Real-Time Intensity-Image Reconstruction for Event Cameras

Thomas Pock

AIT Austrian Institute of Technology GmbH
Digital Safety & Security Department

Scientific vision days, November 9th, 2016

Joint work with:
Christian Reinbacher and Gottfried Graber

Event Cameras vs. Conventional Cameras

Event cameras are a paradigm shift in digital camera technology

- Low Latency
- Low Bandwidth
- High Dynamic Range

Event Cameras vs. Conventional Cameras

Event cameras are a paradigm shift in digital camera technology

- Low Latency
- Low Bandwidth
- High Dynamic Range

- High Resolution
- Images

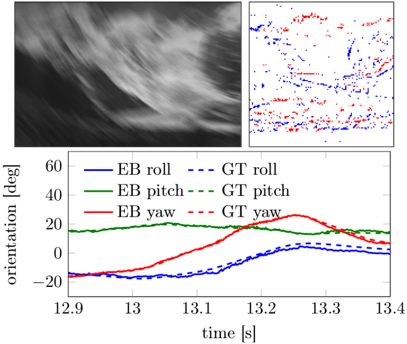
Motivation

Motivation

⟶

Related Work - Computer Vision with Event Cameras

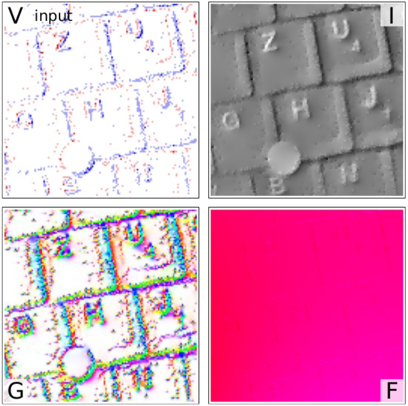
- Camera Tracking
- Optical Flow
- Image Reconstruction



Event-based, 6-DOF Camera Tracking for High-Speed Applications [Gallego et al. '16]

Related Work - Computer Vision with Event Cameras

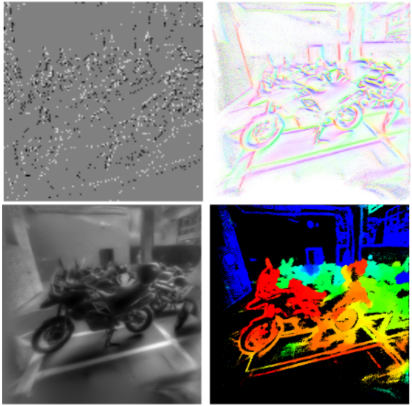
- Camera Tracking
- Optical Flow
- Image Reconstruction



Interacting Maps for Fast Visual Interpretation [Cook et al. '11]

Related Work - Computer Vision with Event Cameras

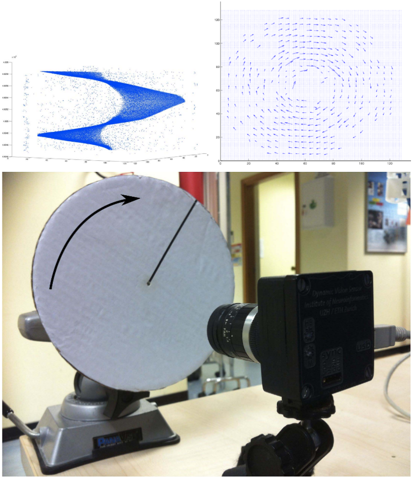
- Camera Tracking
- Optical Flow
- Image Reconstruction



Simultaneous Mosaicing and tracking [Kim et al. '14] and Real-Time 3D Reconstruction and 6-DoF Tracking with an Event Camera [Kim et al. '16]

Related Work - Computer Vision with Event Cameras

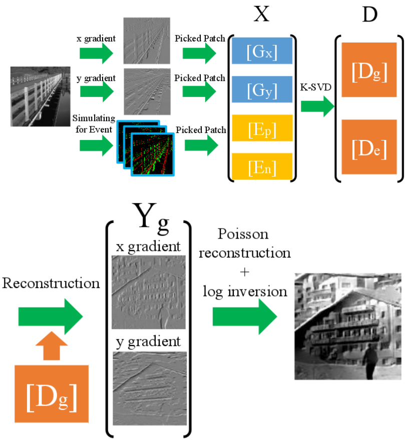
- Camera Tracking
- Optical Flow
- Image Reconstruction



Event-Based Visual Flow [Benosman et al. '14]

Related Work - Computer Vision with Event Cameras

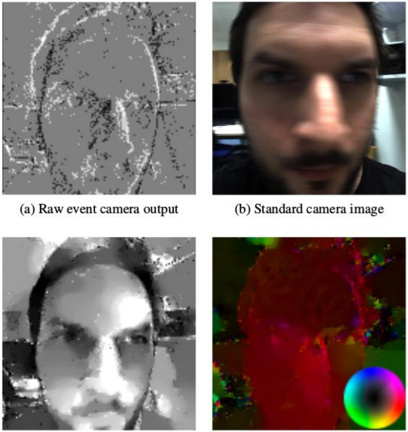
- Camera Tracking
- Optical Flow
- Image Reconstruction



Face Detection and Video Reconstruction from Event Cameras [Miyatani et al. '16]

Related Work - Computer Vision with Event Cameras

- Camera Tracking
- Optical Flow
- Image Reconstruction



(a) Raw event camera output (b) Standard camera image

(c) Intensity estimate from events (d) Optical flow from events

Simultaneous Optical Flow and Intensity Estimation from an Event Camera [Bardow et al. '16]

Image Formation Process

Event cameras report an event $e^n = \{x^n, y^n, \theta^n, t^n\}$ if pixel intensity at (x^n, y^n) has changed by a threshold $\Delta^{\pm}$ in log-space. For the underlying image in intensity space this means:

$$f^n(x^n, y^n) = u^{n-1}(x^n, y^n) \cdot \begin{cases} c_1 & \text{if } \theta^n > 0 \\ c_2 & \text{if } \theta^n < 0 \end{cases},$$

with $c_1 = \exp(\Delta^+)$, $c_2 = \exp(-\Delta^-)$.

Reconstruction by Denoising

Problems:

- u^0 is unknown and can not be recovered
- Noise in events (possibly not iid)

Goal of this work

Recover a denoised u^n from f^n (fast).

Reconstruction by Denoising

Problems:

- u^0 is unknown and can not be recovered
- Noise in events (possibly not iid)

Goal of this work

Recover a denoised u^n from f^n (fast).

Our Solution:

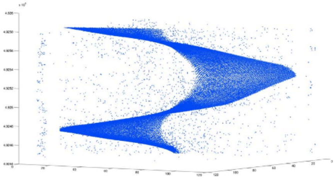
$$u^n = \operatorname*{argmin}_{u \in C^1(\Omega, \mathbb{R}_+)} \left[E(u) = D(u, f^n) + R(u) \right] ,$$

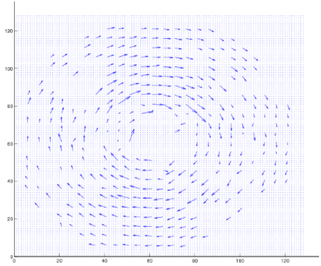
- where $D(u, f^n)$ models camera noise
- and $R(u)$ enforces regularity in the solution

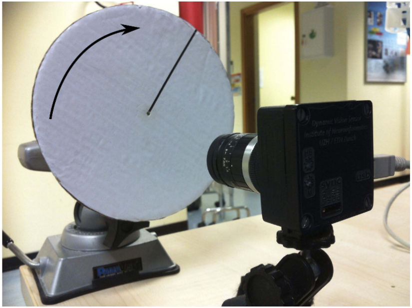
Remaining questions

How to choose D and R and what about the timestamps t^n?

Surface of Active Events

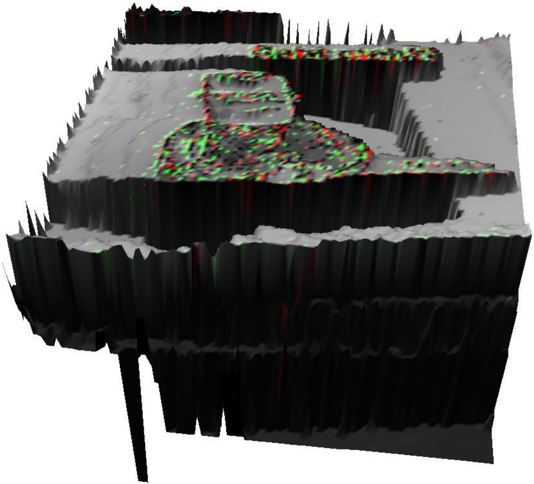






[Benosman et al. '14]

Surface of Active Events



surface $S \subset \mathbb{R}^3$ as the graph of a scalar function $t(x, y)$

$$X = \varphi(x, y) = \begin{bmatrix} x, & y, & t(x, y) \end{bmatrix}^T$$

Defining a Regulariser on the SAE

Given a smooth function $s \in C^1(S, \mathbb{R})$ on the manifold, we can define $\mathrm{d}s(Y) = \langle \nabla_g s, Y \rangle_g \quad \forall Y \in T_X\mathcal{M}$ [Lee et al. '97], with

$$\nabla_g s = \left(g^{11} s_x + g^{12} s_y\right) \varphi_x + \left(g^{21} s_x + g^{22} s_y\right) \varphi_y,$$

where g^{ij} denotes the components of the inverse of g, the metric tensor

$$g = \begin{bmatrix} \langle \varphi_x, \varphi_x \rangle & \langle \varphi_x, \varphi_y \rangle \\ \langle \varphi_x, \varphi_y \rangle & \langle \varphi_y, \varphi_y \rangle \end{bmatrix}.$$

Defining a Regulariser on the SAE

Given a smooth function $s \in C^1(S, \mathbb{R})$ on the manifold, we can define $\mathrm{d}s(Y) = \langle \nabla_g s, Y \rangle_g \quad \forall Y \in T_X\mathcal{M}$ [Lee et al. '97], with

$$\nabla_g s = \left(g^{11} s_x + g^{12} s_y\right) \varphi_x + \left(g^{21} s_x + g^{22} s_y\right) \varphi_y,$$

where g^{ij} denotes the components of the inverse of g, the metric tensor

$$g = \begin{bmatrix} \langle \varphi_x, \varphi_x \rangle & \langle \varphi_x, \varphi_y \rangle \\ \langle \varphi_x, \varphi_y \rangle & \langle \varphi_y, \varphi_y \rangle \end{bmatrix}.$$

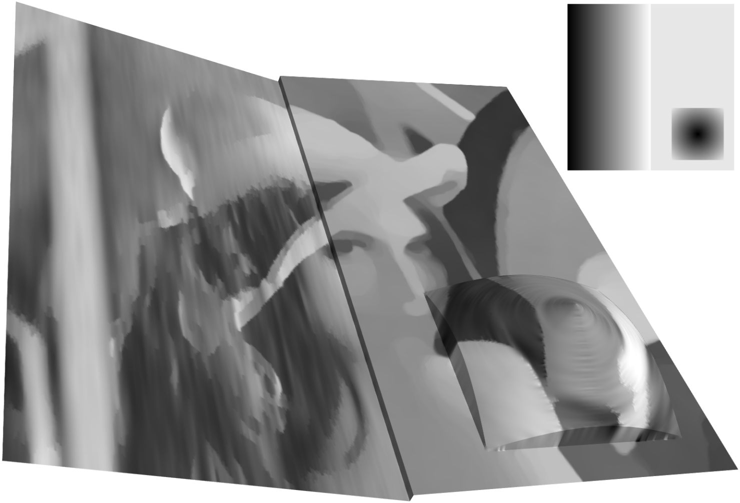
This allows us to define the TV norm on S as

$$TV_g(s) = \int_S |\nabla_g s| \, \mathrm{d}s = \int_\Omega |\nabla_g s| \sqrt{det(g)} \, \mathrm{d}x\mathrm{d}y.$$

Effect of Regularization on a Surface

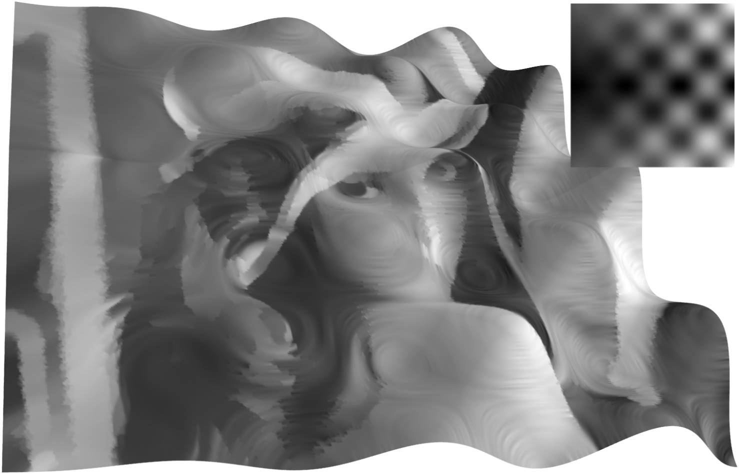
ROF denoising on a flat surface

Effect of Regularization on a Surface



ROF denoising on a ramp

Effect of Regularization on a Surface



ROF denoising on a sine wave

Data Term

Measures the distance of the optimized image u to the current data f^n under some camera noise model assumptions.

$\mathbf{D(u, f^n)}$	**Result**	**Convex**
$\|u - f^n\|$	✗	✓

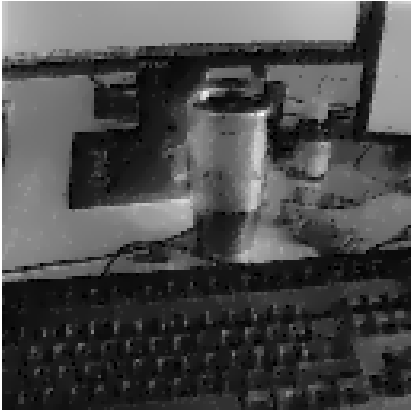
Data Term

Measures the distance of the optimized image u to the current data f^n under some camera noise model assumptions.

$\mathbf{D(u, f^n)}$	**Result**	**Convex**
$\|u - f^n\|$	✗	✓
$\|u - f^n\|^2$	(✗)	✓

Data Term

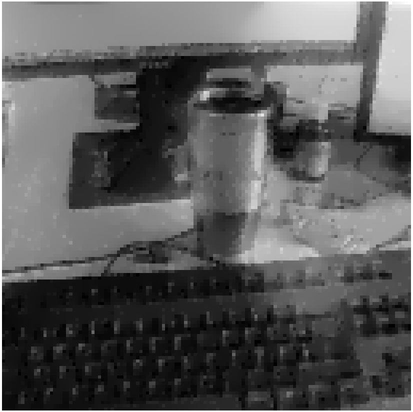
Measures the distance of the optimized image u to the current data f^n under some camera noise model assumptions.



$\mathbf{D(u, f^n)}$	**Result**	**Convex**
$\|u - f^n\|$	✗	✓
$\|u - f^n\|^2$	(✗)	✓
$\sum_i \log\left(\frac{u_i}{f_i^n}\right)^2$	✓	✗

Data Term

Measures the distance of the optimized image u to the current data f^n under some camera noise model assumptions.



$\mathbf{D(u, f^n)}$	**Result**	**Convex**
$\|u - f^n\|$	✗	✓
$\|u - f^n\|^2$	(✗)	✓
$\sum_i \log\left(\frac{u_i}{f_i^n}\right)^2$	✓	✗
$\sum_i u_i - f_i^n \log u_i$	✓	✓

Data Term

We therefore choose:

$$\lambda \int_S (u - f^n \log u)\, \mathrm{d}s = \lambda \int_\Omega (u - f^n \log u) \sqrt{G}\, \mathrm{d}x\mathrm{d}y \ ,$$

known as *generalised Kullback-Leibler divergence*.

- Minimiser is the ML-estimate when assuming Poisson noise between u and f (dependent on the absolute light intensity) [Ratner and Schechner '07]
- Convex $\rightarrow$ easy to minimise

Energy Minimisation

We minimze the original energy using the primal-dual energy formulation

$$\min_u \max_p \left[D(u, f^n) + \langle L_g u, p \rangle - R^*(p) \right], \tag{1}$$

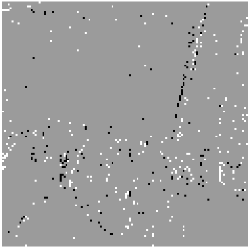
with L_g being the discretised operator ∇_g with the algorithm of [Chambolle and Pock '11].

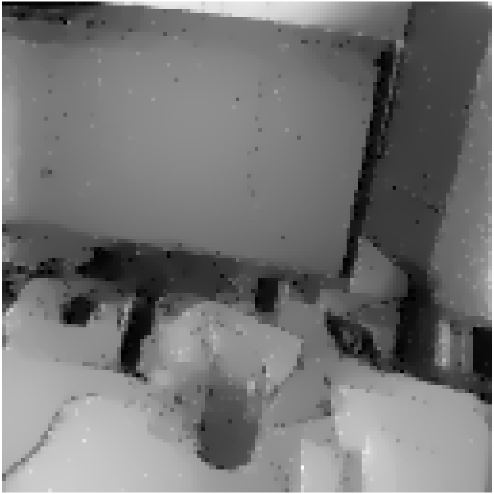
- Algorithm is still defined in pixel-space $\rightarrow$ parallelizable on GPUs
- Converges in less than 50 iterations due to small image size of 128×128

Experiments - Timing

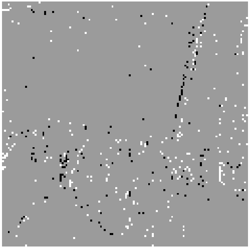
- NVidia GTX 780 ti - TitanX
- $\approx$ 600 frames/sec
- $\approx$ 500.000 events/sec $\rightarrow$ collect 500-1000 events before denoising

Experiments - Influence of Manifold Regularization

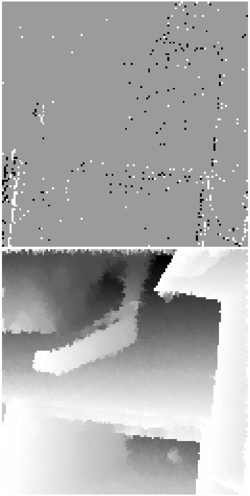




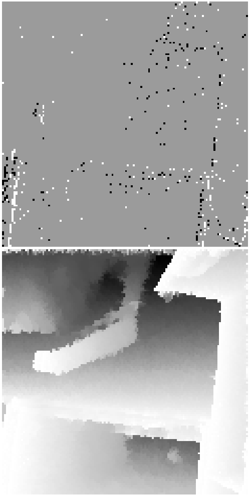
Experiments - Influence of Manifold Regularization



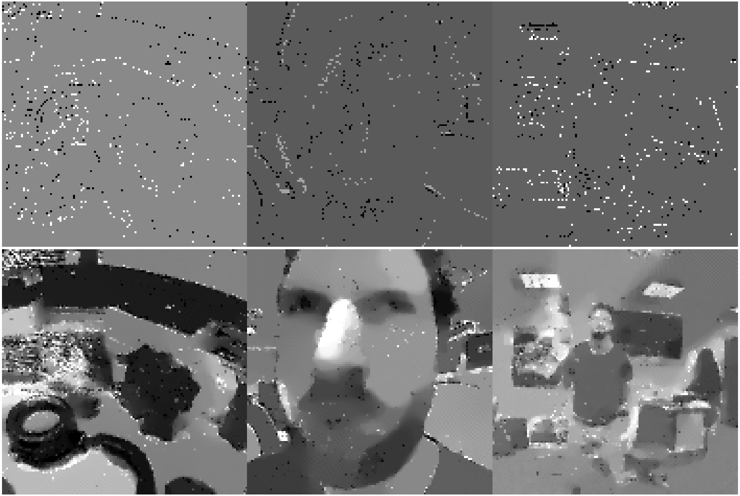
Experiments - Influence of Manifold Regularization



Experiments - Influence of Manifold Regularization

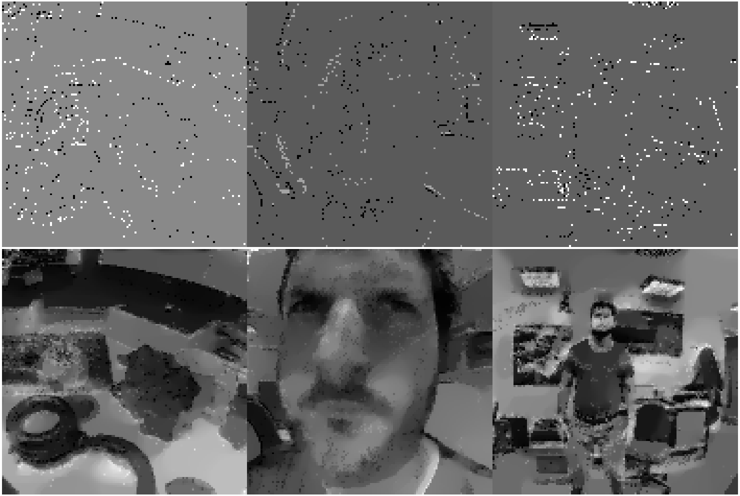


Experiments - Comparison to [Bardow et al. '16]



[Bardow et al. '16]

Experiments - Comparison to [Bardow et al. '16]



OURS

Experiments - Video

Conclusion

- Image Reconstruction without camera tracking
- Fast enough to achieve Real-Time performance
- Exploitation of the *Surface of Active Events*

Software for DVS128 and DAVIS240 can be downloaded from
https://github.com/VLOGroup

Outlook and Open Questions

- Definition of data term (real camera noise model still unknown)
- Quantitative evaluation of the result (blind Image Quality Analysis or comparison to ground truth)
- Weighting of data term dependent on the number of events per reconstructed image

Thank you for your attention!

Log L2 Data Term

$$\breve{d}(t) = \begin{cases} d(t) & \text{if } \log t < 1 \\ d(e) + d'(e)(t-e) = -\frac{1}{2} + \frac{t}{e} & \text{else} \end{cases}$$