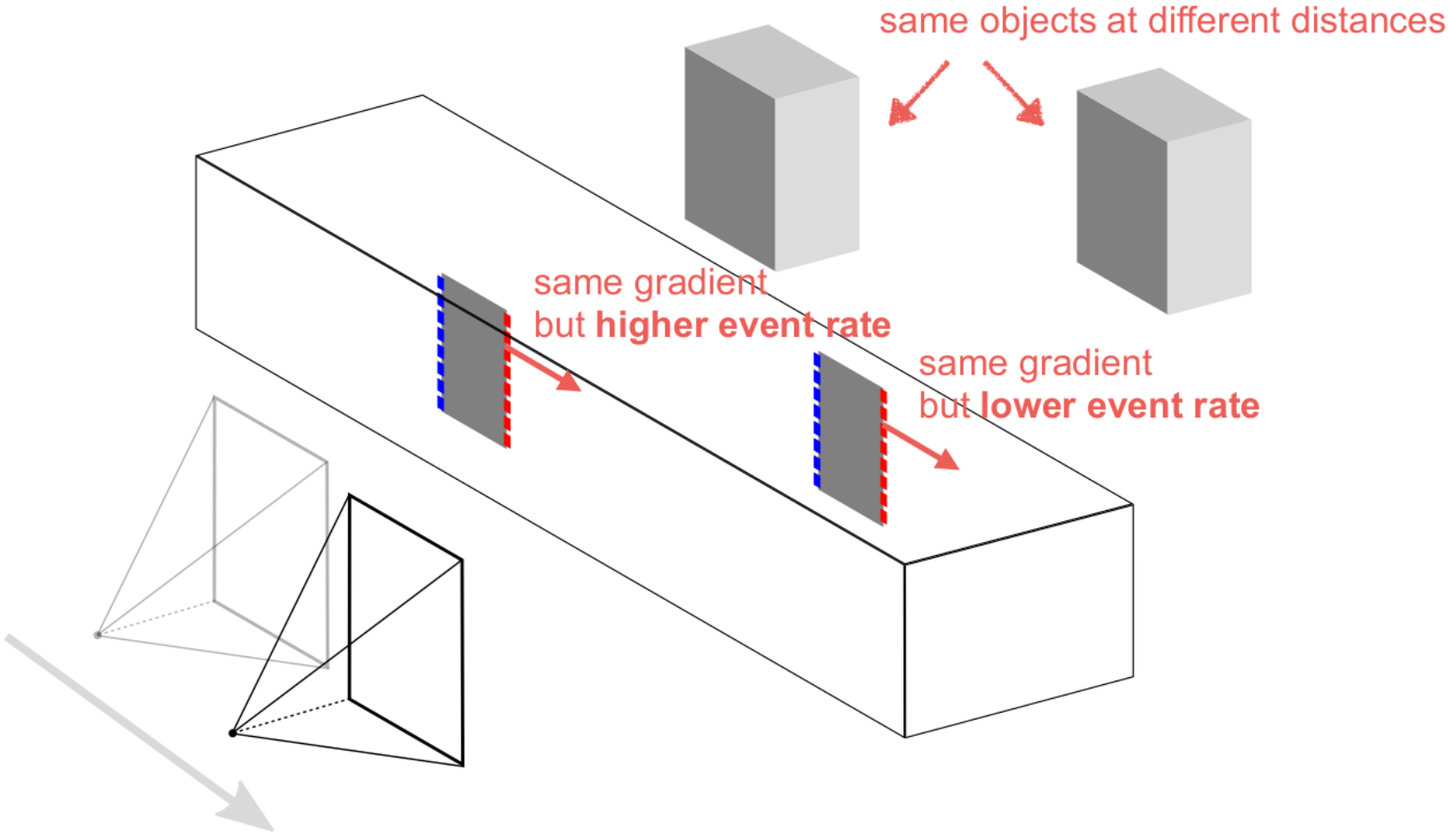


localisation
pure rotation motion

mapping
panoramic mosaic

3D scene
unknown & unstructured

Events Induced by Translation Motion

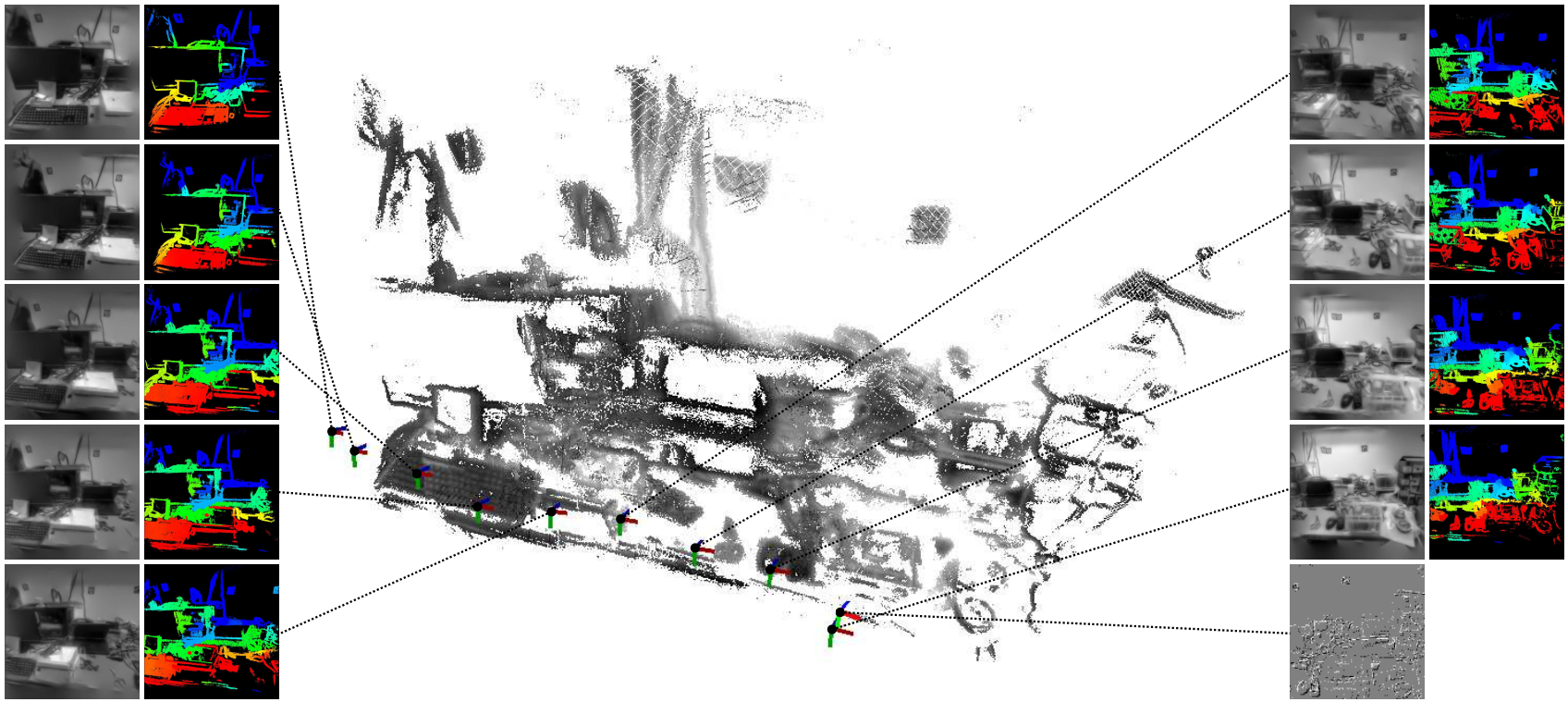


localisation
6-DoF hand-held motion

mapping
3D

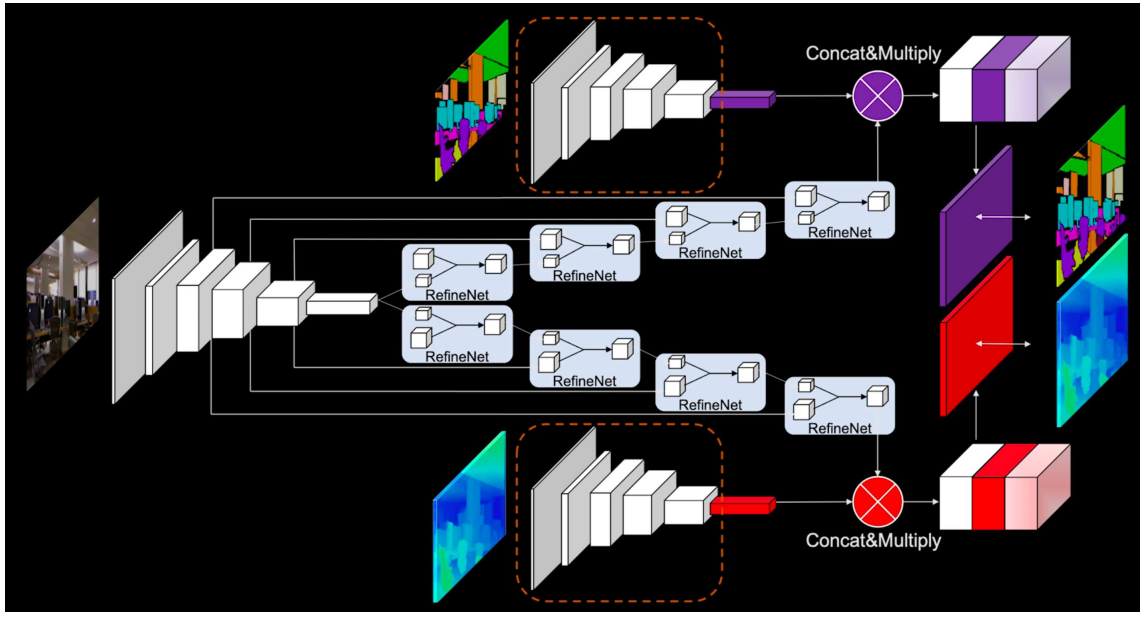
3D scene
unknown & unstructured

3D Motion, Structure and Intensity from Event Data



- Kim, Leutenegger, Davison, ECCV 2016 (Best Paper).

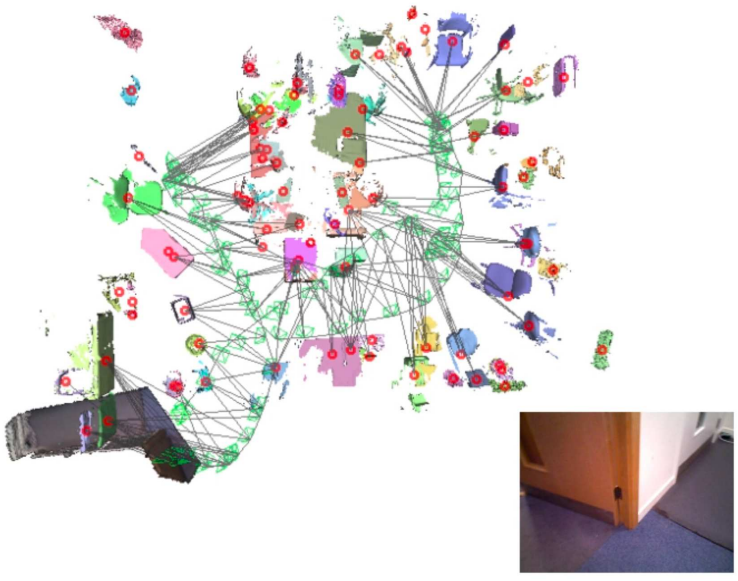
New Representations for Spatial AI



SceneCode, Zhi, Bloesch, Leutenegger, Davison, CVPR 2019.

- Small learned per-frame codes for depth and semantics allow coherent and efficient multi-view semantic fusion.

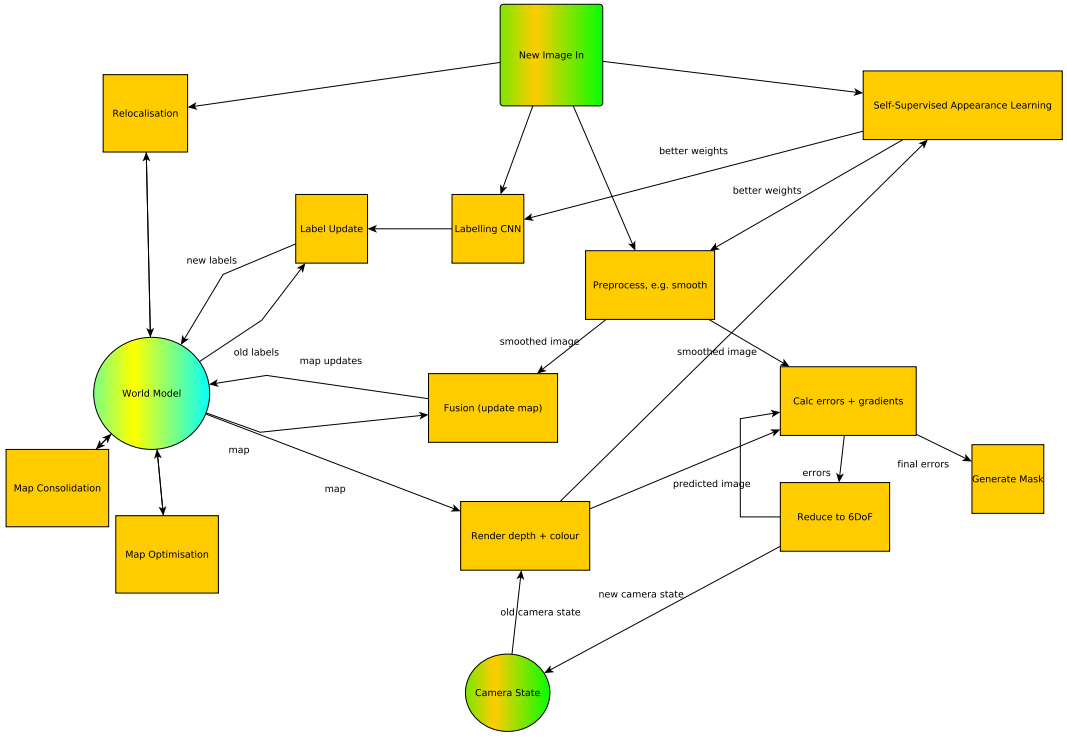
Fusion++: Volumetric Object-Level SLAM



Fusion++: McCormac, Clark, Bloesch, Davison, Leutenegger, 3DV 2018.

- A graph of objects with 6DoF transforms is the map, as in SLAM++.
- Objects are identified and segmented using Mask-RCNN.
- Object models are individually fused as in KinectFusion.

Semantic SLAM Computation Graph



Complicated, loopy storage, computation and communication.

- **This** (and more) is what we need to optimise to get to true Spatial AI. Learned and designed modules.
- Lines of attack: 1. Hardware; 2. Representation.

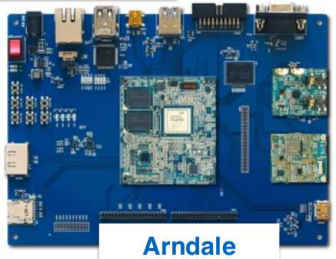
SLAM on a Range of Processors

TITAN
4998 GFLOPS
< 400 W

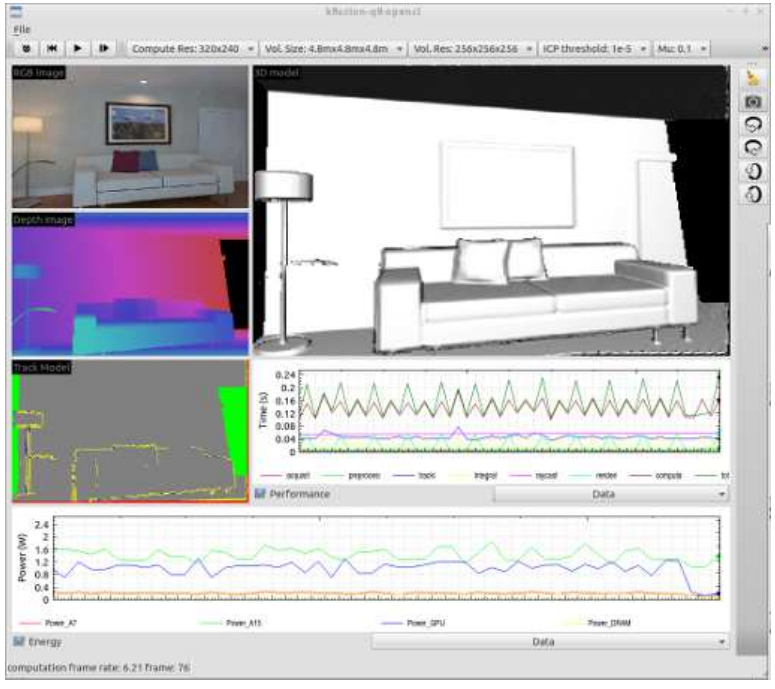
GTX 870M
2827 GFLOPS
< 100 W

TK1
404 GFLOPS
< 20 W

ODROID
170 GFLOPS
< 10 W

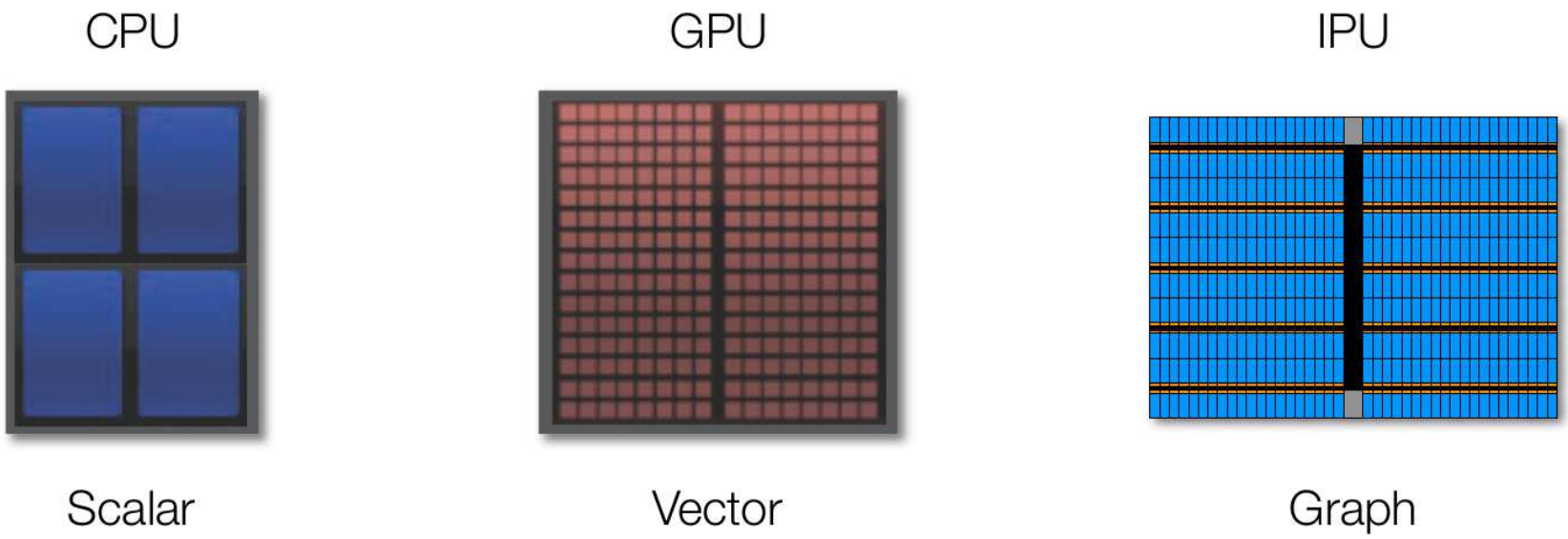


Arndale
87 GFLOPS
< 5 W



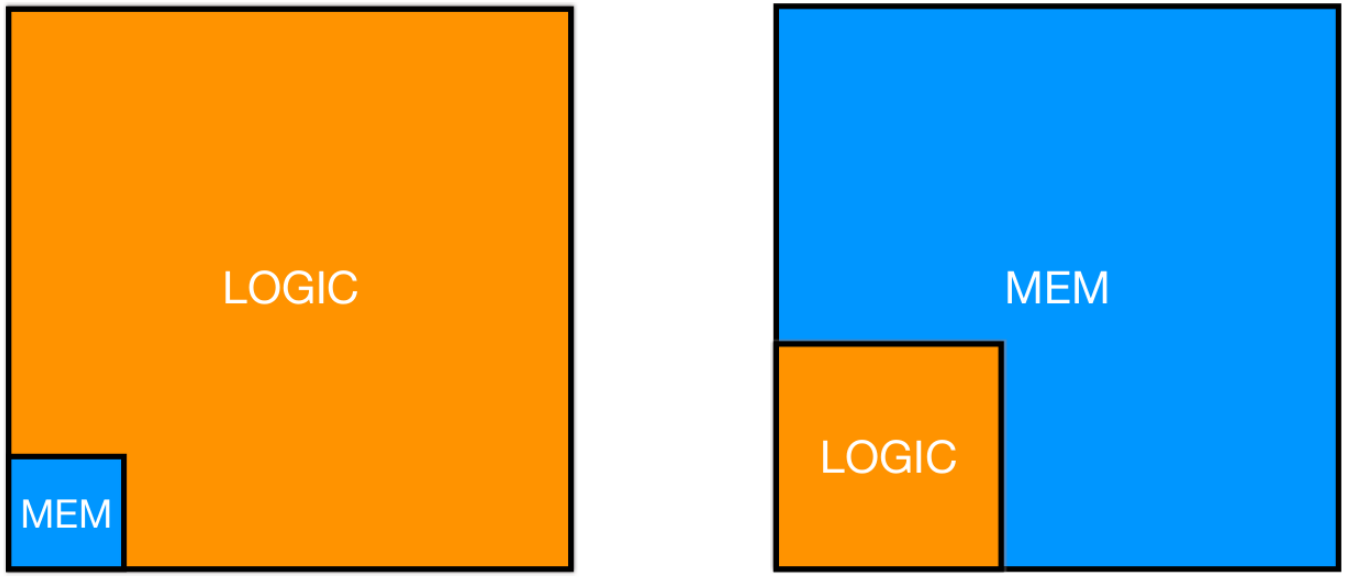
- SLAMBench (PAMELA Project, Universities of Manchester, Edinburgh and Imperial College). Opening up research in the joint development of real-time vision algorithms, programming tools and architecture.

Graph Processing (e.g. SpiNNaker, Graphcore)



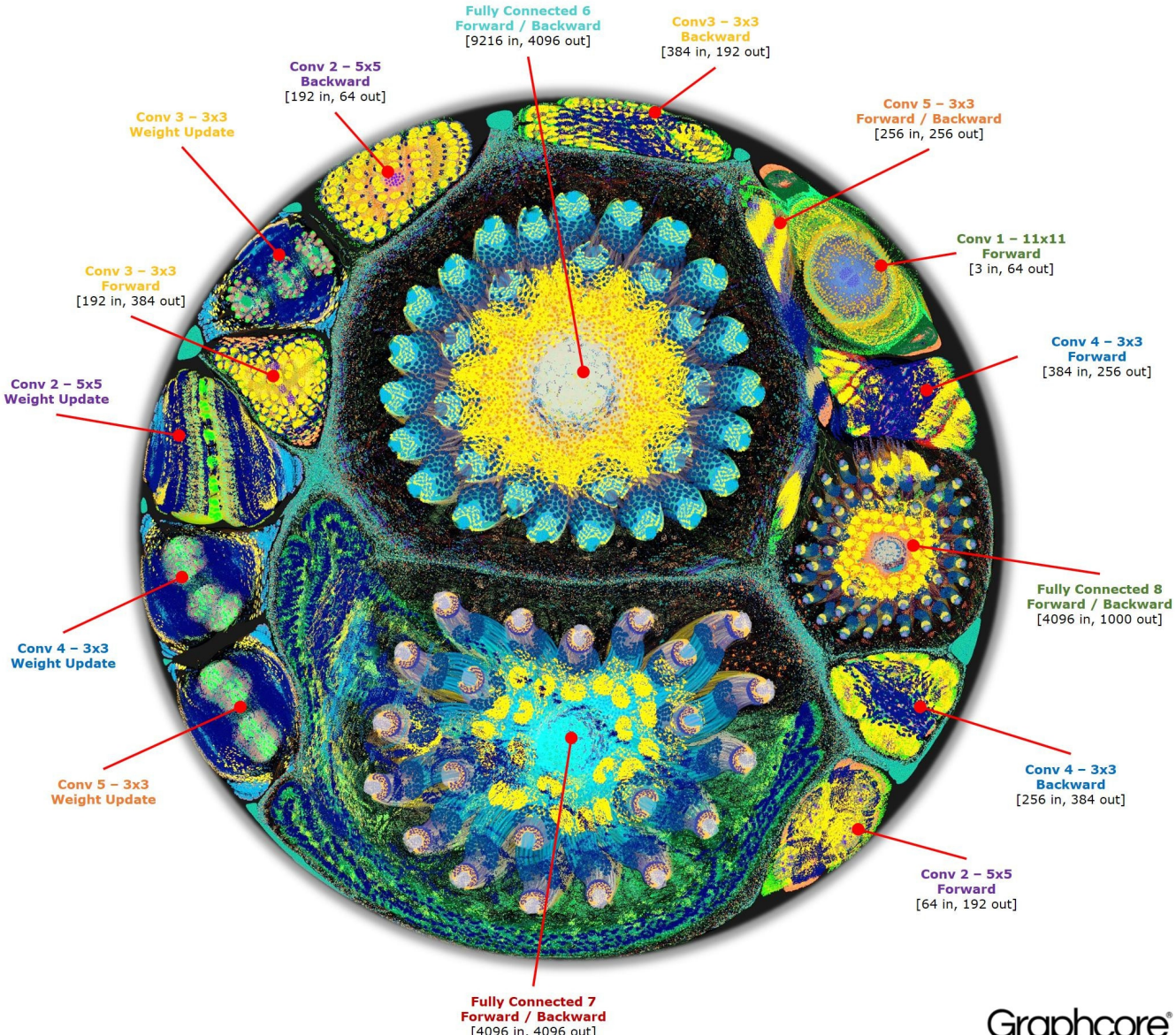
- AI, including vision, presents a new type of workload which suits neither CPUs nor GPUs: too much moving of data in and out of memory.
- Towards better use of silicon and power by algorithms bring together processing and local memory and communicate by message passing.

Graphcore's 'IPU' or Graph Processor



- Thousands of pure distributed scalar multiprocessors on a single chip (digital, synchronous).
- Memory should dominate the die to enable rapid, temporary, distributed communication. Memory uses only 2–10% of the power of logic.
- Other related projects (e.g. IBM Truenorth, Brainchip) are more explicitly neuromorphic.

Visualising the Processing Graph of a Neural Network



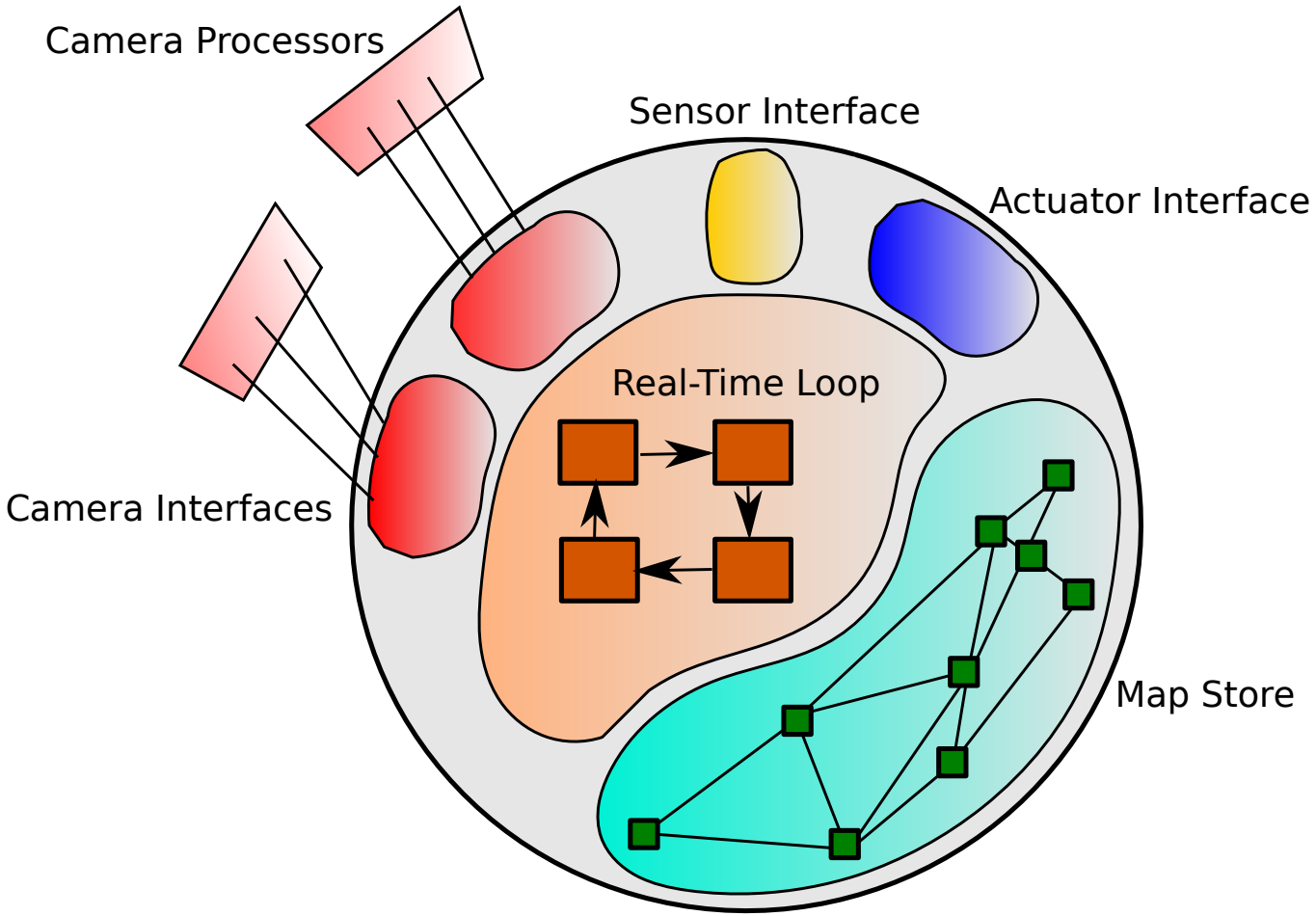
Graphcore

Finding the Graphs in Spatial AI

What are the graphs in a Spatial AI system?

- Image Graph (regular grid, usually).
- Map Graph (general locality-linked map structure; not necessarily simple connectivity with visual sensing).
- Main real-time processing graph.

‘Spatial AI Brain’!



Where to do the processing?

Close-to-Sensor Processing

- Local motion estimation (optical flow).
- Segmentation/labelling.
- Detecting features; abstracting image.
- Multi-purpose CNN.
- SIMT-type operations that could be extremely efficient on on in-plane processor or eventual stacked processors.

Real-Time Loop

- Prediction of views (rendering).
- Tracking against map (alignment).
- Local map update (geometry and label fusion).
- Mostly highly parallelisable, but usually with some reduction elements which currently run on the CPU.

Where to do the processing?

Map Store

- Feature clustering; object segmentation and identification.
- Loop closure detection.
- Loop closure optimisation.
- Map regularisation (smoothing).
- Unsupervised clustering to discover new semantic categories.
- Generally very graph-like operations.
- Difficult issue: connection between the real-time loop and the map is dynamic. Do we need special ‘router’ nodes?

Some experiments in distributable processing

Gaussian Belief Propagation

- Master representation is the factor graph; can be completely dynamic.
- Global entities can be estimated with completely local processing and storage and message passing.
- Linear, Gaussian GaBP converges to correct means, over-confident ellipses.
- Interesting progress towards non-linear and robust factor graphs.
- Implements recompute instead of store. New measures of algorithm performance needed.
- Paper coming soon!

Conclusions

Spatial AI Research!

- Efficient and usable 3D scene representations.
- Co-design of processors, sensors and algorithms.
- Graph-based algorithms for estimation (e.g. Belief Propagation) and machine learning (e.g. Graph Networks).
- 'FutureMapping: The Computational Structure of Spatial AI Systems', A. J. Davison, arXiv:1803.11288, 2018.

Dyson Robotics Lab at Imperial College

- Elite academic lab with openings for post-docs and PhD students.

SLAMcore: applied Spatial AI

SLAMCORE

- London-based venture-backed startup.