Reconstruction, Motion Estimation and SLAM from Events

Andrew Davison

Robot Vision Group and Dyson Robotics Laboratory
Department of Computing
Imperial College London
www.google.com/+AndrewDavison

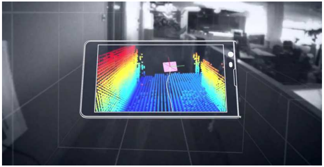
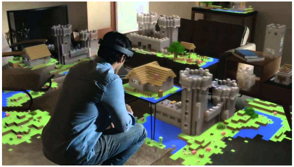
June 2, 2017

Visual SLAM and General Spatial Perception

- Simultaneous Localisation and Mapping: how does a robot or device understand the space it is moving through using on-board sensors?
- My focus: a single camera in a small area; real-time, closed loop systems (MonoSLAM 2003).

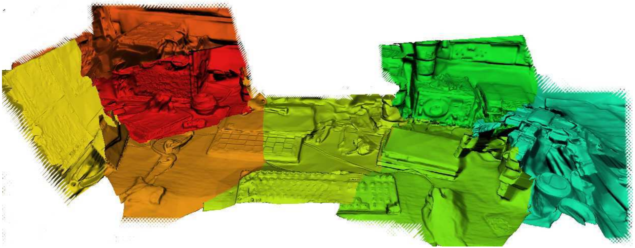
- Initial emphasis on localisation as an output; now increasingly dense mapping and semantic understanding.
- SLAM is evolving towards general real-time spatial perception, a crucial layer for AI or IA, (but it's still SLAM!)

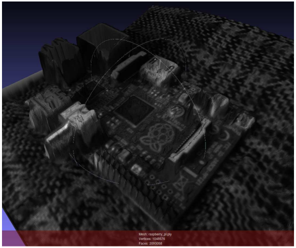
Modern Products and Systems

		
Dyson 360 Eye	Google Project Tango	Microsoft HoloLens

- Positioning and sparse/semi dense reconstruction now rather mature. . . and entering real products.
- But much is still to be done to enable much more capable and widespread devices.

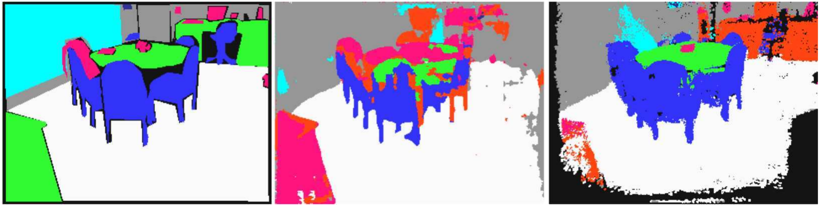
Dense SLAM





- ICCV 2011: DTAM (Dense Tracking and Mapping), Newcombe, Lovegrove and Davison. Dense mapping alternating with dense tracking (every pixel).
- 3DV 2016: Height Map Fusion with Dynamic Level of Detail, Zienkiewicz, Tsiotsios, Davison, Leutenegger.

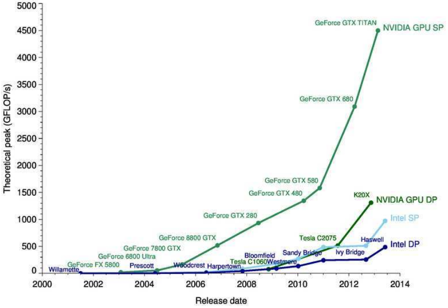
Towards Real-Time Semantic Labelling



- ICRA 2017: SemanticFusion: McCormac, Handa, Davison, Leutenegger.

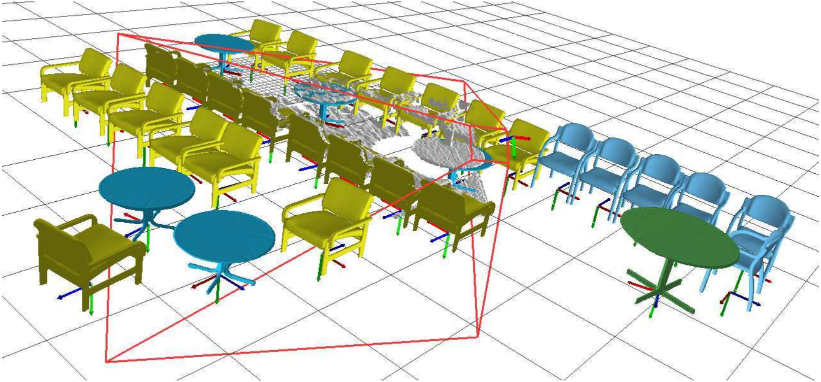
Brute Force Vision

- Rising processing allows increasingly computationally expensive computer vision algorithms to be brought into play in robot vision.
- Bundle adjustment; image retrieval; regularised dense reconstruction; CNNs.
- However... real embedded applications need low power, compactness and real-world robustness and **usefulness**.



Towards Pure Object-Level SLAM

- SLAM++ (Salas-Moreno, *et al* CVPR 2013): bring object recognition to the front of SLAM, and directly build a map at that level to benefit from strong predictions immediately.



- Predict, measure, update will be even stronger with object or even whole scene priors.

The Future of SLAM

What is the perception capability we need?

- Always on and aware geometric and semantic perception of everything of importance to complete a task.
- Hypothesis: a recognisable SLAM system building a persistent, (metric) representation will still be useful.
- Evaluate via task-oriented performance measures.

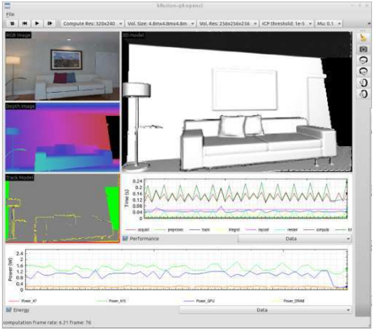
Methodology

- Many components of the SLAM system learned rather than designed, but overall architecture to remain familiar.
- Model-based, closed loop prediction and inference to enable efficiency and robustness.

What are the Constraints of Real Products in AI or IA?

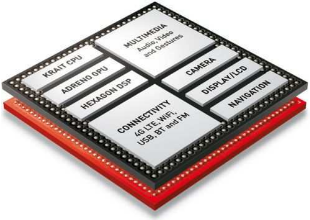
- Power usage; sensor/processor size, complexity and cost.

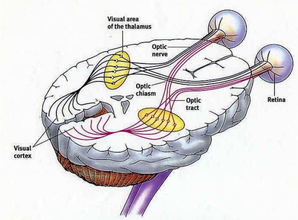
The Need for Efficiency in Advanced Real-Time Vision



- 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.
- Looking towards unified design of algorithms, processors. . . and sensors.

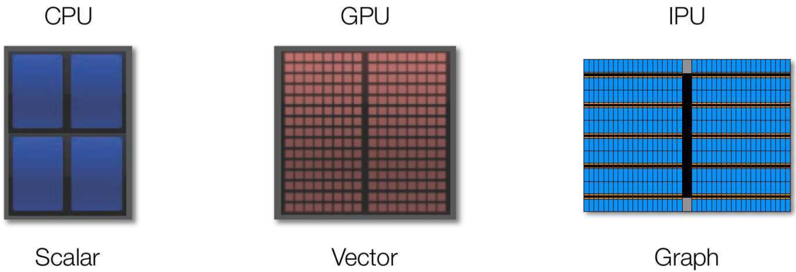
Future Embedded Vision





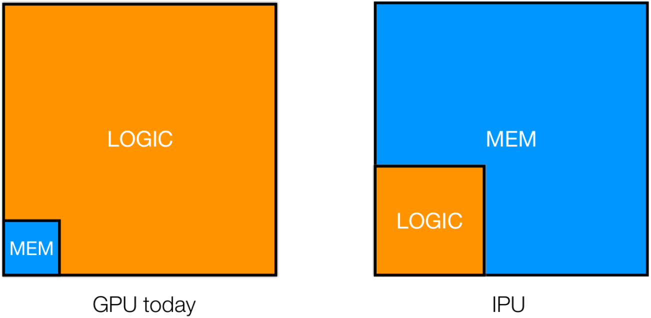
- Smartphone system-on-chip technology will provide the template for low power smart devices in the near term — and computer vision will be a major driver.
- CPUs, GPUs and increasingly specialised application-specific 'ASIC' chips and low power vision processors (e.g. Movidius).
- But how will we achieve always on operation in tiny devices with all day battery life?
- I believe that the long-term way forward is to increasingly look to neuromorphic principles.

Graph Processing (e.g. SpiNNaker, Graphcore)



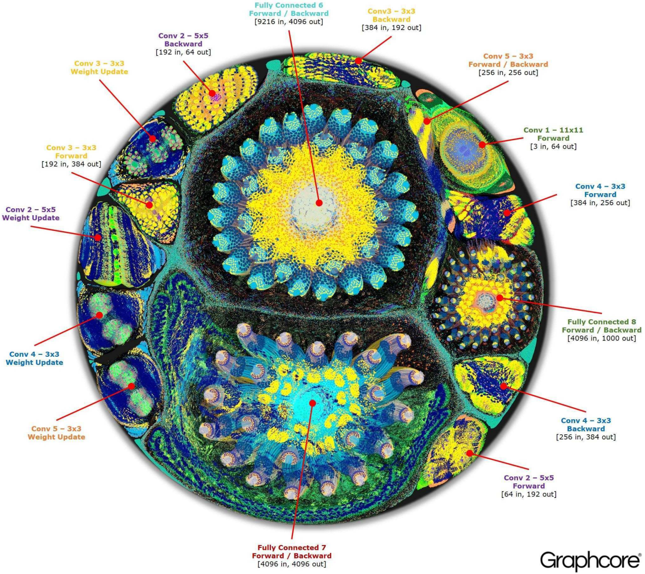
- AI, including vision, presents a new type of workload which suits neither CPUs nor GPUs.
- Sparse graph data structures and message passing algorithms.

Graphcore's 'IPU' or Graph Processor



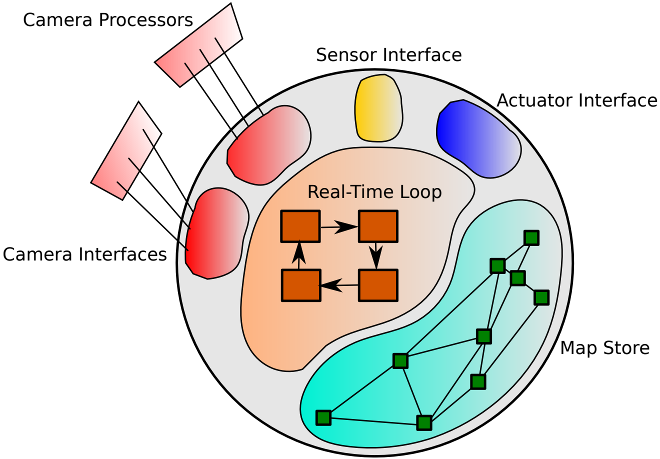
- Thousands of pure distributed scalar multiprocessors on a single chip (digital, synchronous).
- Memory should dominate the die to enable rapid, temporary, distributed communication. Memory uses only 2–10% of the power of logic.
- Other related projects (e.g. IBM Truenorth, Brainchip) are more explicitly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

Visualising the Processing Graph of a Neural Network



Graphcore®

Computational Structure of the Robot Vision Problem



- What are the graphs in a generic SLAM system?
- Components could be based on estimation or learned, but the computational structure would be similar.

Our Approach to Event-Based SLAM

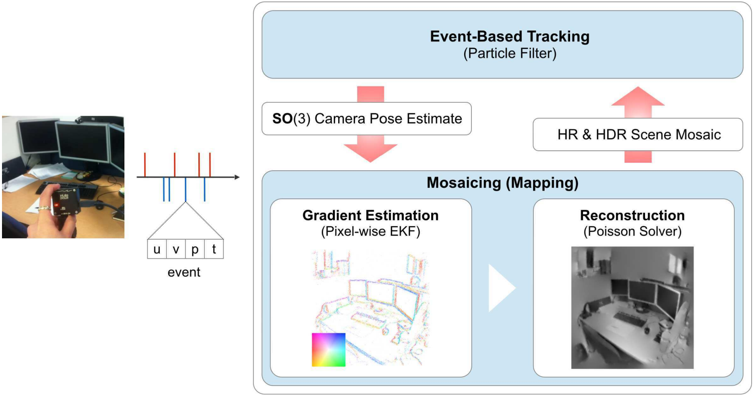
Generative

- 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

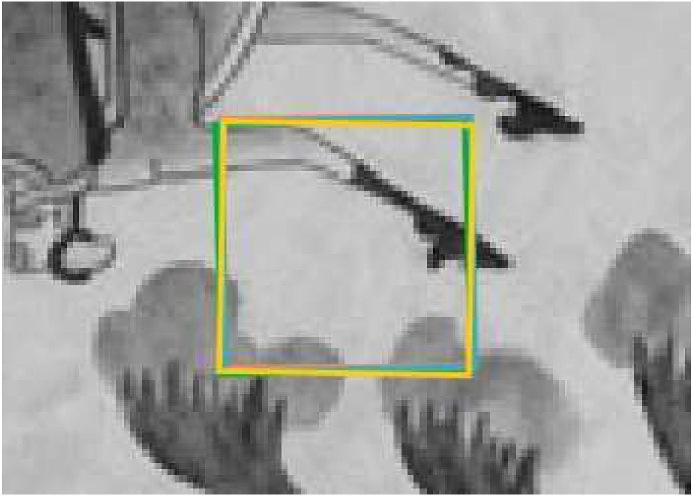
- Purely event-based; minimise latency; use filtering methods.
- Dealing with very high event rates one by one is tough on a CPU, but this problem should go away with future integrated sensor/processor architectures.

Simultaneous Mosaicing and Tracking from Events

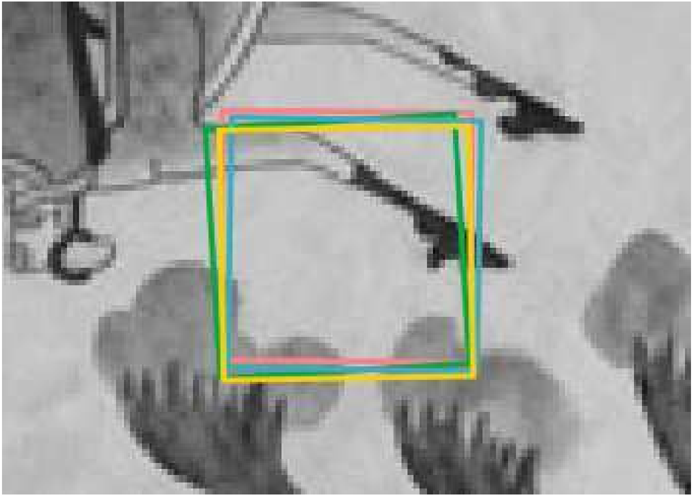


- Alternating filters to estimate tracking and dense scene gradient (upgraded to intensity). Using DVS128.
- Kim, Handa, Ieng, Benosman, Davison, BMVC 2014 (Best Industry Paper).

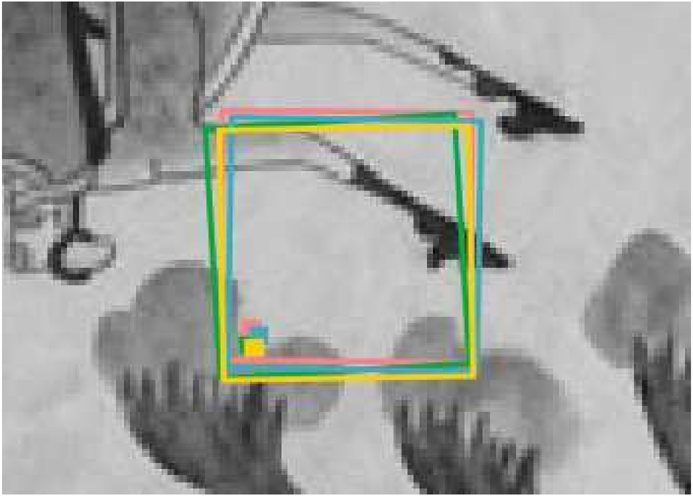
Tracking Filter Update from One Event



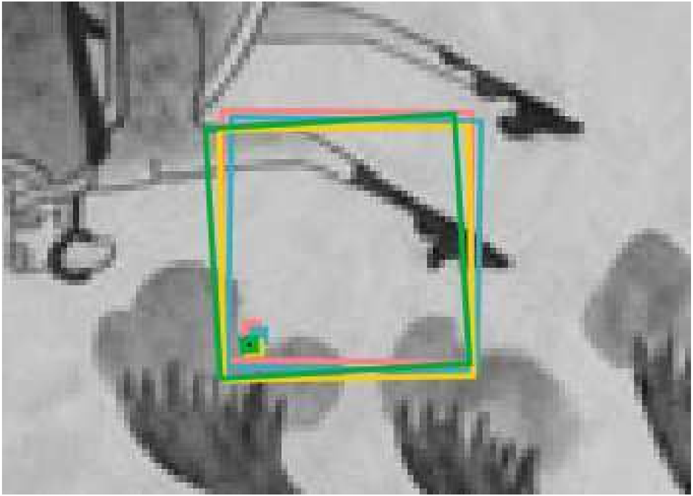
Tracking Filter Update from One Event



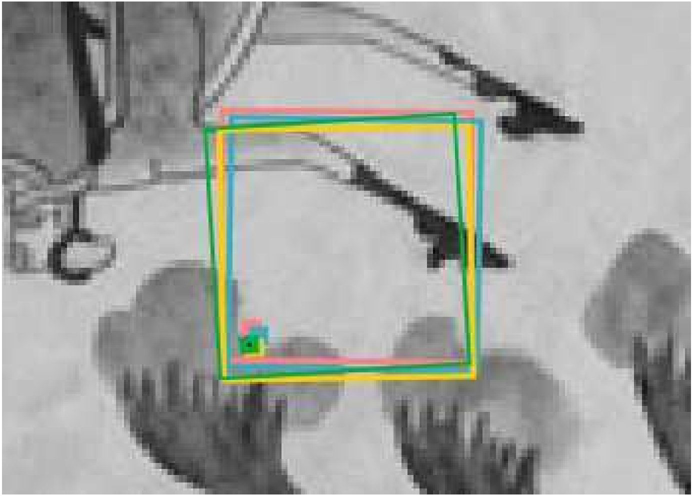
Tracking Filter Update from One Event



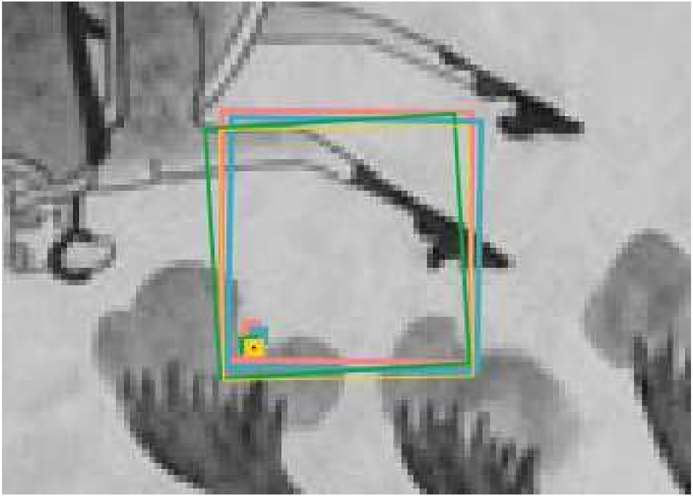
Tracking Filter Update from One Event



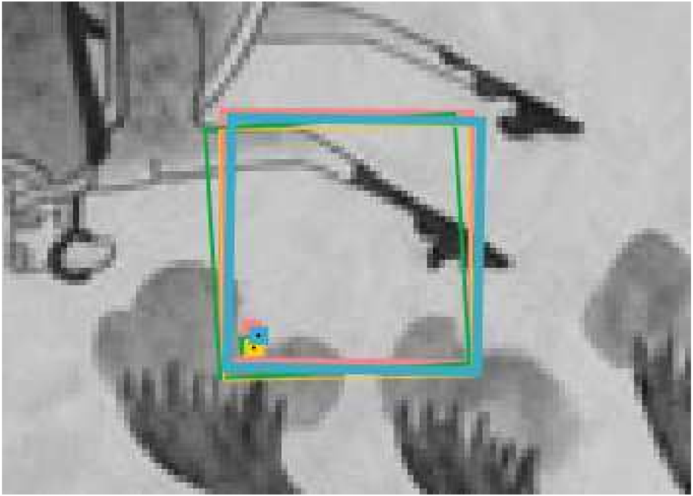
Tracking Filter Update from One Event



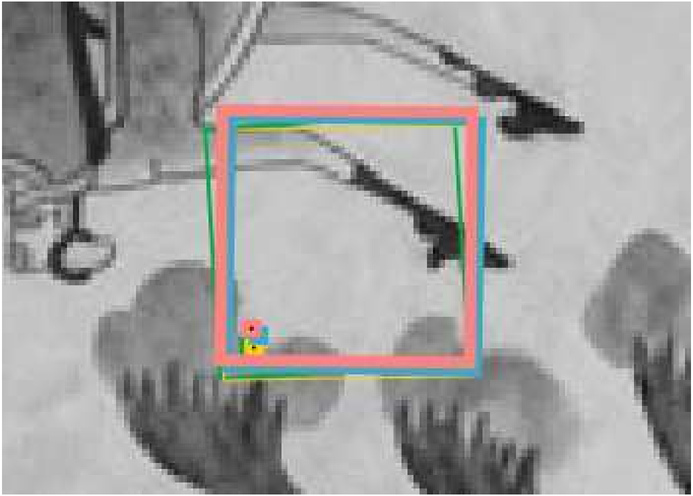
Tracking Filter Update from One Event



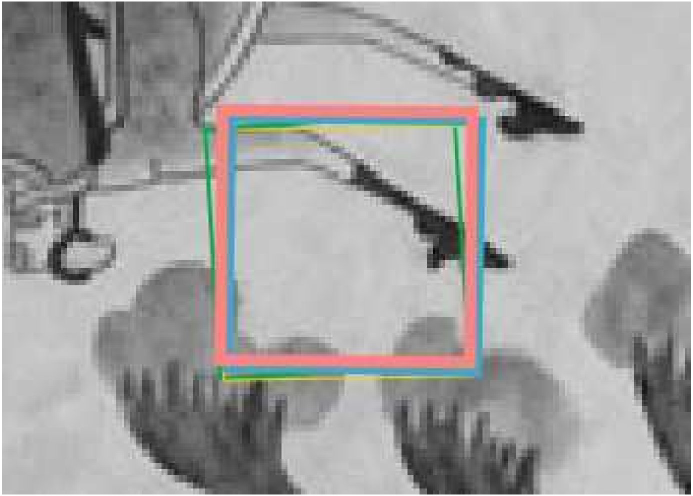
Tracking Filter Update from One Event



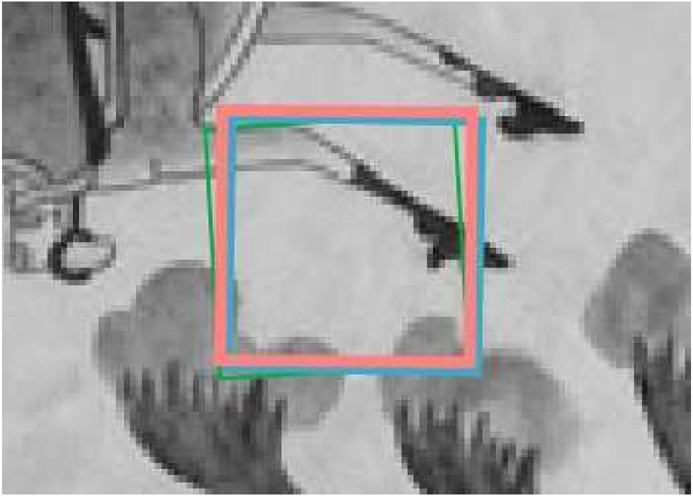
Tracking Filter Update from One Event



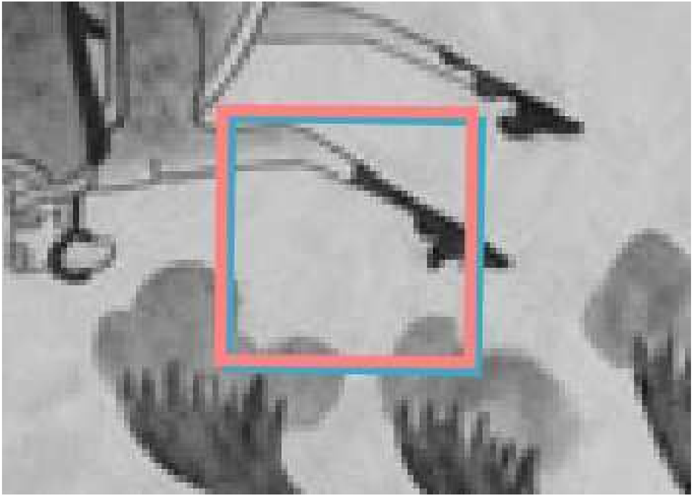
Tracking Filter Update from One Event



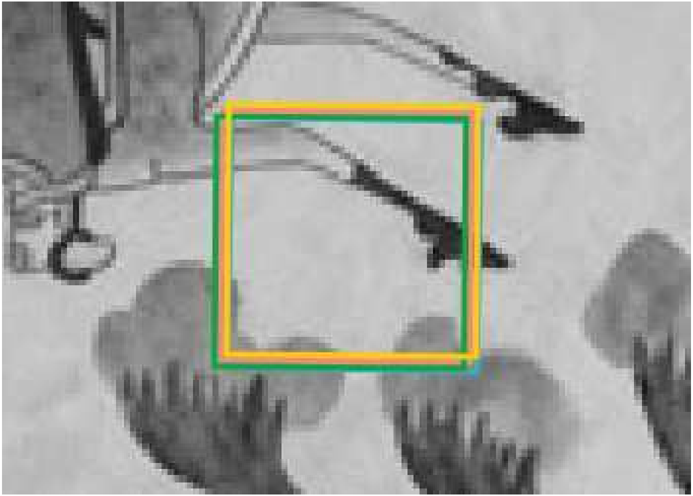
Tracking Filter Update from One Event



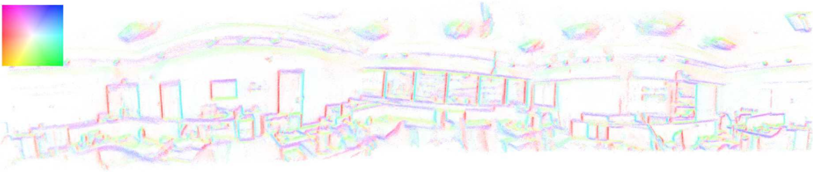
Tracking Filter Update from One Event



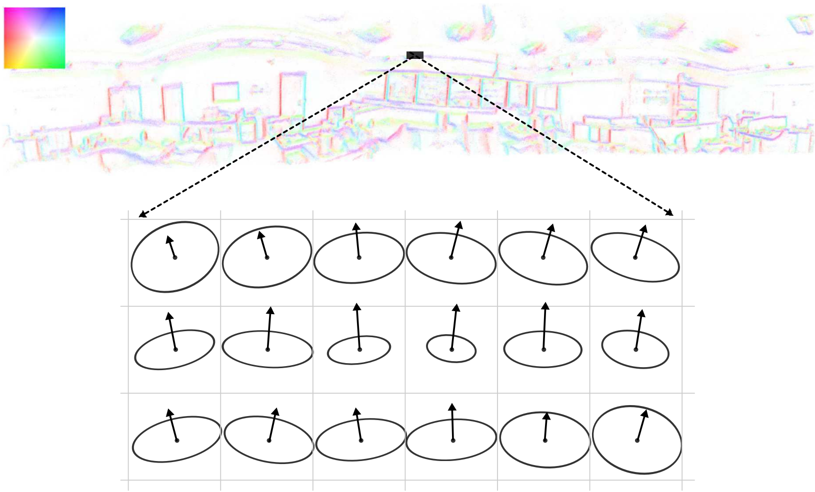
Tracking Filter Update from One Event



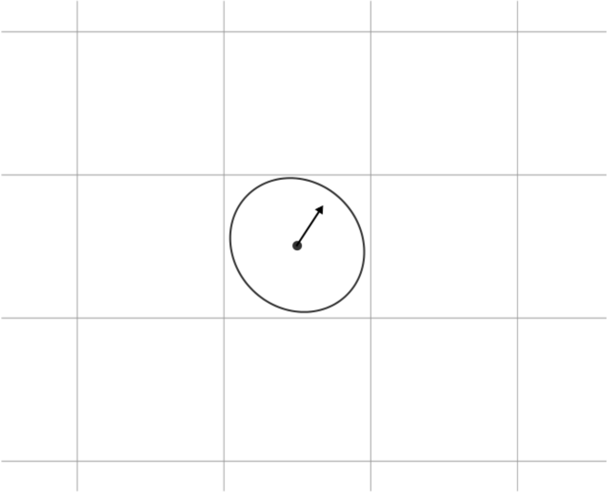
Gradient Estimation



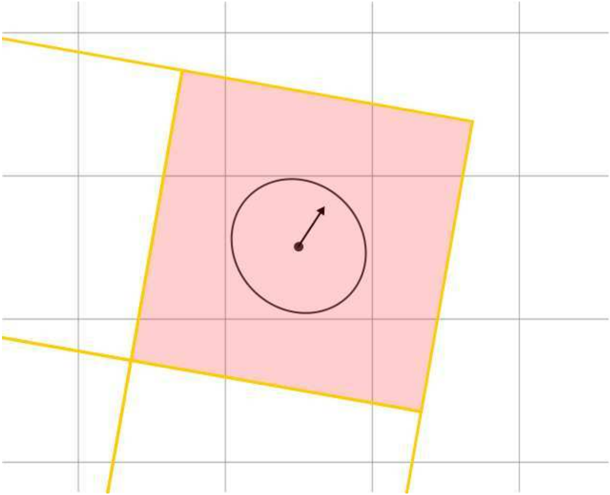
Pixel-wise EKF Gradient Estimation



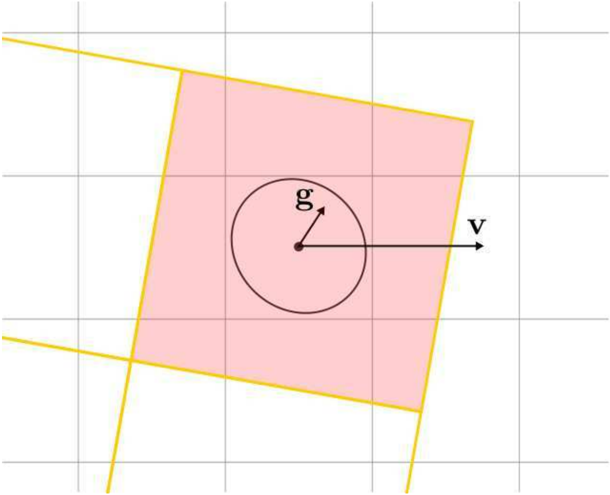
Pixel-wise EKF Gradient Estimation



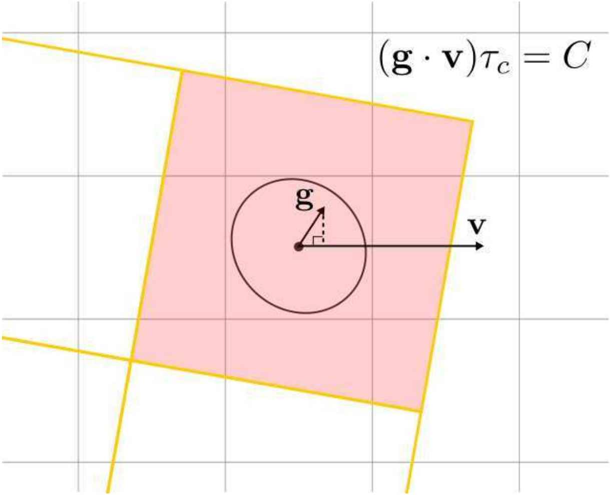
Pixel-wise EKF Gradient Estimation



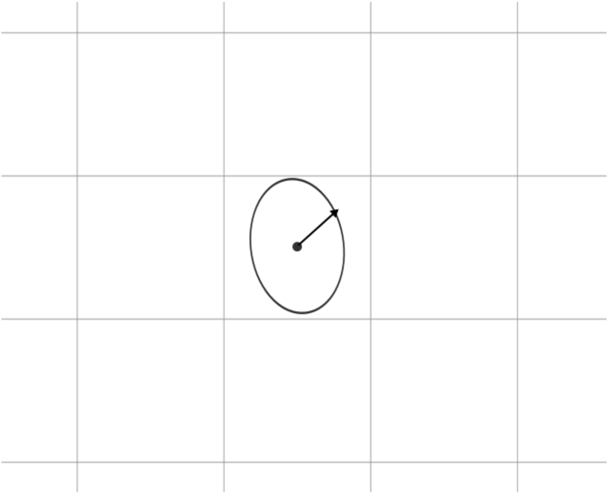
Pixel-wise EKF Gradient Estimation



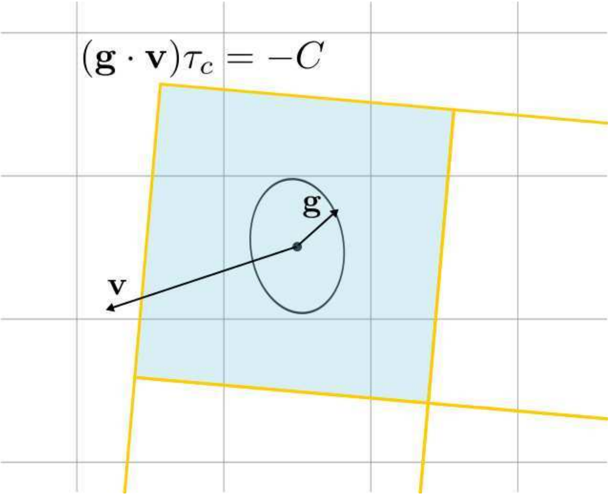
Pixel-wise EKF Gradient Estimation



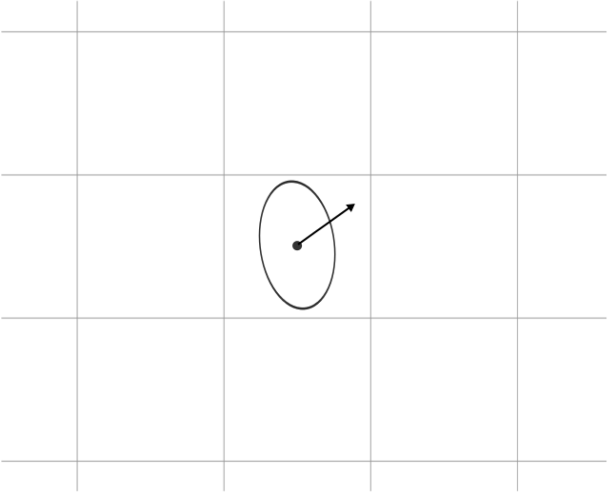
Pixel-wise EKF Gradient Estimation



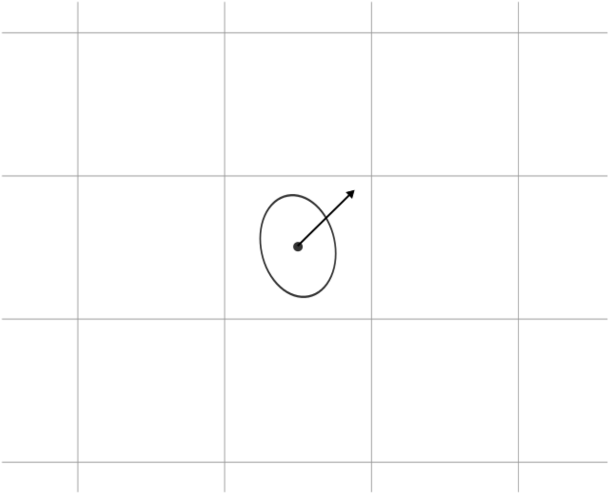
Pixel-wise EKF Gradient Estimation



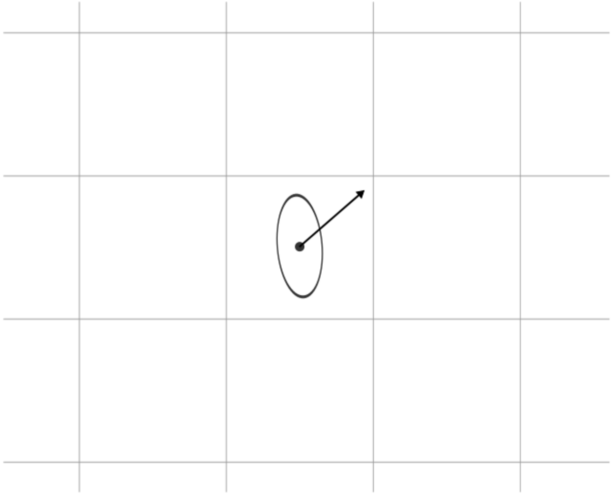
Pixel-wise EKF Gradient Estimation



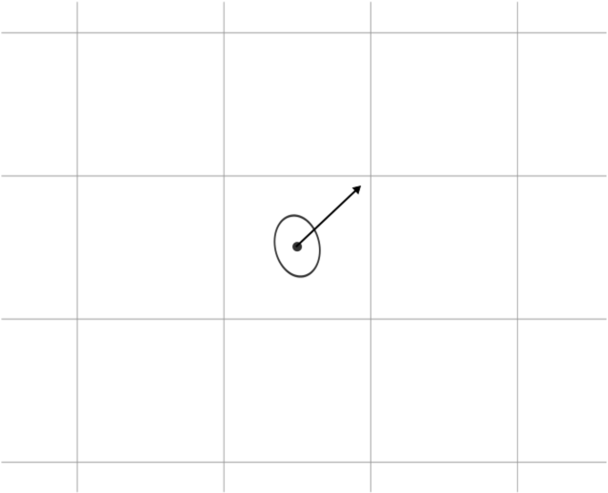
Pixel-wise EKF Gradient Estimation



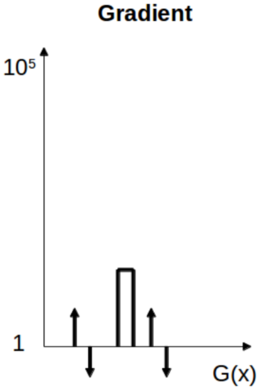
Pixel-wise EKF Gradient Estimation



Pixel-wise EKF Gradient Estimation

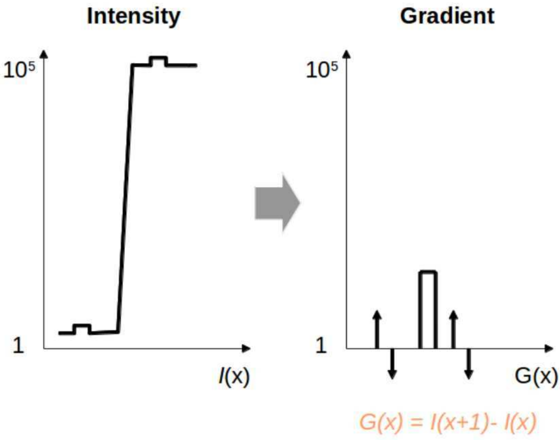


Reconstruction from Gradients in 1D



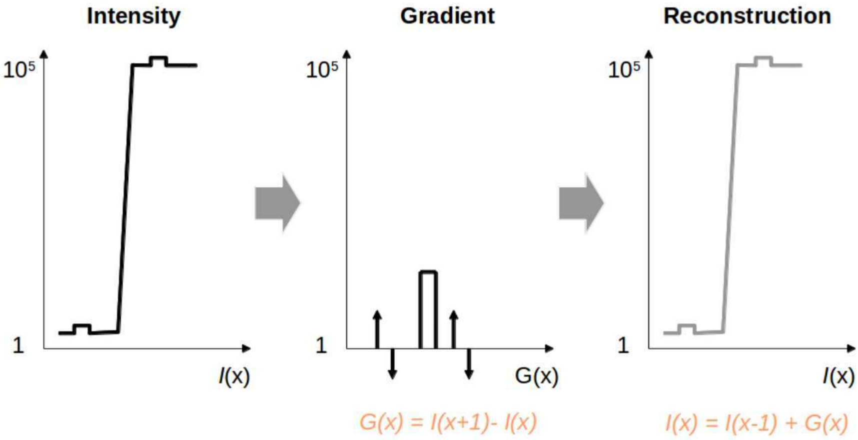
Agrawal, A. and Rasker, R., 2007

Reconstruction from Gradients in 1D



Agrawal, A. and Rasker, R., 2007

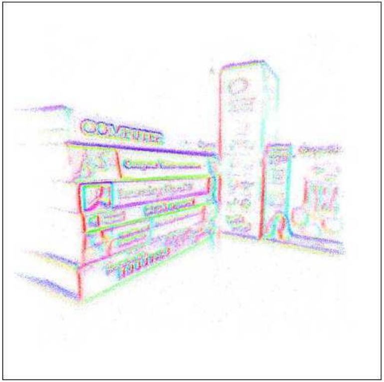
Reconstruction from Gradients in 1D



Agrawal, A. and Rasker, R., 2007

Intensity Reconstruction from Gradients in 2D

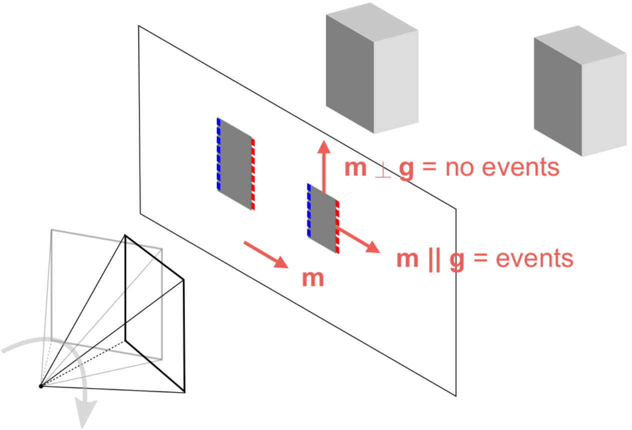
$$\min_{\mathbf{I}_l} \left\{ \int_{\Omega} ||\mathbf{g}(\mathbf{p}_k) - \nabla \mathbf{I}_l(\mathbf{p}_k)||^h_{\epsilon_d} + \lambda ||\nabla \mathbf{I}_l(\mathbf{p}_k)||^h_{\epsilon_r} \mathrm{d}\mathbf{p}_k \right\}$$



gradient estimation

intensity reconstruction

Events Induced by Pure Rotation Motion

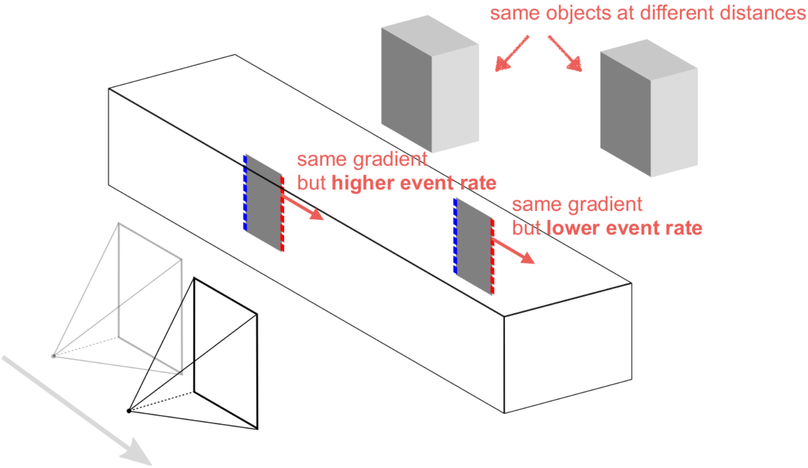


localisation
pure rotation motion

mapping
panoramic mosaic

3D scene
unknown & unstructured

Events Induced by Translation Motion

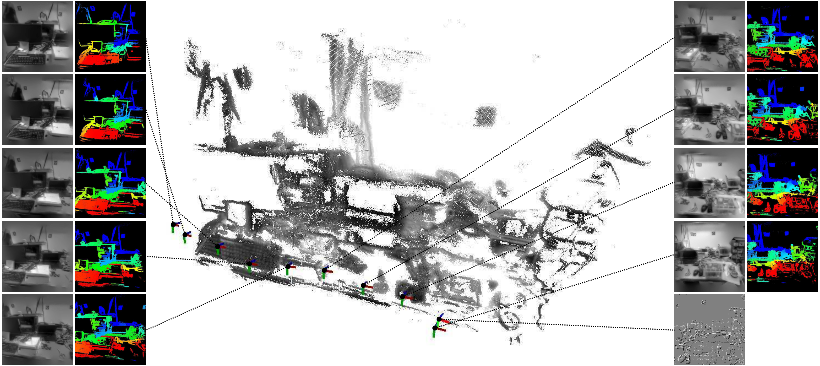


localisation
6-DoF hand-held motion

mapping
3D

3D scene
unknown & unstructured

3D Motion, Structure and Intensity from Event Data



- Change pose representation to 6DoF; add depth map estimation.
- Kim, Leutenegger, Davison, ECCV 2016 (Best Paper).

Conclusions

Future Directions

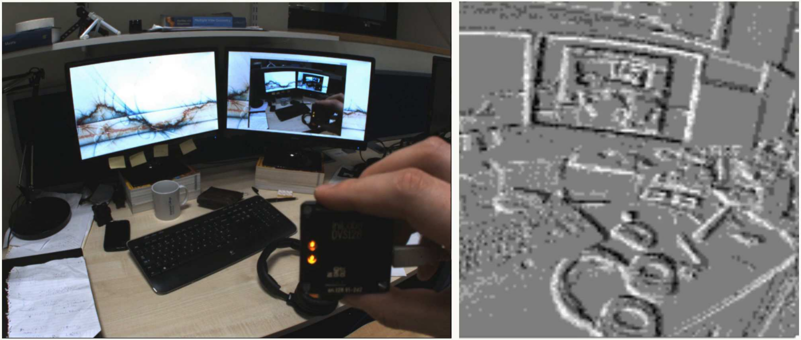
- SLAM will continue to evolve into general real-time spatial perception for embedded AI and IA.
- Co-design of sensors, processors and algorithms is the path to the performance and efficiency we need.
- Event-based reconstruction, SLAM and motion estimation is a fascinating and important research direction towards these goals.

Simultaneous Optical Flow and Intensity Estimation from an Event Camera

Patrick Bardow, Andrew Davison and Stefan Leutenegger
27/06/2016

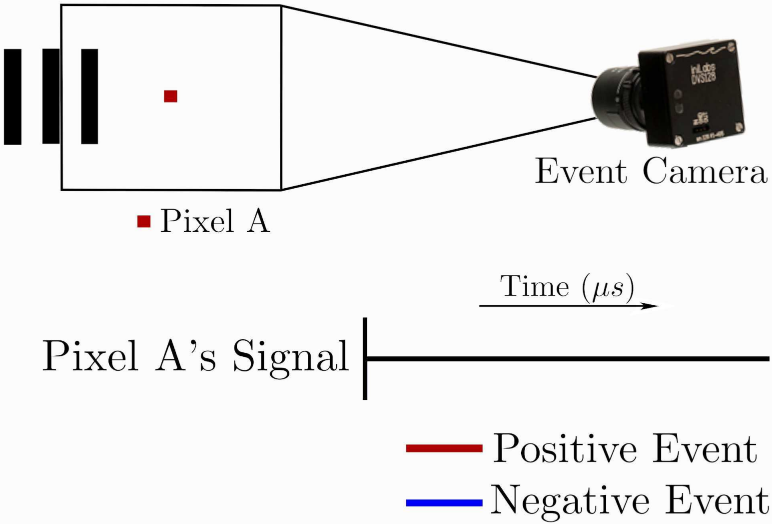
Dyson Robotics Laboratory, Imperial College London

Event Camera



Camera Scene Event Camera

Event Camera Scheme

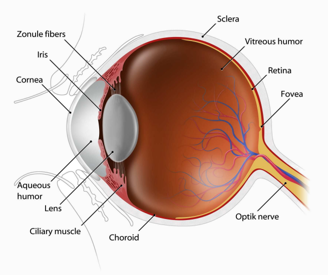


Event Cameras are

- bio-inspired
- captures only intensity changes
- asynchronous pixel (no frames)

1

DVS and EDVS from iniLabs



2

[1][Untitled photograph of DVS128 and EDVS]. (n.d.). Retrieved June 01, 2016, from http://inilabs.com

[2][Untitled anatomic drawing of an eye]. (n.d.). Retrieved May 10, 2016, from http://www.savesightcentre.com

Events cameras do not suffer from:

 [3]

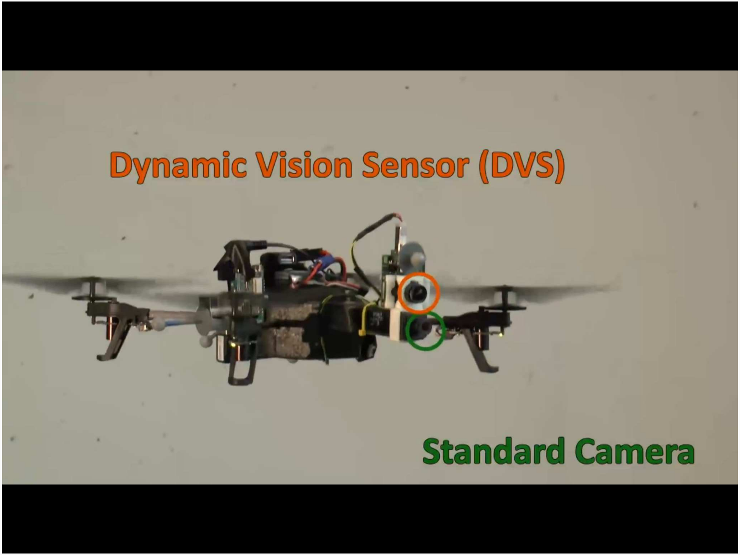
Overexposure

[3] Dangers of Driving Into Sun. (n.d.). Retrieved May 10, 2016, from http://echange.aaa.com

Events cameras do not suffer from:

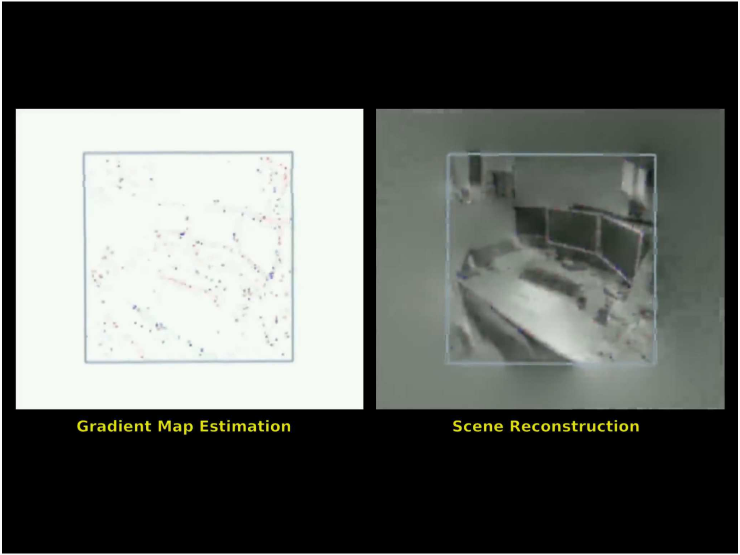
Motion Blur

Event-based, 6-DOF Pose Tracking for High-Speed Maneuvers[3]



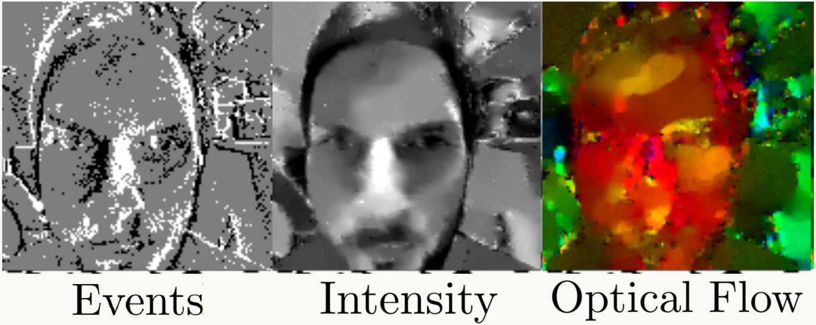
[3] Elias Mueggler, Basil Huber, and Davide Scaramuzza. "Event-based, 6-DOF Pose Tracking for High-Speed Maneuvers". In: *Intelligent Robots and Systems (IROS), 2014 IEEE/RSJ International Conference on*. 2014.

Simultaneous Mosaicing and Tracking with an Event Camera[4]



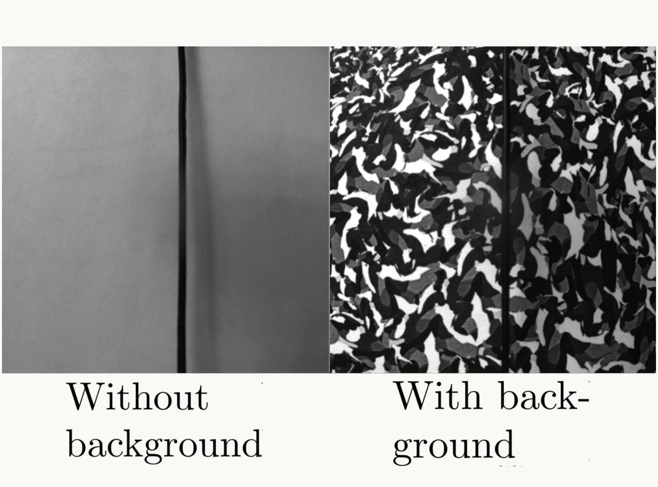
[4]Hanme Kim et al. "Simultaneous Mosaicing and Tracking with an Event Camera". In: *Proceedings of the British Machine Vision Conference*. BMVA Press, 2014.

Our Method

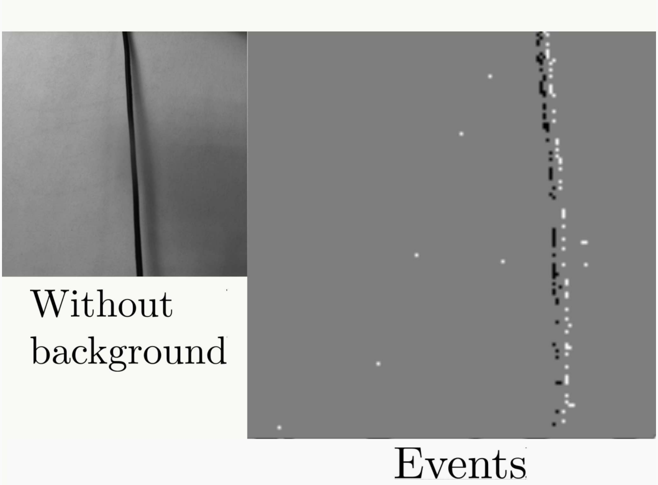


Proposed Method

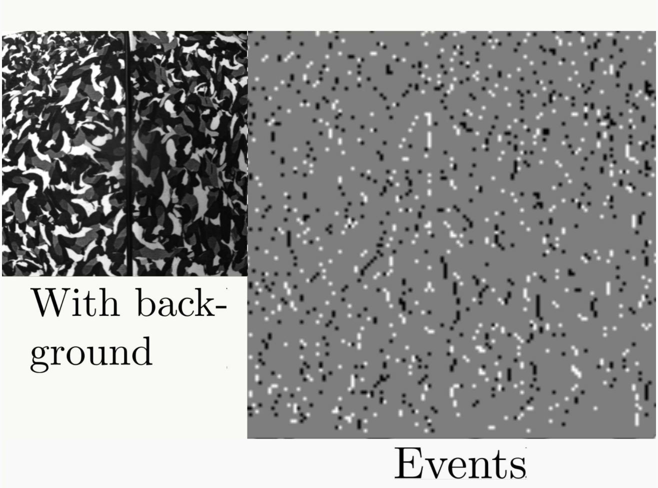
Detecting Motion with an Event Camera



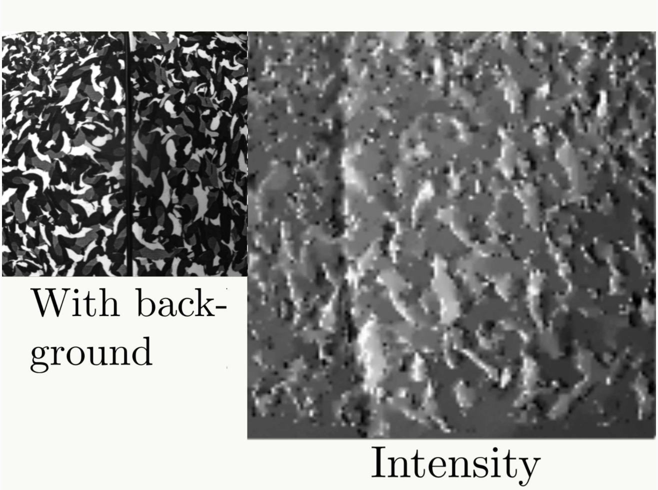
Detecting Motion with an Event Camera



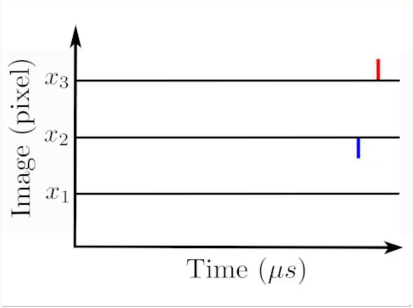
Detecting Motion with an Event Camera



Detecting Motion with an Event Camera

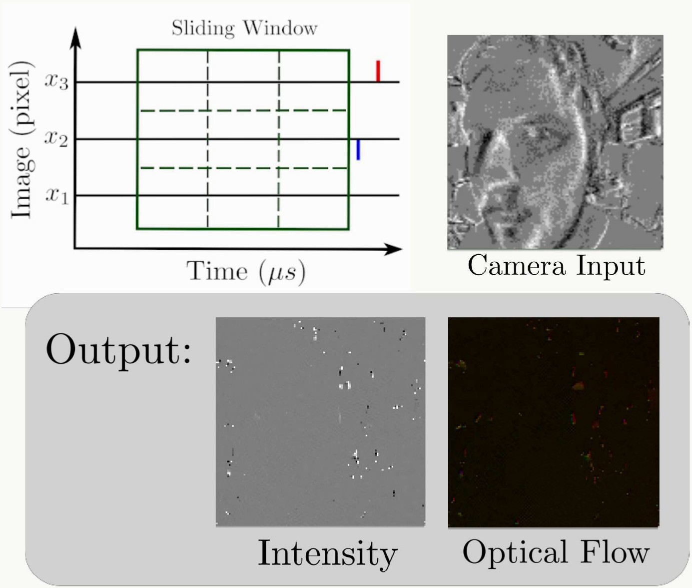


Sliding Window



Camera Input

Sliding Window



$$\min_{\mathbf{u},L} \int_\Omega \int_T \Big(\boxed{\lambda_1 \|\mathbf{u}_\mathbf{x}\|_1 + \lambda_2 \|\mathbf{u}_\mathbf{t}\|_1 + \lambda_3 \|L_\mathbf{x}\|_1} + \boxed{\lambda_4 \|\langle L_\mathbf{x}, \delta_t \mathbf{u}\rangle + L_t\|_1} + \boxed{\lambda_5\, h_\theta(L - L(t_p))} \Big)\, \mathrm{d}t\, \mathrm{d}\mathbf{x} + \int_\Omega \sum_{i=2}^{|P(\mathbf{x})|} \boxed{\|L(t_i) - L(t_{i-1}) - \theta \rho_i\|_1} \mathrm{d}\mathbf{x},$$

- Smoothness terms
- Optical flow term
- No-event term
- Event term

Cost Function

$$\min_{\mathbf{u},L} \int_\Omega \int_T \Big(\boxed{\lambda_1 \|\mathbf{u_x}\|_1 + \lambda_2 \|\mathbf{u_t}\|_1 + \lambda_3 \|L_\mathbf{x}\|_1} + \lambda_4 \| \langle L_\mathbf{x}, \delta_t \mathbf{u} \rangle + L_t \|_1 + \lambda_5 \, h_\theta(L - L(t_p)) \Big) \mathrm{d}t \, \mathrm{d}\mathbf{x} + \int_\Omega \sum_{i=2}^{|P(\mathbf{x})|} \|L(t_i) - L(t_{i-1}) - \theta \rho_i \|_1 \mathrm{d}\mathbf{x},$$

- Smoothness terms

Cost Function

$$\min_{\mathbf{u},L} \int_\Omega \int_T \Big(\boxed{\lambda_1 \|\mathbf{u}_\mathbf{x}\|_1 + \lambda_2 \|\mathbf{u}_\mathbf{t}\|_1 + \lambda_3 \|L_\mathbf{x}\|_1} + \boxed{\lambda_4 \|\langle L_\mathbf{x}, \delta_t \mathbf{u}\rangle + L_t\|_1} + \lambda_5\, h_\theta(L - L(t_p)) \Big) \mathrm{d}t\, \mathrm{d}\mathbf{x} + \int_\Omega \sum_{i=2}^{|P(\mathbf{x})|} \|L(t_i) - L(t_{i-1}) - \theta\rho_i\|_1 \mathrm{d}\mathbf{x},$$

- Smoothness terms
- Optical flow term

Cost Function

$$\min_{\mathbf{u},L} \int_\Omega \int_T \Big(\boxed{\lambda_1 \|\mathbf{u_x}\|_1 + \lambda_2 \|\mathbf{u_t}\|_1 + \lambda_3 \|L_\mathbf{x}\|_1} + \boxed{\lambda_4 \|\langle L_\mathbf{x}, \delta_t \mathbf{u} \rangle + L_t\|_1} + \boxed{\lambda_5 \, h_\theta(L - L(t_p))} \Big) \mathrm{d}t \, \mathrm{d}\mathbf{x}$$
$$+ \int_\Omega \sum_{i=2}^{|P(\mathbf{x})|} \|L(t_i) - L(t_{i-1}) - \theta \rho_i\|_1 \mathrm{d}\mathbf{x},$$

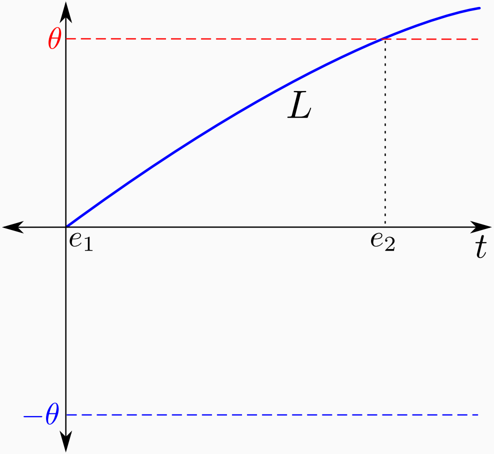
- Smoothness terms
- Optical flow term
- No-event term

Cost Function

$$
\begin{aligned}
\min_{\mathbf{u}, L} \int_\Omega \int_T \Big(& \lambda_1 \|\mathbf{u_x}\|_1 + \lambda_2 \|\mathbf{u_t}\|_1 + \lambda_3 \|L_\mathbf{x}\|_1 + \\
& \lambda_4 \| \langle L_\mathbf{x}, \delta_t \mathbf{u} \rangle + L_t \|_1 + \lambda_5 \, h_\theta(L - L(t_p)) \Big) \, \mathrm{d}t \, \mathrm{d}\mathbf{x} \\
& + \int_\Omega \sum_{i=2}^{|P(\mathbf{x})|} \| L(t_i) - L(t_{i-1}) - \theta \rho_i \|_1 \mathrm{d}\mathbf{x},
\end{aligned}
$$

- Smoothness terms
- Optical flow term
- No-event term
- Event term

Event term and no-event term

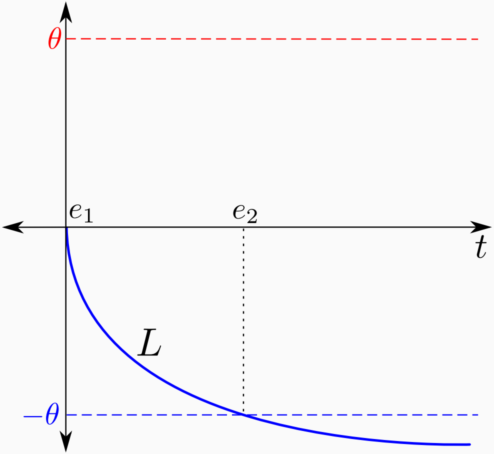


Event fire condition:

$$|L(t') - L(e_1)| \geq \theta,$$

$$L := \log(I + b)$$

Event term and no-event term

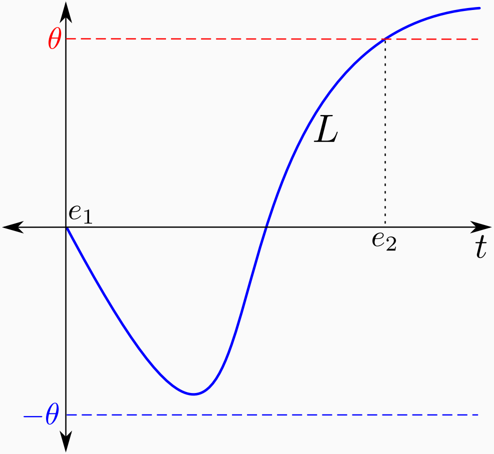


Event fire condition:

$$|L(t') - L(e_1)| \geq \theta,$$

$$L := \log(I + b)$$

Event term and no-event term

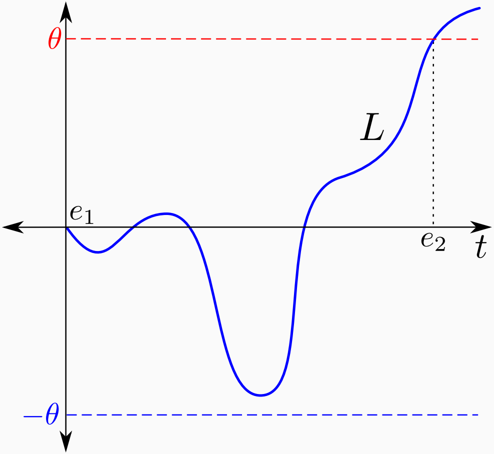


Event fire condition:

$$|L(t') - L(e_1)| \geq \theta,$$

$$L := \log(I + b)$$

Event term and no-event term

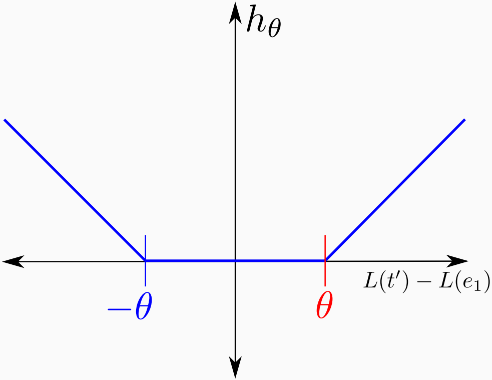


Event fire condition:

$$|L(t') - L(e_1)| \geq \theta,$$

$$L := \log(I + b)$$

Event term and no-event term

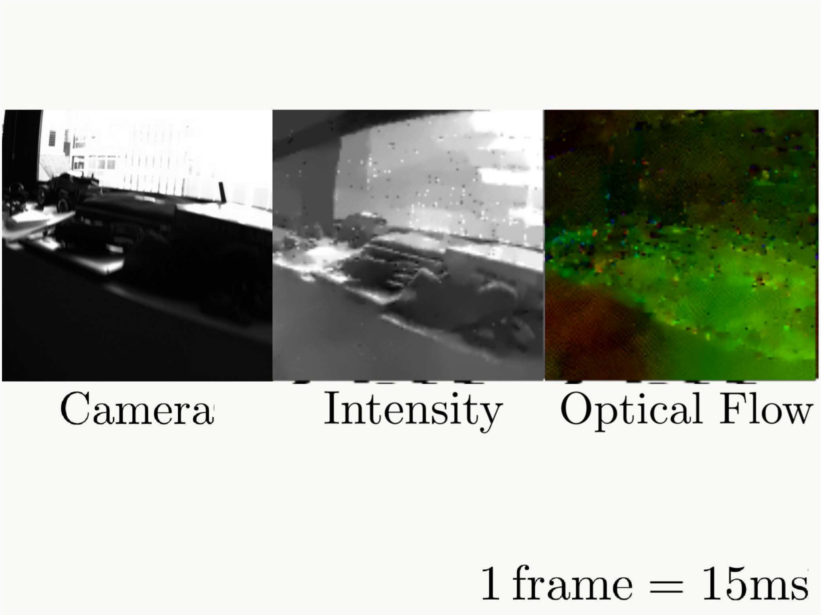


No-event term:

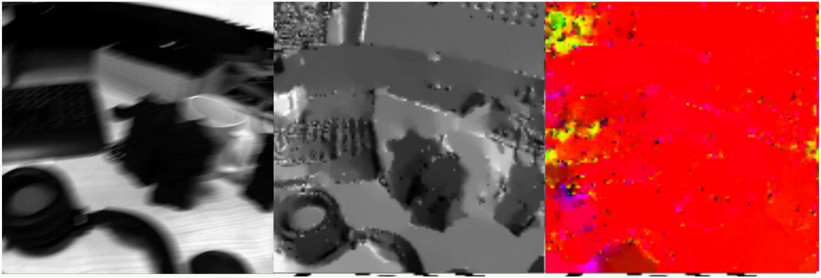
$$h_\theta(L(t') - L(e_1)),$$
$$L := \log(I + b)$$

Results

High Dynamic Range Sequence



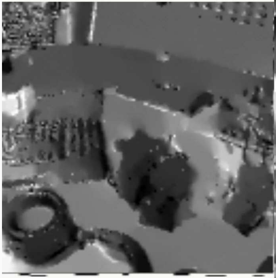
Rapid Motion Sequence



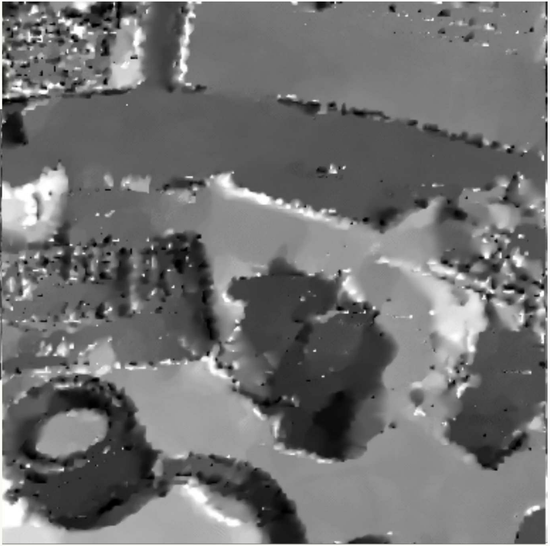
Camera Intensity Optical Flow

1 frame = 5ms

Super-Resolution



128×128



384×384

Conclusion

- New bio-inspired camera paradigm which overcomes issues of motion blur and overexposure.
- Sends only essential data by discarding the concept of frames
- Potential benifits to large areas of robotic vision

Conclusion

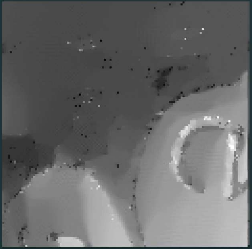
- New bio-inspired camera paradigm which overcomes issues of motion blur and overexposure.
- Sends only essential data by discarding the concept of frames
- Potential benifits to large areas of robotic vision

- First method for dense optical and intensity reconstruction with any image motion from an event camera
- A first proof-of-concept that the event data contains enough image information
- Bridges the gap between event data and traditional computer vision algorithms

Thank You

Reconstructing spinning fan blades

Standard Camera



500fps Reconstruction

References

[1] Hanme Kim et al. "Simultaneous Mosaicing and Tracking with an Event Camera". In: *Proceedings of the British Machine Vision Conference*. BMVA Press, 2014.

[2] Elias Mueggler, Basil Huber, and Davide Scaramuzza. "Event-based, 6-DOF Pose Tracking for High-Speed Maneuvers". In: *Intelligent Robots and Systems (IROS), 2014 IEEE/RSJ International Conference on*. 2014.