Imperial College London

Qualcomm Augmented Reality Lecture Series:

Visual SLAM with an Event-based Camera

Hanme Kim

Supervisor: Prof. Andrew Davison

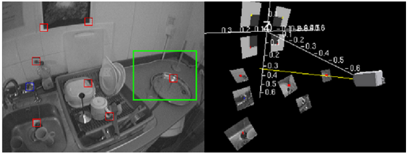
Robot Vision Group
Department of Computing
Imperial College London

January 27, 2015

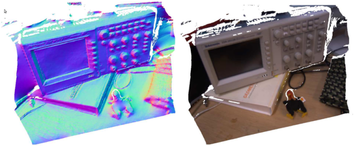
Imperial College London

Robot Vision Group

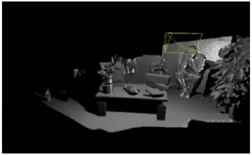
- Principle investigator: Prof. Andrew Davison.



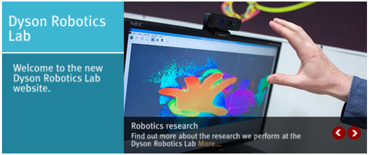
MonoSLAM - Davison, ICCV 2003



DTAM - Newcombe et al., ICCV 2011



KinectFusion - Newcombe et al., ISMAR 2011



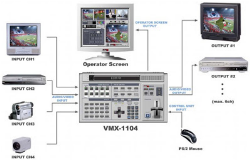
Dyson Robotics Lab

- Closely linked to the Dyson Robotics Laboratory.

Self Introduction

- 2nd year PhD student with industrial experience (about 12 years).

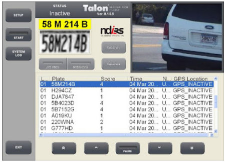
CCTV



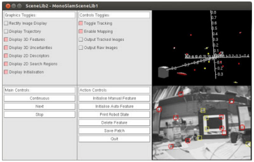
A/V Mixer

ISP

Sensor Guided Robot



ANPR

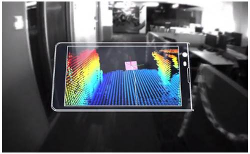


SceneLib2

Visual SLAM Applications

- We are interested in visual SLAM applications.

Dyson 360 Eye



Google Tango

Amazon PrimeAir

- They require:
  - fast control feedback;
  - high dynamic range;
  - low power consumption and hardware complexity.

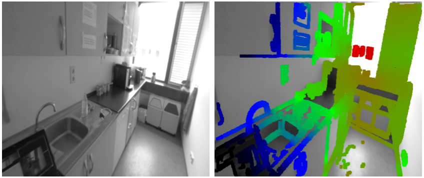
Limited by Conventional Imaging Sensors

- Only work well with controlled camera motion and scene condition.
- High power consumption (e.g. Hands-free Google Glass with a battery pack on the hand!).
- High hardware complexity (e.g. powerful GPU requirement).

www.techradar.com

Sophisticated State-of-the-Art Algorithms

- Sophisticated algorithms which mainly reduce their computational burden by selecting and processing only informative data.



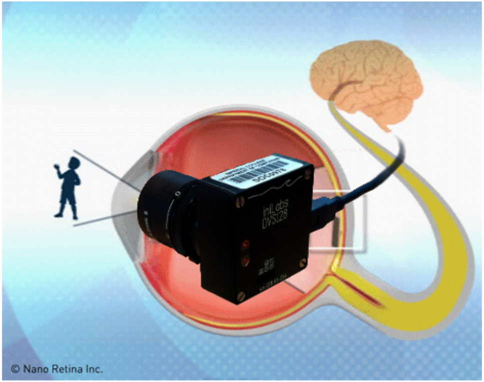
Semi-Dense VO, Engel, J. et al., ICCV 2013

Semi-Direct VO, Forster, C. et al., ICRA 2014

- They still rely on conventional imaging sensors, therefore they still suffer from some of the limitations (e.g. blind between frames, low dynamic range, high power consumption, etc.).

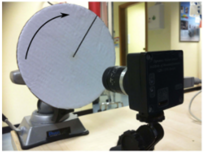
Motivation

- How to satisfy these requirements of real-time SLAM applications?
- Can **bio-inspired silicon retinas** from Neuromorphics be a solution?

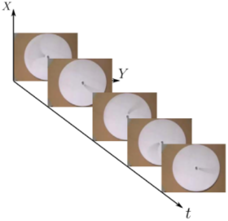


DVS128, iniLabs & Background from Nano Retina Inc.

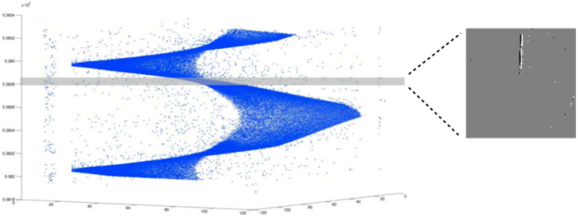
Event-based Camera



Frame-based Camera

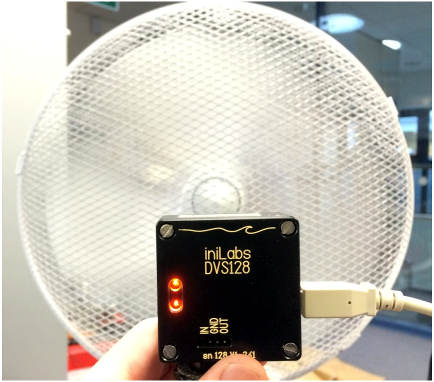


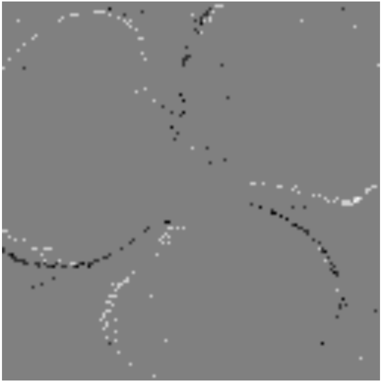
Event-based Camera



Benosman, R. et al., 2014

DVS (Dynamic Vision Sensor) Live Demo

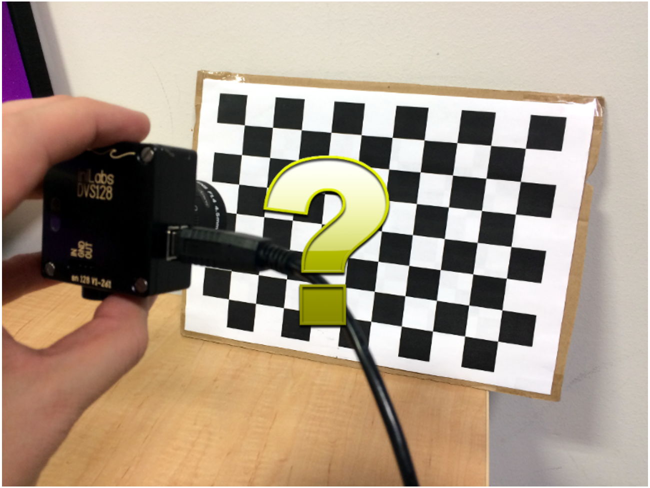




Event-based Camera

- Advantages:
  - Asynchronous and fast visual measurements, low latency.
  - High dynamic range.
  - Compressed visual information requires lower transmission bandwidth, storage capacity, processing time, and power consumption.
- Has potential to overcome the limitations of conventional imaging sensors.
- Requires totally new computer vision algorithms.

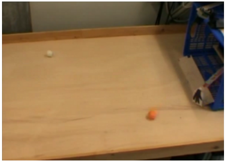
e.g. How To Calibrate It?



Related Work:

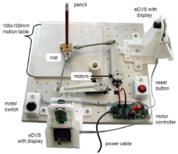
Tracking and Recognition

- Tracking algorithms showing its low latency capability.



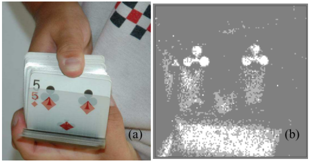
Delbruck, T. and Lichtsteiner, P., 2007

DVS Laser Tracker, 2008

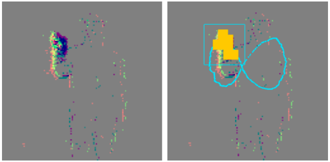


Conradt, J. et al., 2009

- Combined with biologically inspired learning approaches.



Pérez-Carrasco, J. A. et al., PAMI 2013

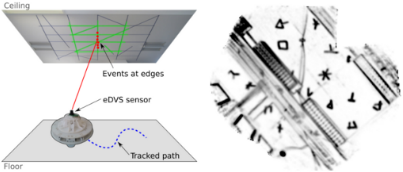


Lee, J. et al., ISCAS 2012

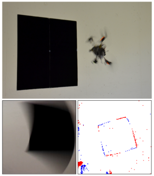
Related Work:

SLAM Applications

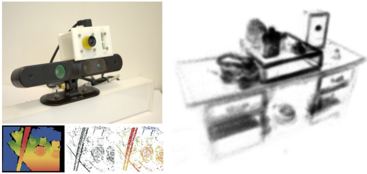
- SLAM with limitations or extra sensors.



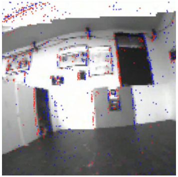
Weikersdorfer, D. et al., ICVS 2013



Mueggler, E. et al., IROS 2014



Weikersdorfer, D. et al., ICRA 2014



Censi, A. and Scaramuzza, D., ICRA 2014

What We Want To Achieve

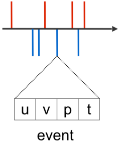
- 3D Visual SLAM with a single event-based camera:
    - able to track extremely fast 6 DoF camera motion and reconstruct 3D scenes;
    - requires low computational cost, hardware complexity and power consumption;
    - suitable for real world applications.

ETAM (Event-based Tracking and Mapping), conceptual drawing

Simultaneous Mosaicing and Tracking with an Event Camera

- Hanme Kim, Ankur Handa, Ryad Benosman, Sio-Hoi Ieng, Andrew J. Davison.
- Published at BMVC (British Machine Vision Conference) 2014.
- Oral presentation (7.7% acceptance rate).
- Best industry paper.

Proposed Algorithm



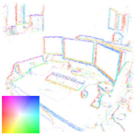
Event-Based Tracking
(Particle Filter)

SO(3) Camera Pose Estimate

HR & HDR Scene Mosaic

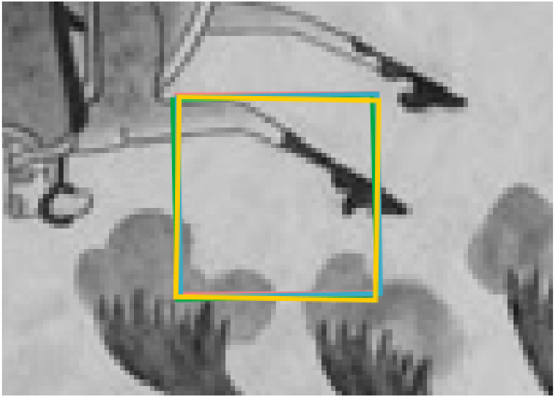
Mosaicing (Mapping)

Gradient Estimation
(Pixel-wise EKF)



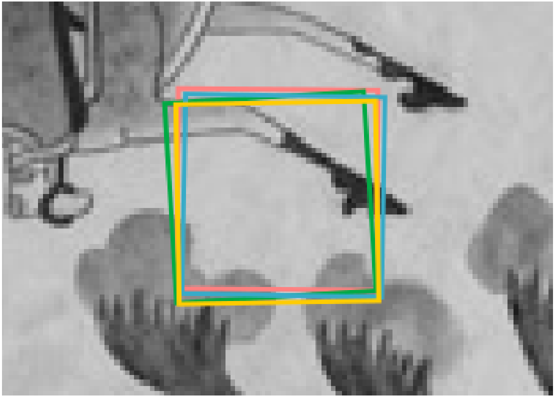
Reconstruction
(Poisson Solver)

Event-based Tracking



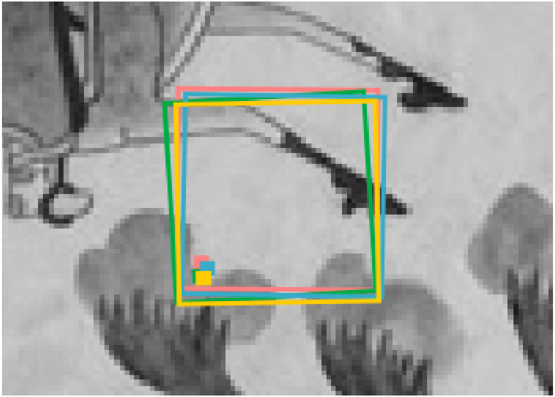
$$\{\mathbf{p}_1^{(t)}, \mathbf{p}_2^{(t)}, \mathbf{p}_3^{(t)}, \mathbf{p}_4^{(t)}\}, \mathbf{p}_i^{(t)} = \{\mathrm{R}_i^{(t)} \in \mathbf{SO}(3), w_i^{(t)}\}$$

Event-based Tracking



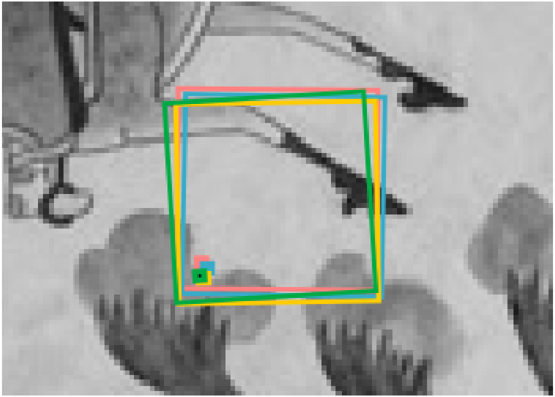
$$\mathtt{R}_i^{(t)} = \mathtt{R}_i^{(t-\tau)} \exp\left(\sum_{k=1}^{3} \mathbf{n}_k \mathtt{G}_k\right), n_i \sim \mathcal{N}(0, \sigma_i^2)$$

Event-based Tracking



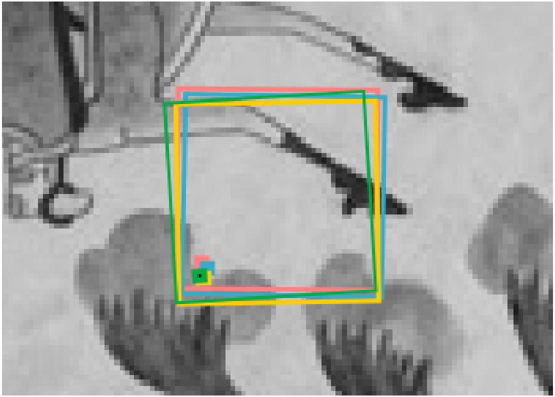
Imperial College London

Event-based Tracking



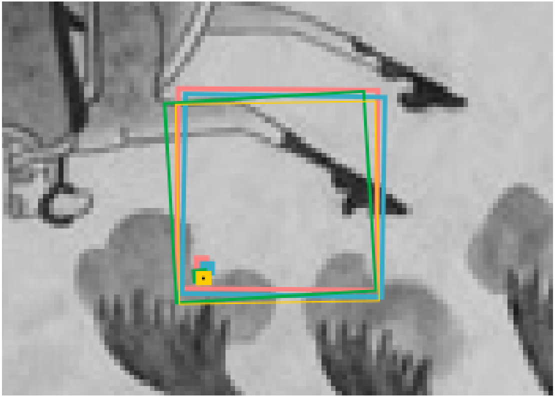
$$w_1^{(t)} = P(z|\mathtt{R}_1^{(t)})w_1^{(t-\tau)}, z = \log(\mathtt{M}(\mathbf{p}_m^{(t)})) - \log(\mathtt{M}(\mathbf{p}_m^{(t-\tau_c)}))$$

Event-based Tracking



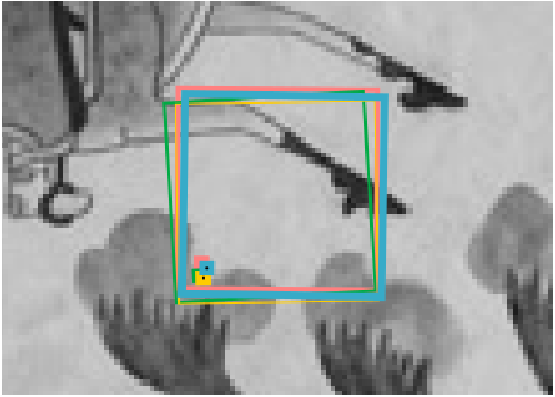
$$w_1^{(t)} = P(z|\mathtt{R}_1^{(t)})w_1^{(t-\tau)}, z = \log(\mathtt{M}(\mathbf{p}_m^{(t)})) - \log(\mathtt{M}(\mathbf{p}_m^{(t-\tau_c)}))$$

Event-based Tracking



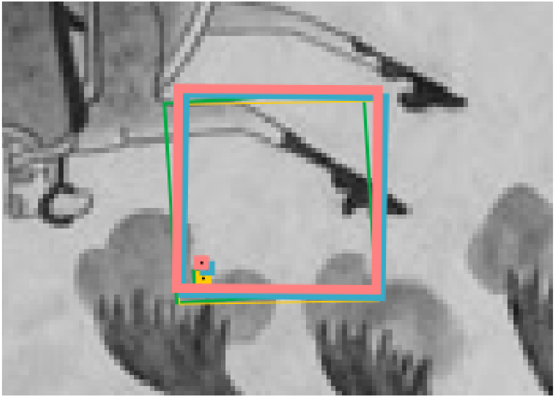
$$w_2^{(t)} = P(z|\mathtt{R}_2^{(t)})w_2^{(t-\tau)}, z = \log(\mathtt{M}(\mathbf{p}_m^{(t)})) - \log(\mathtt{M}(\mathbf{p}_m^{(t-\tau_c)}))$$

Event-based Tracking



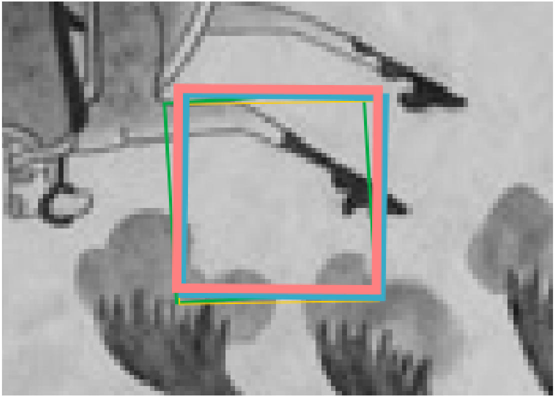
$$w_3^{(t)} = P(z|\mathrm{R}_3^{(t)})w_3^{(t-\tau)}, z = \log(\mathtt{M}(\mathbf{p}_m^{(t)})) - \log(\mathtt{M}(\mathbf{p}_m^{(t-\tau_c)}))$$

Event-based Tracking

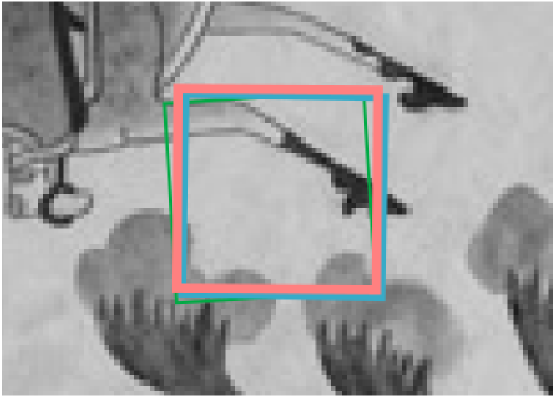


$$w_4^{(t)} = P(z|\mathtt{R}_4^{(t)})w_4^{(t-\tau)}, z = \log(\mathtt{M}(\mathbf{p}_m^{(t)})) - \log(\mathtt{M}(\mathbf{p}_m^{(t-\tau_c)}))$$

Event-based Tracking

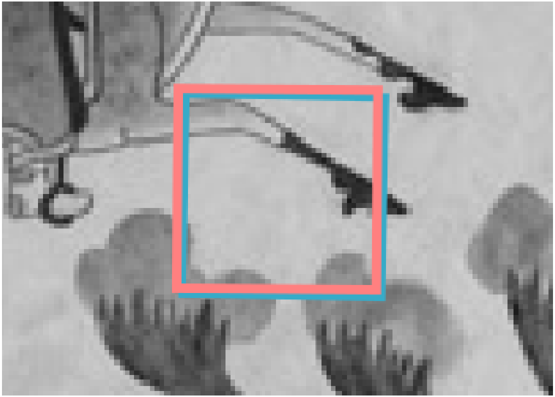


Event-based Tracking

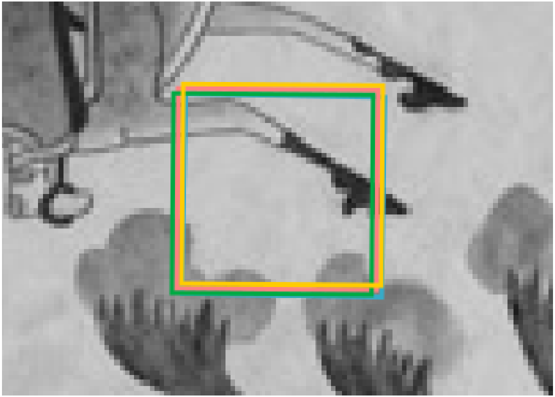


Imperial College London

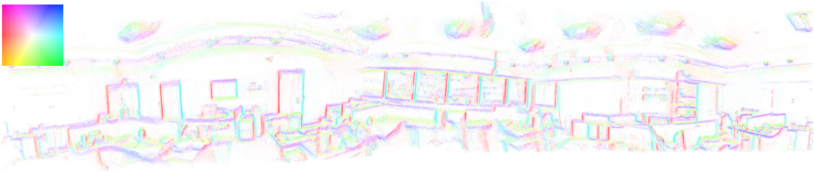
Event-based Tracking



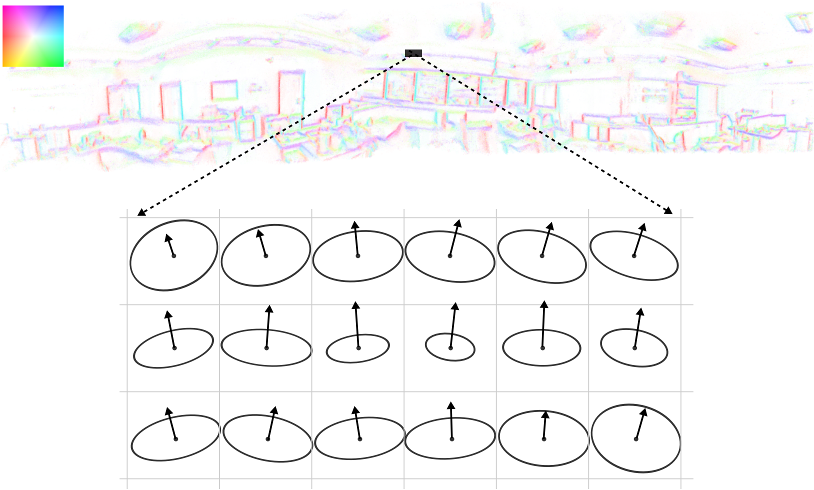
Event-based Tracking



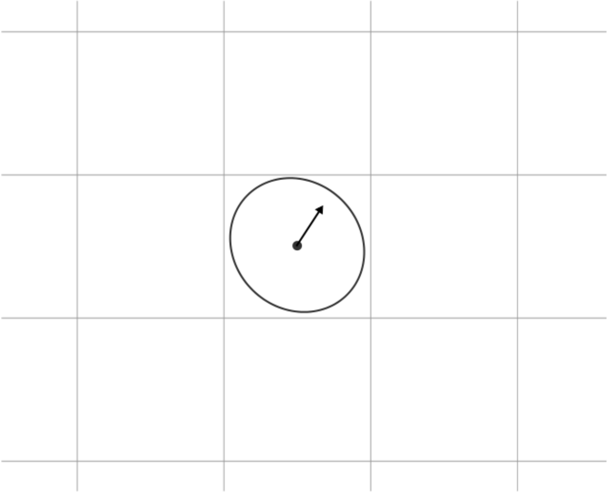
Gradient Estimation



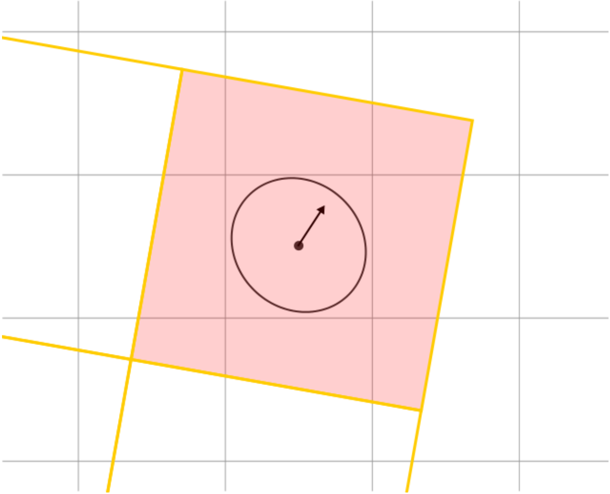
Pixel-wise EKF Gradient Estimation



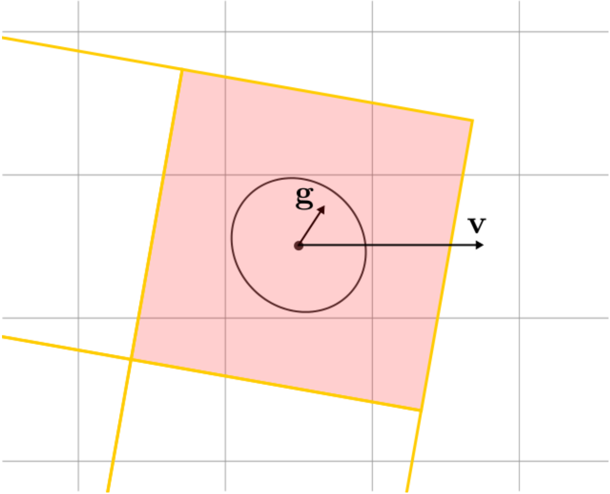
Pixel-wise EKF Gradient Estimation



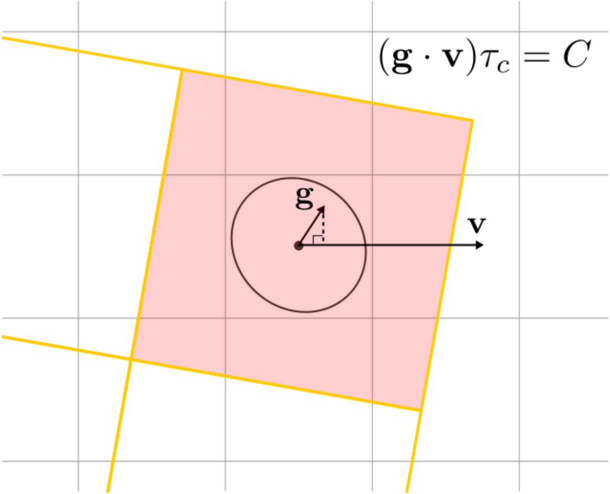
Pixel-wise EKF Gradient Estimation



Pixel-wise EKF Gradient Estimation



Pixel-wise EKF Gradient Estimation



Pixel-wise EKF Gradient Estimation

- Measurement z and measurement model h.

$$z^{(t)} = \frac{1}{\tau_c}, \quad h^{(t)} = \frac{\mathbf{g}^{(t)} \cdot \mathbf{v}^{(t)}}{C}$$

- Update the gradient vector and its covariance matrix using standard EKF equation.

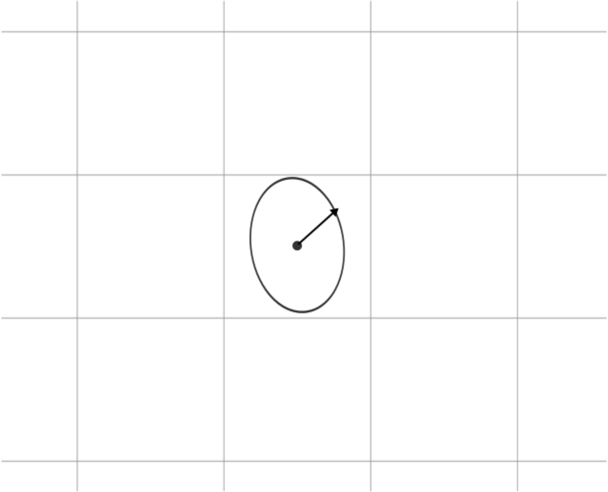
$$\mathbf{g}^{(t)} = \mathbf{g}^{(t-\tau_c)} + \mathbf{W}\nu, \quad \mathrm{P}_{\mathbf{g}}^{(t)} = \mathrm{P}_{\mathbf{g}}^{(t-\tau_c)} - \mathbf{W}\mathrm{S}\mathbf{W}^\top$$

$$\nu = z^{(t)} - h^{(t)}$$

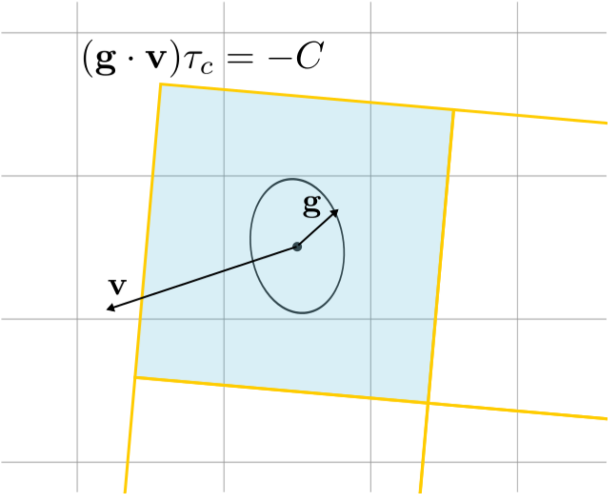
$$\mathbf{W} = \mathrm{P}_{\mathbf{g}}^{(t-\tau_c)} \frac{\partial h}{\partial \mathbf{g}}^\top \mathrm{S}^{-1}$$

$$\mathrm{S} = \frac{\partial h}{\partial \mathbf{g}} \mathrm{P}_{\mathbf{g}}^{(t-\tau_c)} \frac{\partial h}{\partial \mathbf{g}}^\top + \mathrm{R}$$

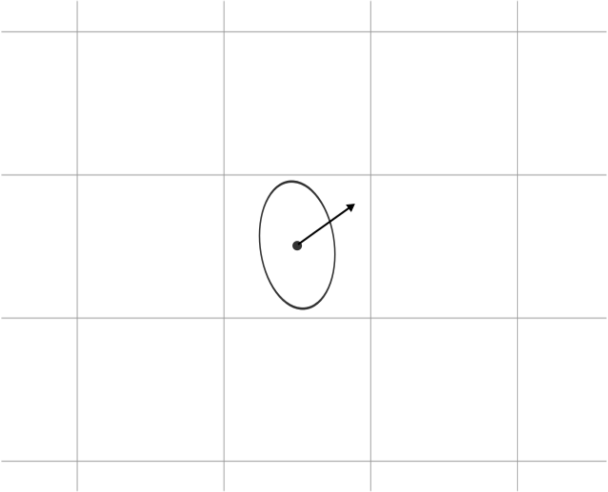
Pixel-wise EKF Gradient Estimation



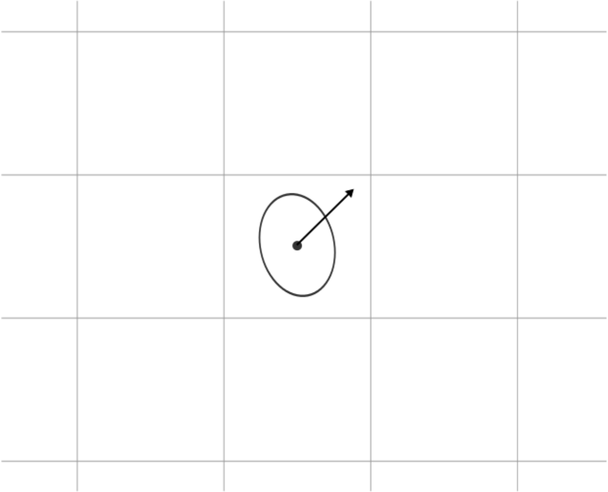
Pixel-wise EKF Gradient Estimation



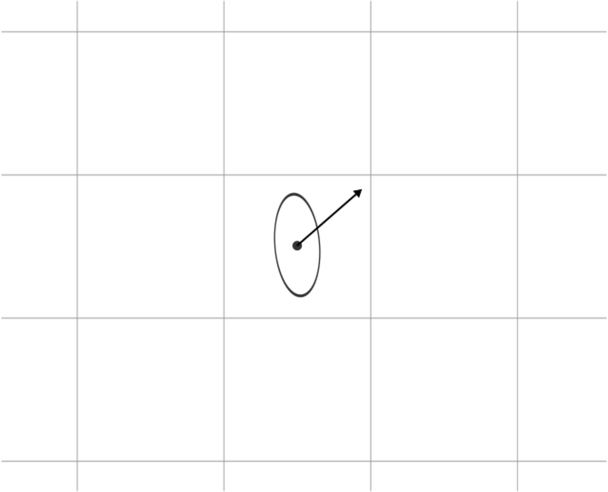
Pixel-wise EKF Gradient Estimation



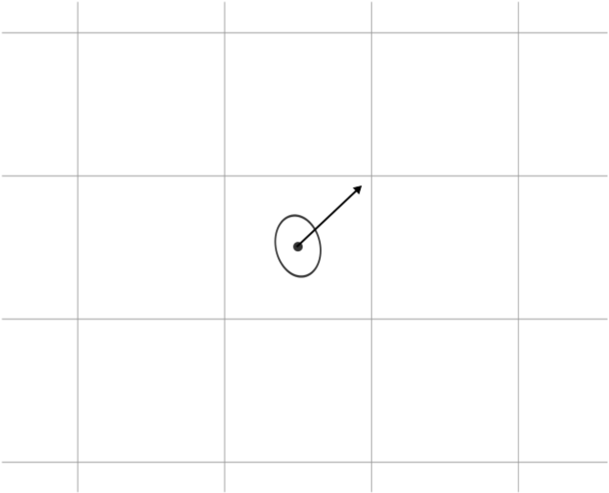
Pixel-wise EKF Gradient Estimation



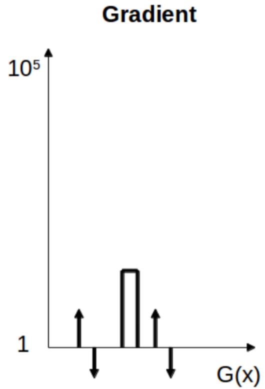
Pixel-wise EKF Gradient Estimation



Pixel-wise EKF Gradient Estimation

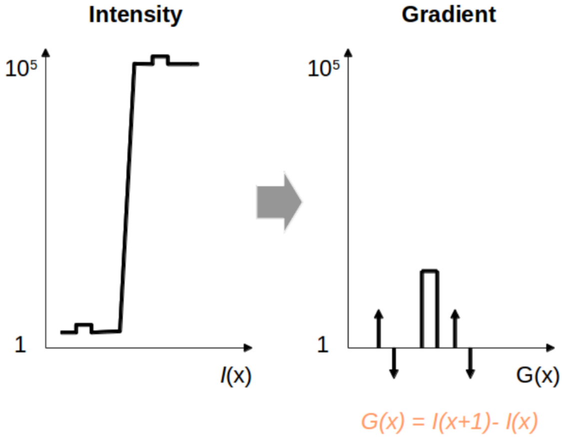


Reconstruction from Gradients in 1D



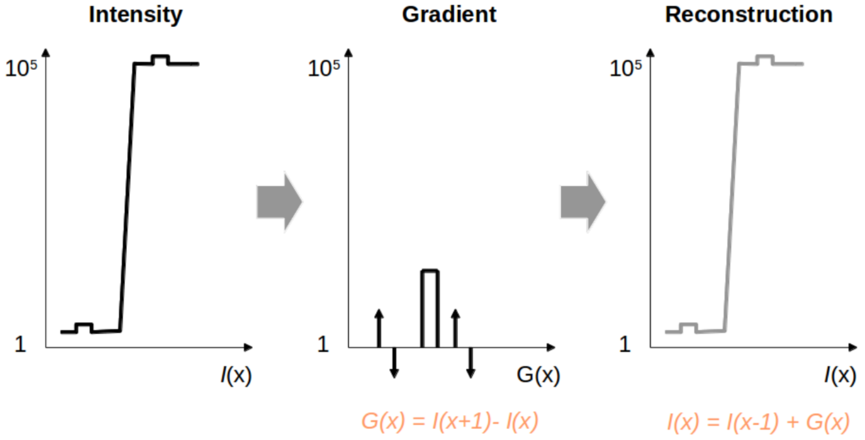
Agrawal, A. and Rasker, R., 2007

Reconstruction from Gradients in 1D



Agrawal, A. and Rasker, R., 2007

Reconstruction from Gradients in 1D



Agrawal, A. and Rasker, R., 2007

Reconstruction from Gradients in 2D

- Reconstruct the log intensity of the image whose gradients $\mathtt{M}_x$ and $\mathtt{M}_y$ across the whole image domain are close to the estimated gradient g_x and g_y in a least squares sense (Tumblin, J. *et al.*, 2005):

  $J(\mathtt{M}) = \int\int (M_x - g_x)^2 + (M_y - g_y)^2 dxdy.$

  The Euler-Lagrange equation to minimise $J(\mathtt{M})$ is:

  $\frac{\partial J}{\partial \mathtt{M}} - \frac{\mathrm{d}}{\mathrm{d}x}\frac{\partial J}{\partial M_x} - \frac{\mathrm{d}}{\mathrm{d}y}\frac{\partial J}{\partial M_y} = 0$

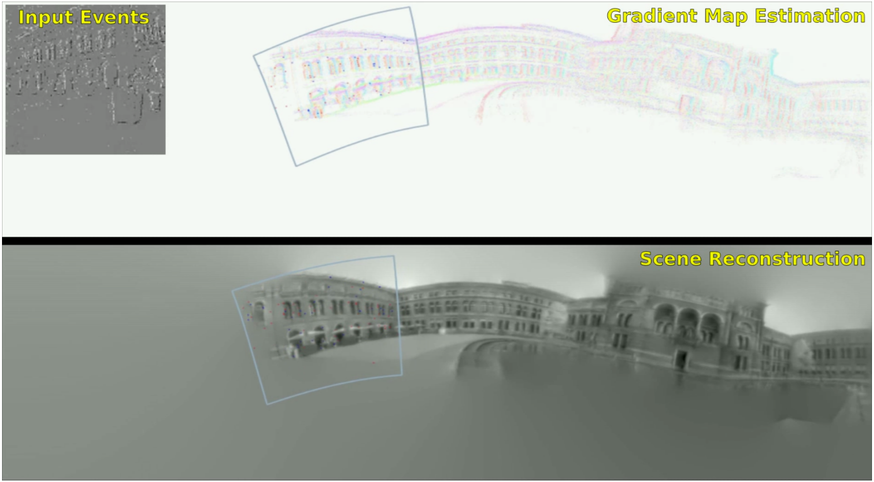
  which leads to the well known Poisson equation:

  $\boldsymbol{\nabla}^2\mathtt{M} = \frac{\partial}{\partial x}g_x + \frac{\partial}{\partial y}g_y.$

- We use a sine transform based method to solve the Poisson equation (Agrawal, A. *et al.*, 2005 and 2006).

Imperial College London

Demo Video



Contributions and Limitations

- **Contributions**
  - Track camera rotation while building a mosaic of a scene purely based on an event stream with no additional sensing.
  - Reconstruct high resolution and dynamic range scenes by harnessing the characteristics of event cameras.
  - Show that all visual information is in the event stream by reconstructing scenes.
- **Limitations**
  - Processing time depends on the # of particles and map resolution.
  - No proper bootstrapping.

Current Extensions

- Real-time operation.
  - Passion equation $\rightarrow$ parallelizable primal-dual equation.
  - Particle filter based tracking $\rightarrow$ EKF based tracking.
  - Track against an estimated gradient map directly.
- Performance improvement.
  - Stronger motion model (e.g. constant velocity or acceleration).
  - Bootstrapping.
  - Continuous-time representation.
- Hope to publish and make it as open source soon.

Move Towards 3D SLAM

- 3D simulator development.
  - Generate synthetic 3D scenes (POV-Ray), trajectories and event datasets.
  - Plan to make it as open source.
- Depth estimation.
  - Events are also related to translating motion and depth.

ETAM (Event-based Tracking and Mapping), conceptual drawing

Things Worth Restating

- Useful pixels are determined in hardware, at no computational cost.
- Event camera vs conventional camera.

	Event Camera	Conventional Camera
Data Rate	40-180kB/s	10MB/s
Latency	few μs	few ms
Dynamic Range	about 120dB	about 60dB
Power Consumption	few hundreds mW	few W

- Much more innovation is happening in Neuromorphics.