The interplay between events and frames: A comprehensive explanation

R.B. Benosman,
Meta Reality Lab Research (Surreal)
New York

Meta

Structure vs Statistics

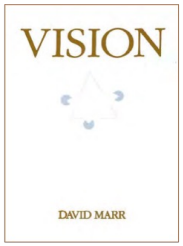
R.B. Benosman,
Meta Reality Lab Research (Surreal)
New York

Historic Timeline

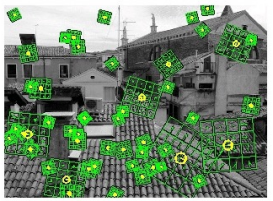
Franck Rosenblatt: Perceptron

Russell's Infant Son: 5cm by 5cm (176x176 array) Portland Art Museum.

MIT summer Vision Project, Seymour Papert, automatically, background/foreground segmentation, extract non-overlapping objects

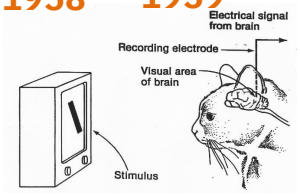


David Marr, "Vision a computational investigation into the human representtion and processing of visual information" vision is hierarchical



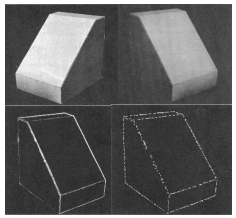
Feature based Pattern recognition. Keypoints
3D reconstruction gets "solved", generic object recognition

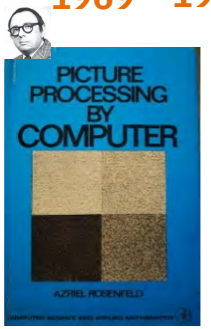
1958 1959 1963 1966 1969 1982 1990' 2000'



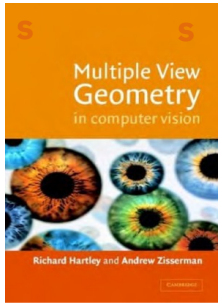
David Hubel and Torsten Wiesel—in 1959. Their publication, entitled "*Receptive fields of single neurons in the cat's striate cortex*"

Lawrence Roberts' "Machine perception of three-dimensional solids". Process 2D photographs to build up 3D representations from lines.





Azriel Rosenfeld Early applications of image analysis



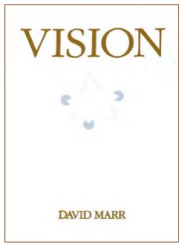
Vision is ruled by Geometry, (projective) 3D reconstruction, fundamental matrix, RANSAC, bundle,

Historic Timeline

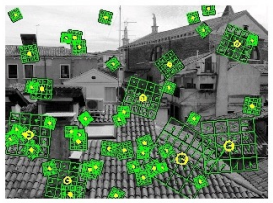
k
lblatt:
otron

Russell's Infant Son: 5cm by 5cm (176x176 array) Portland Art Museum.

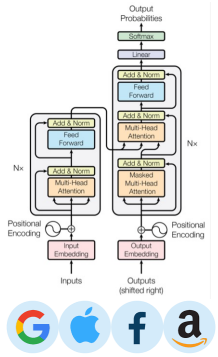
MIT summer Vision Project, Seymour Papert, automatically, background/foreground segmentation, extract non-overlapping objects



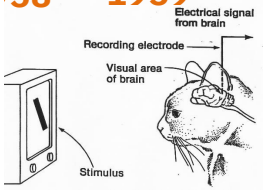
David Marr,
"Vision a computational investigation into the human representtion and processing of visual information" vision is hierarchical



Feature based Pattern recognition.
Keypoints
3D reconstruction gets "solved", generic object recognition

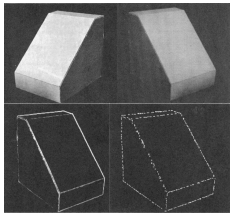


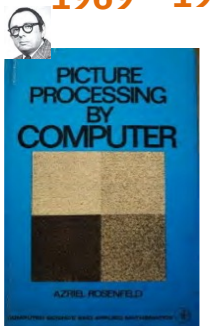
58 **1959** **1963** **1966** **1969** **1982** **1990's** **2000's** **2000's**



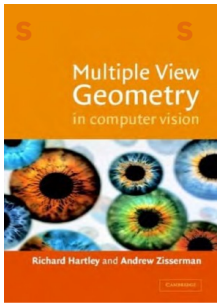
David Hubel and Torsten Wiesel—in 1959. Their publication, entitled "*Receptive fields of single neurons in the cat's striate cortex*"

Lawrence Roberts' "Machine perception of three-dimensional solids". Process 2D photographs to build up 3D representations from lines.

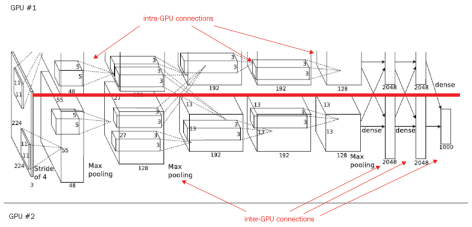




Azriel Rosenfeld
Early applications of image analysis



Vision is ruled by Geometry, (projective) 3D reconstruction, fundamental matrix, RANSAC, bundle,



?

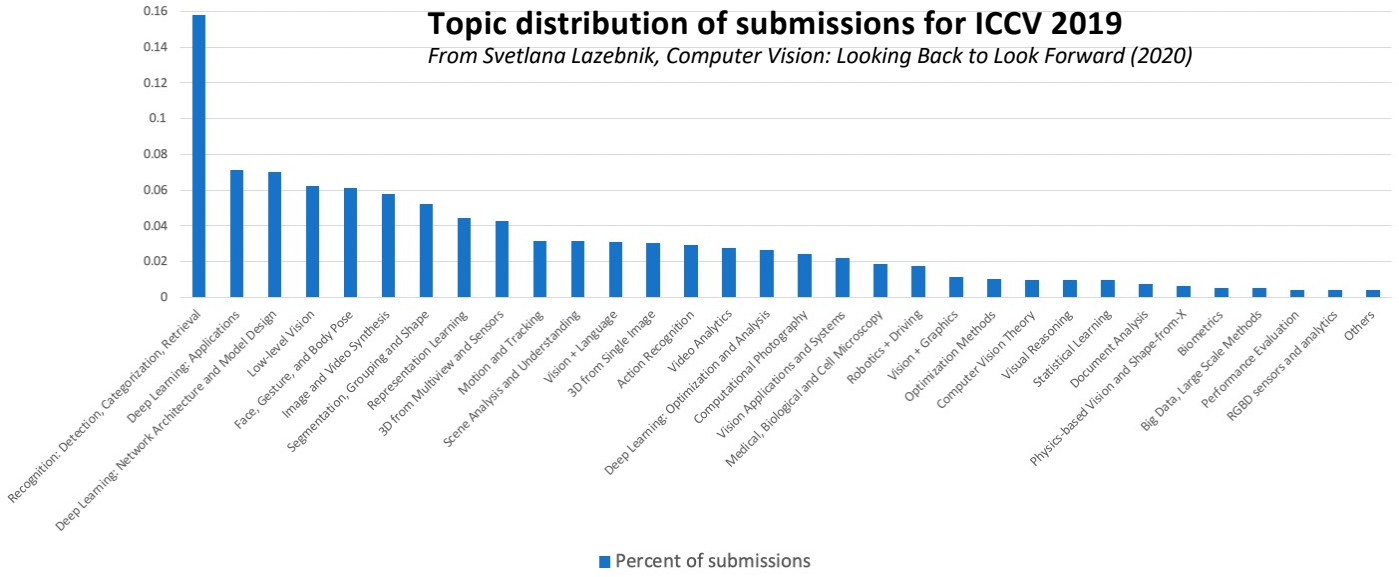
What Went Wrong?

• Computer vision has been reinvented at least three times.

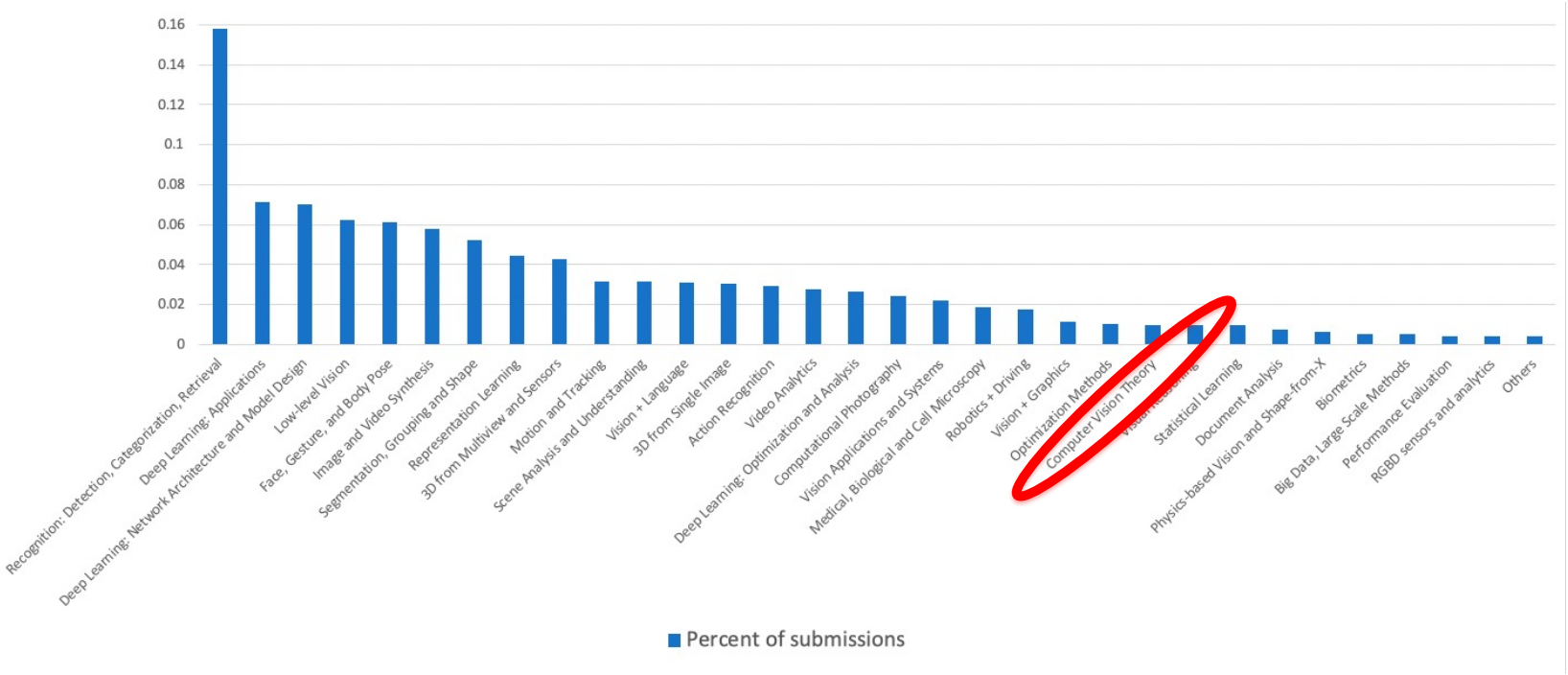
• Too close to the market: **applications based research**

Tendency to **resist novelty** choosing applications over potentially more promising methods that could not yet deliver

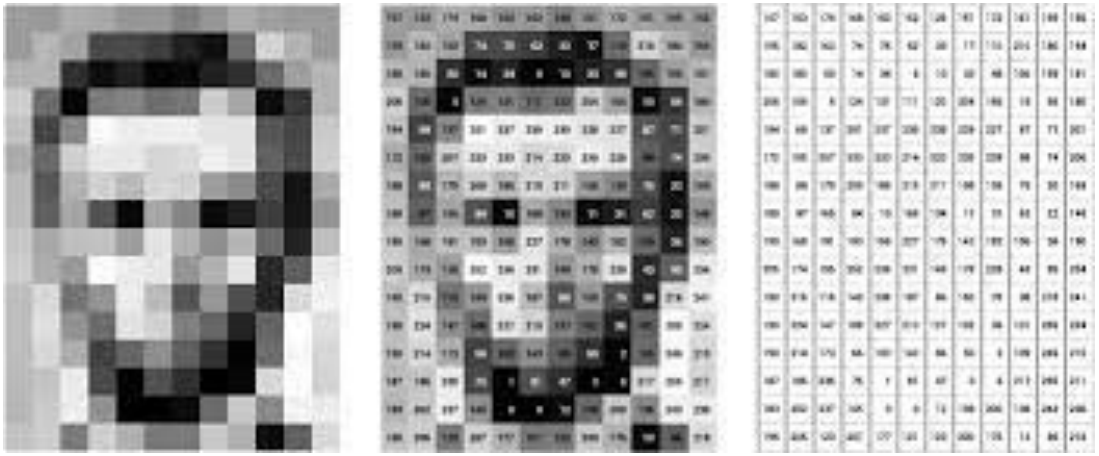
• **Not idea driven**



What Went Wrong?

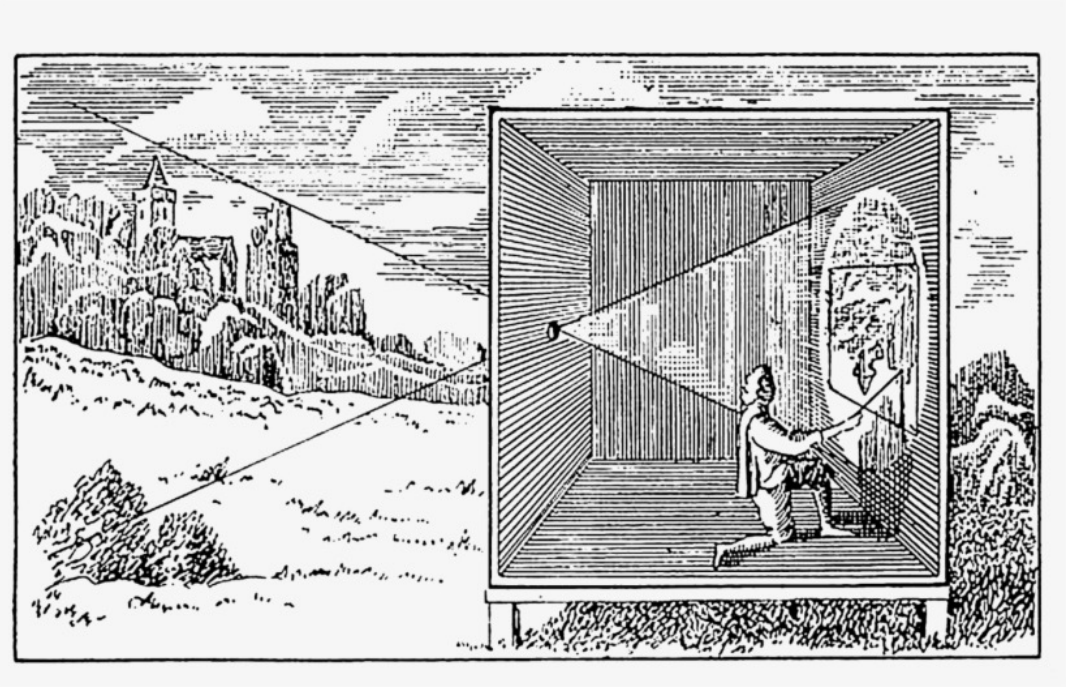


Why are we using images?



- **Images** are the optimal **structure of data**
- **Grey Levels** as **source of information**

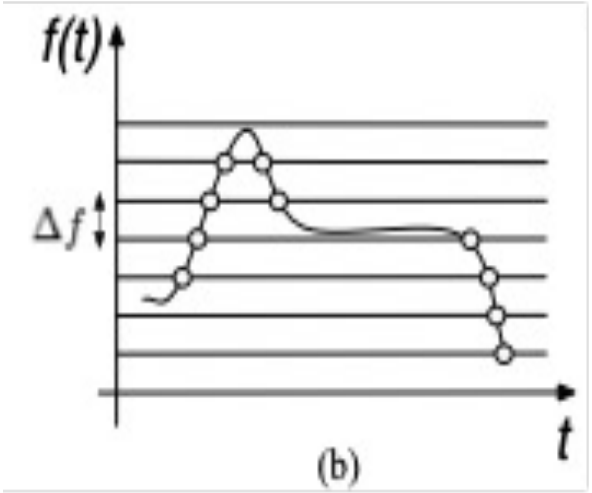
Computer Vision: a Heritage from Art!

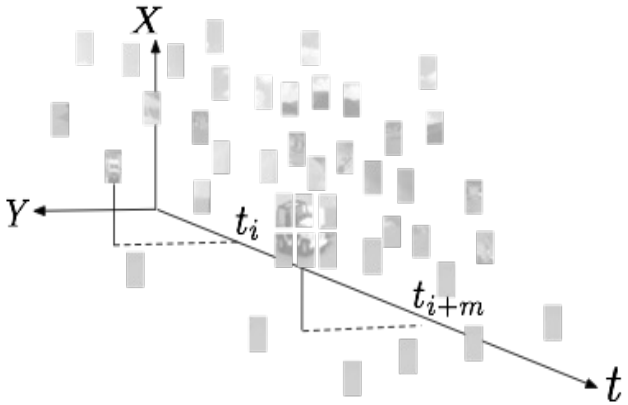


- Invention of the camera obscura in 1544 (L. Da Vinci?)
- The mother of all cameras

Event acquisition

Popular solution: Sample on the amplitude axis of signals



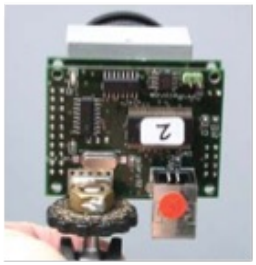


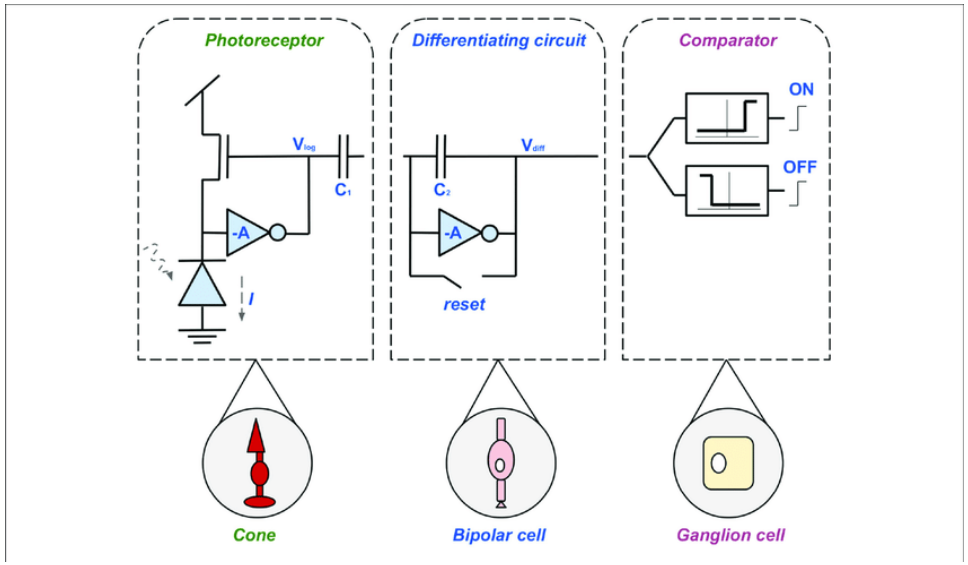
- New Information is detected when it happens
- When nothing happens, nothing is sent or processed
- Sparse information coding

Time is the most valuable information

A 128×128 120dB 15us Latency Asynchronous Temporal Contrast Vision Sensor

Patrick Lichtsteiner, Christoph Posch, and Tobi Delbruck, *Member, IEEE*

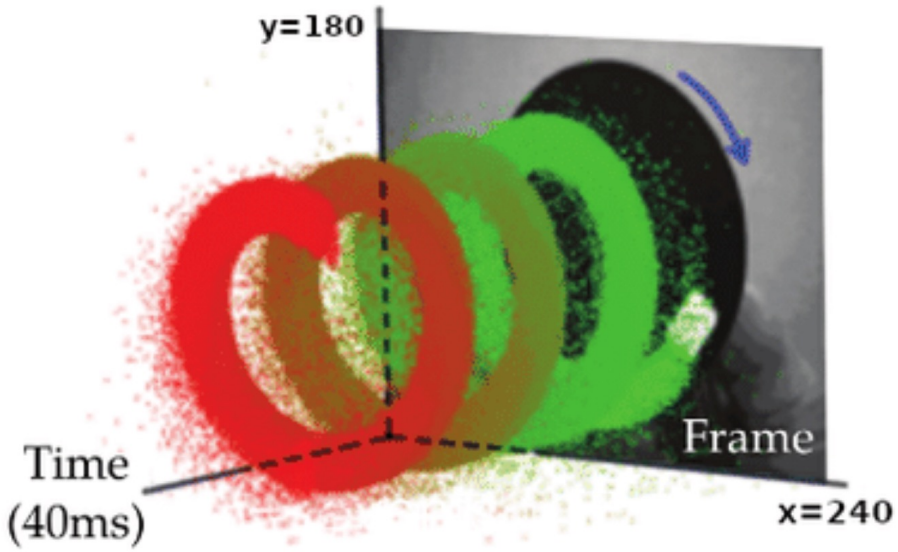




$$\left|\Delta \log I\right| > T$$

A 240×180 10mW 12us latency sparse-output vision sensor for mobile applications

R. Berner, Christian Brandli, +2 authors T. Delbruck • Published 12 June 2013 • Computer Science •
2013 Symposium on VLSI Circuits



Event based is a companion chip for frames applications: deblurring, contrast …..

Event Acquisition

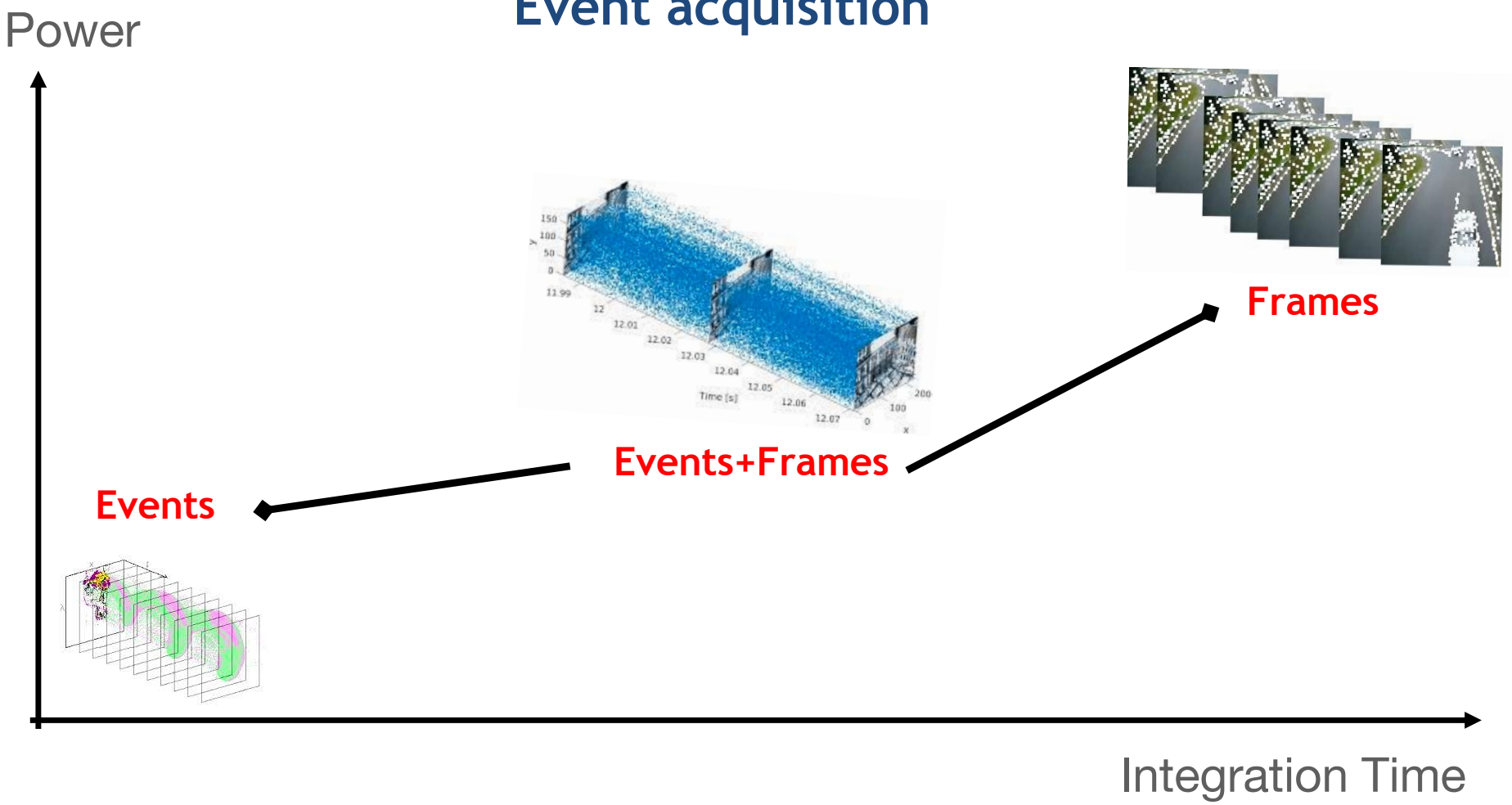
Scopes:

- **Reduce Data Load** and **only** Detect "**meaningful**" events, **at the time they happen!**
- Avoid **burning energy** to acquire, transmit and store information that ends up **being trashed**

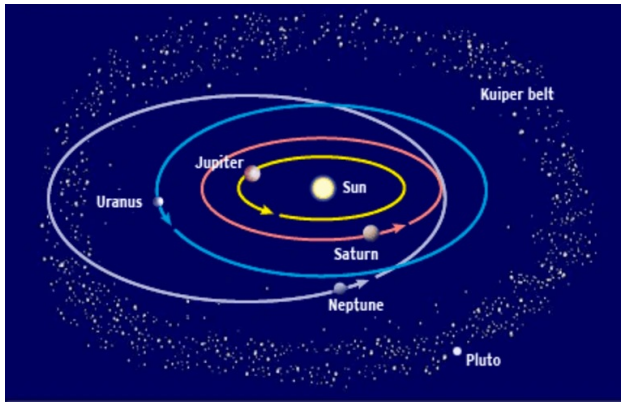
Solutions:

- **No generic solution**,
- There are almost an **infinite number** of solutions to extract events
- Need to be **adapted to the dynamics and nature of the data**

Event acquisition



Structure vs Statistics



Knowing

Structure: explores the systemic **organization**, patterns, and spatial **relationships** of objects, providing **explanations and insights into the mechanisms** of **how things are arranged and interconnected**.

Understanding

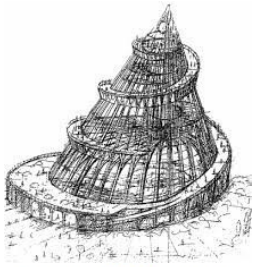
Statistics: explore numerical measures and **distributions**, structure analysis **capturing the appearance phenotype** of the subject being studied.

Structure vs Statistics

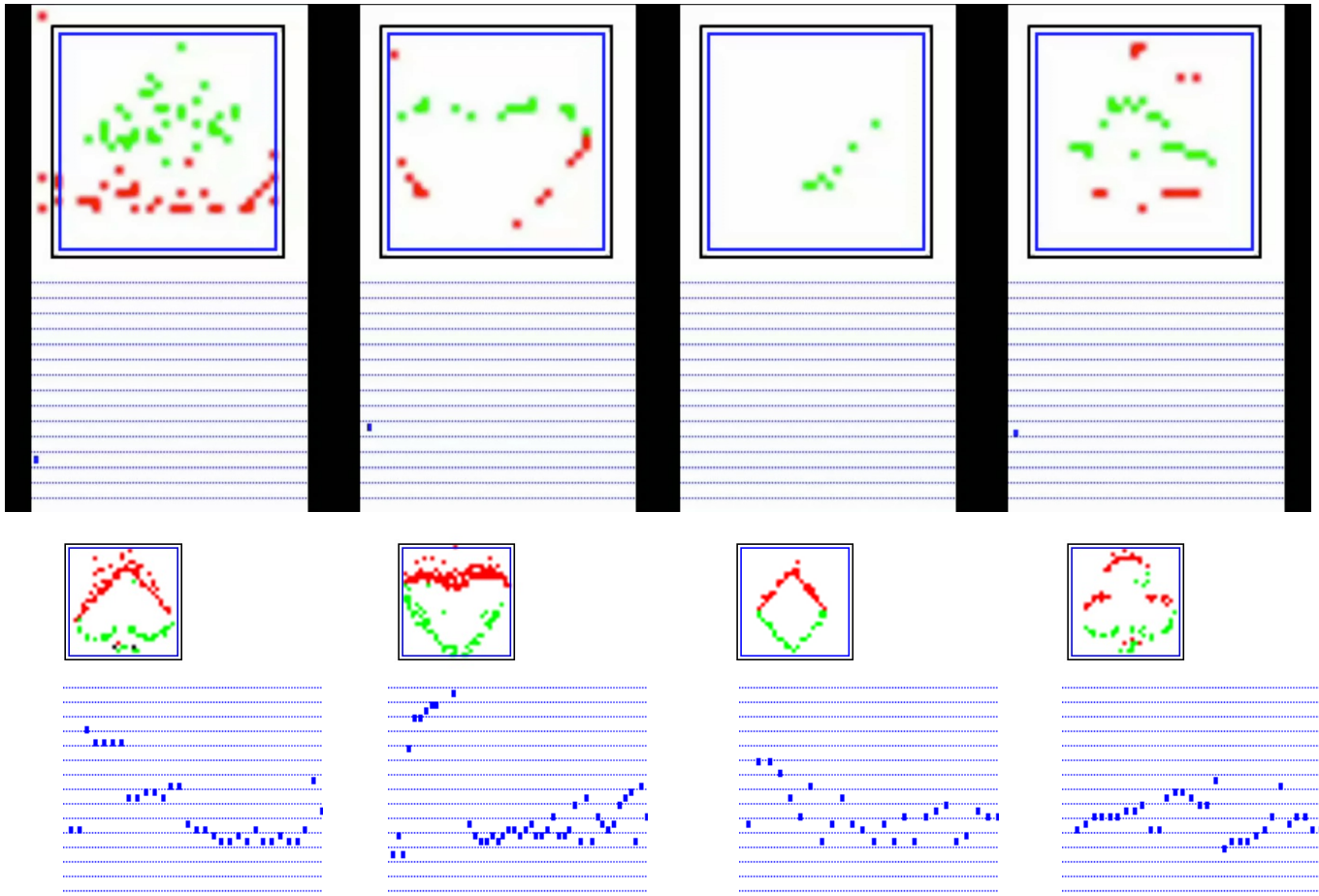
All cats look different in appearance

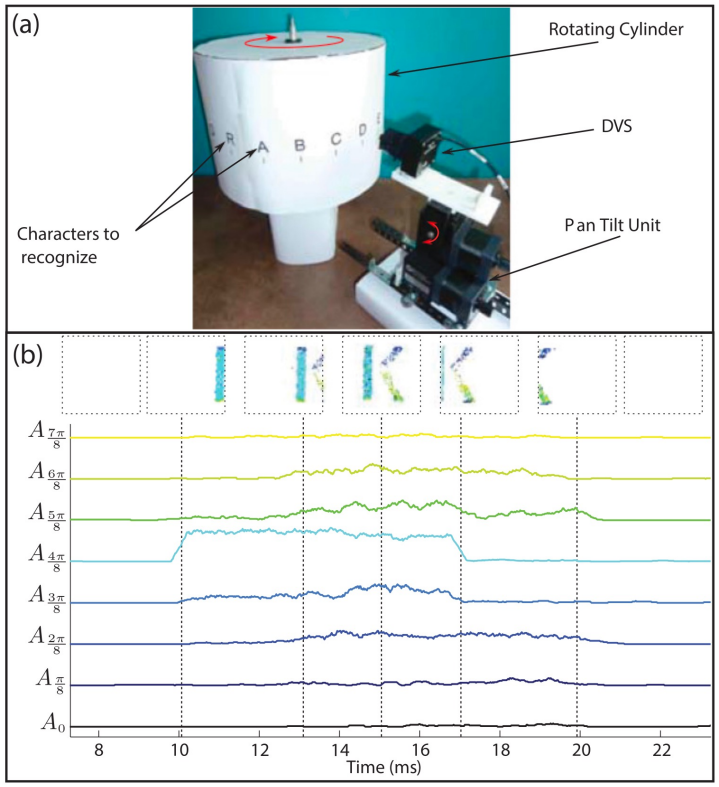
They are structurally identical

1. **Pixel Intensity Statistics:** Mean, Median, Mode,Standard deviation, Minimum, Maximu, Histogram (intensity distribution)
2. **Color Channel Statistics:**Mean, median, mode, standard deviation (for each color channel), Histogram (for each color channel)
3. **Texture Statistics:** Co-occurrence matrices (contrast, correlation, energy, homogeneity), Entropy, Gabor filters, Local binary patterns (LBP), Haralick features
4. **Shape Statistics:** Area, Perimeter, Aspect ratio, Orientation, Eccentricity, Convexity, Circularity
5. **Object Count and Size Statistics:** Object count, Object size distribution, Object centroid coordinates
6. **Spatial Statistics:** Distance measures (Euclidean distance, Manhattan distance), Nearest neighbor analysis, Cluster analysis (k-means, hierarchical clustering), Point pattern analysis (Ripley's K-function, nearest neighbor index)
7. **Frequency Domain Statistics:** Fourier transform, Power spectrum, Frequency histogram, Low-pass and high-pass filtering
8. **Motion Statistics (for video/image sequences):** Optical flow, Velocity, Direction, Motion histogram, Trajectory analysis



Data Structure

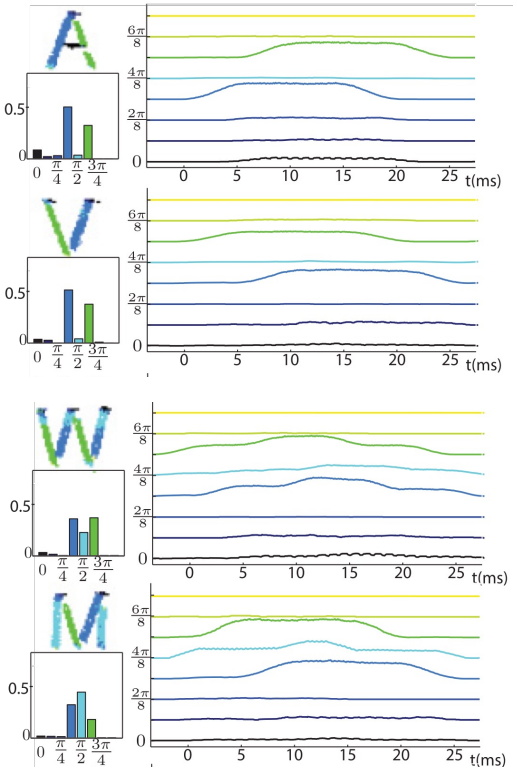




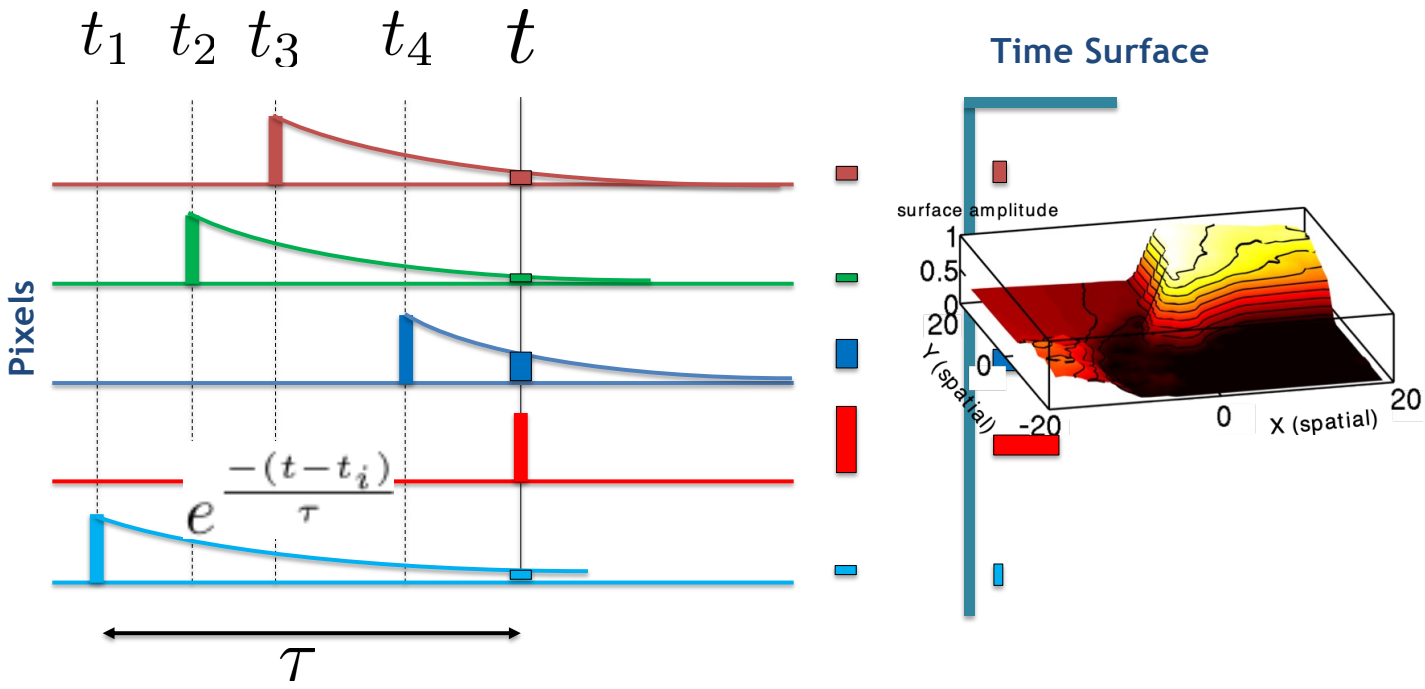
LETTER Communicated by Bertram Shi

What Can Neuromorphic Event-Driven Precise Timing Add to Spike-Based Pattern Recognition?

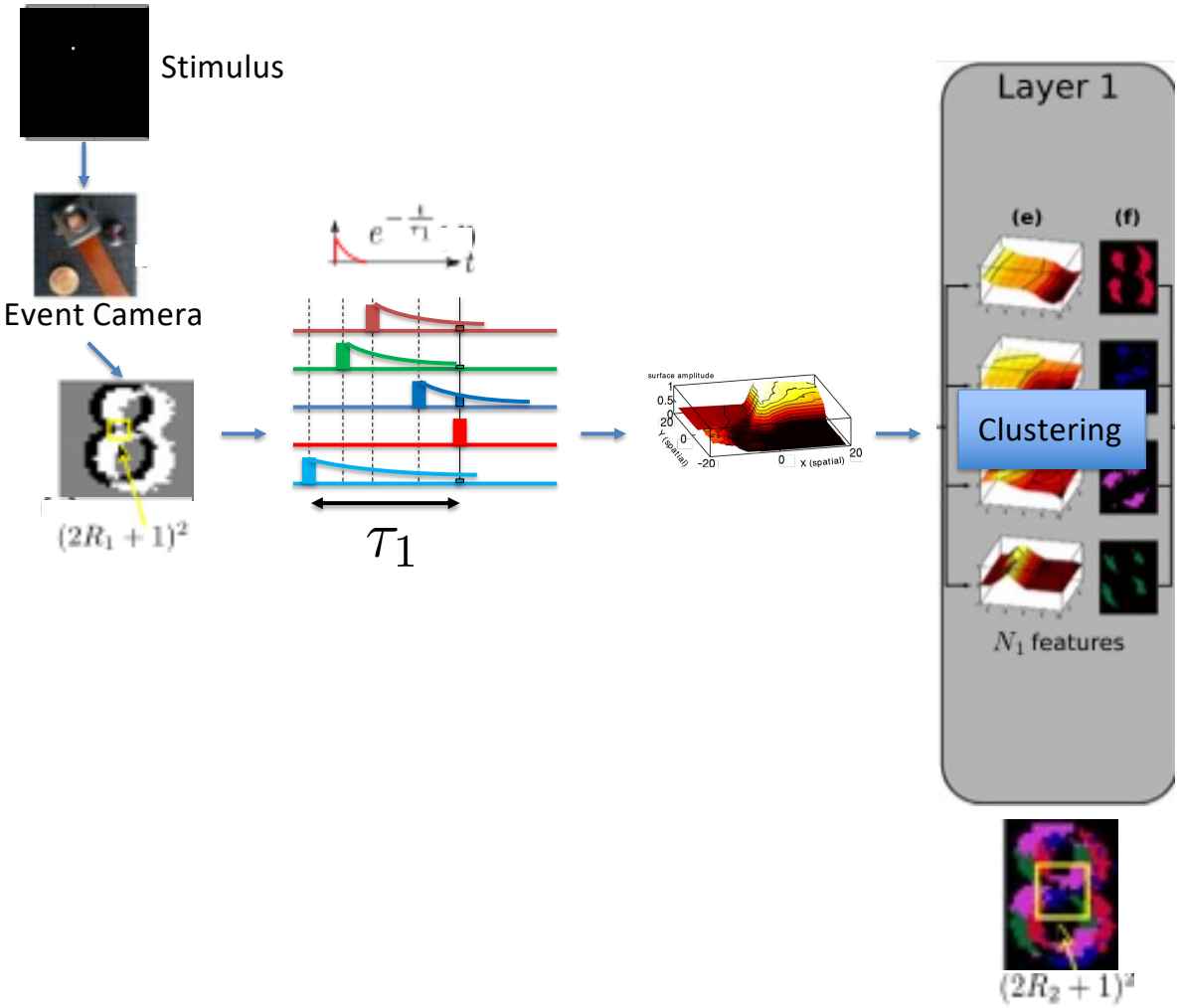
Neural Computation **27, 561–593** (2015)
doi:10.1162/NECO_a_00703

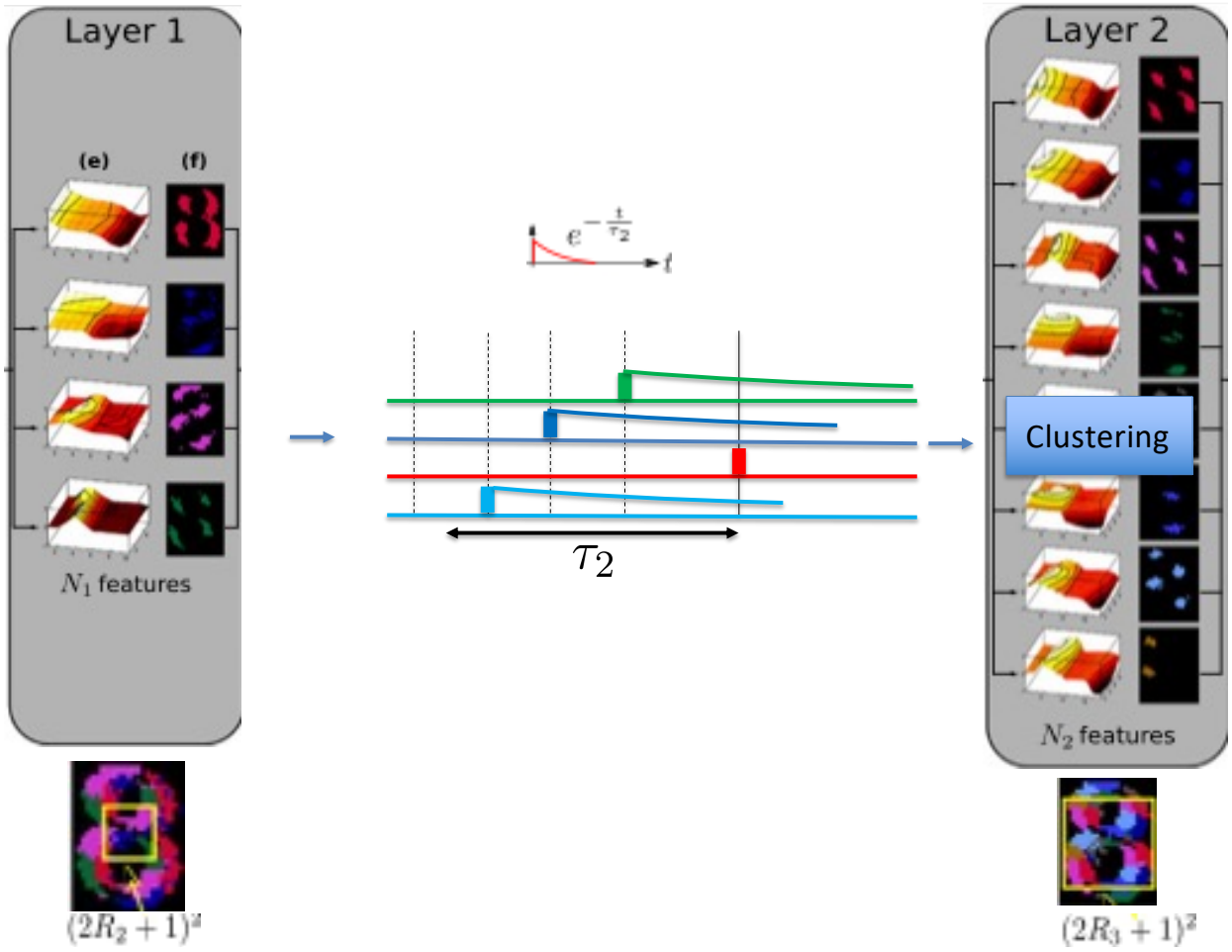


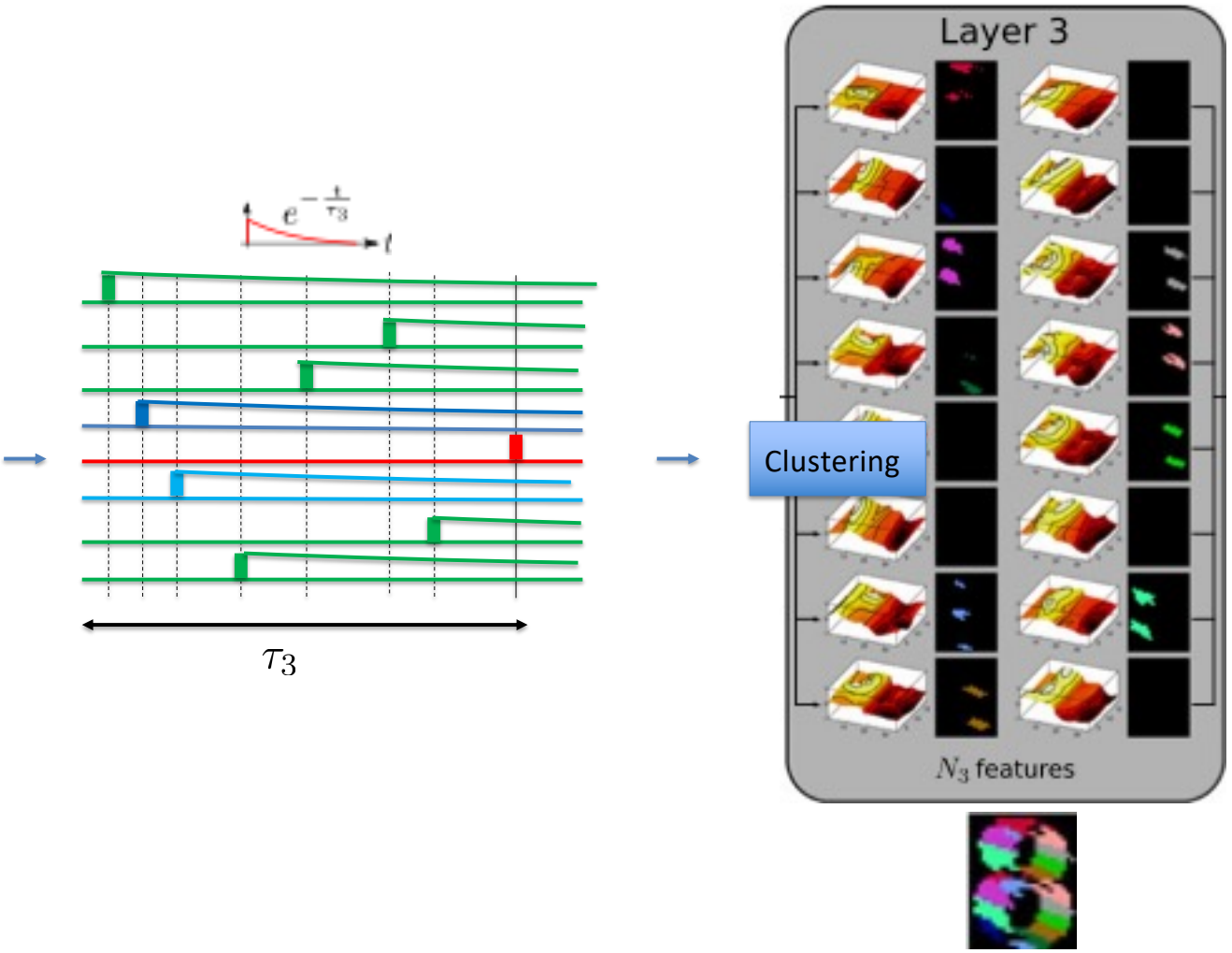
Deep temporal learning: Time Surfaces

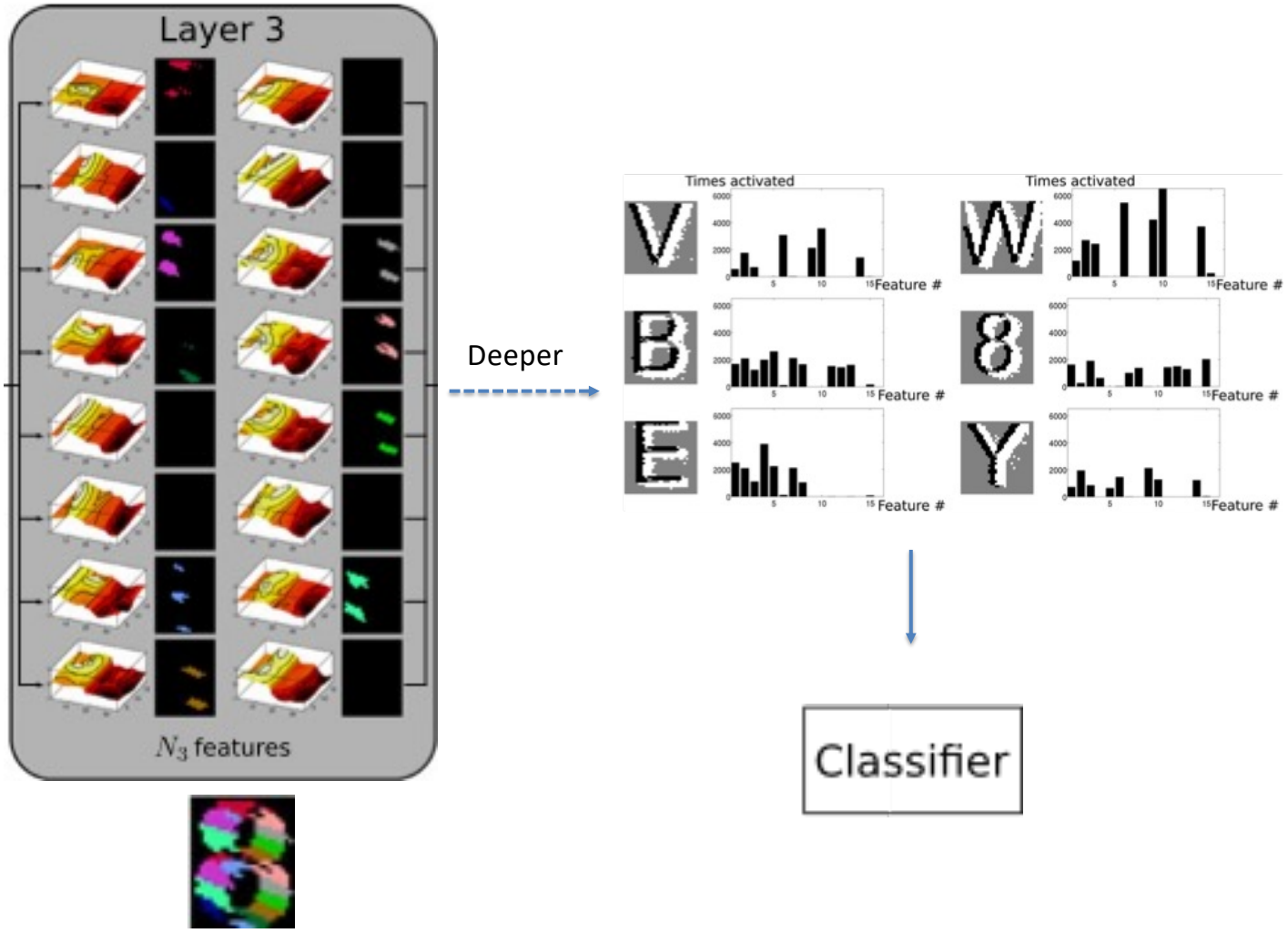


X. Lagorce, G. Orchard, F. Galluppi, B. E. Shi, R.B. Benosman, HOTS : A Hierarchy Of event- based Time-Surfaces for pattern recognition, 2016 July 11, IEEE Transaction on Pattern Analysis and Machine Intelligence 39(7) :1346-1359, doi :10.1109/TPAMI.2016.2574707

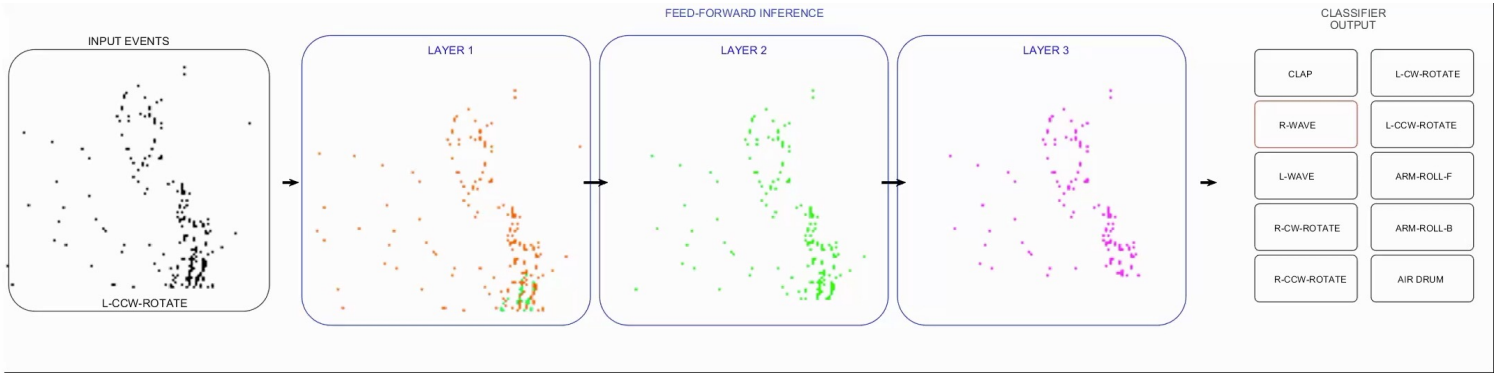

Stimulus
Event Camera
$(2R_1+1)^2$
$e^{-\frac{t}{\tau_1}}$
t
τ_1
surface amplitude
Y (spatial)
X (spatial)
Layer 1
(e)
(f)
Clustering
N_1 features
$(2R_2+1)^2$


Layer 1
(e)
(f)
N_1 features
$(2R_2+1)^2$
$e^{-\frac{t}{\tau_2}}$
t
τ_2
Clustering
Layer 2
N_2 features
$(2R_3+1)^2$

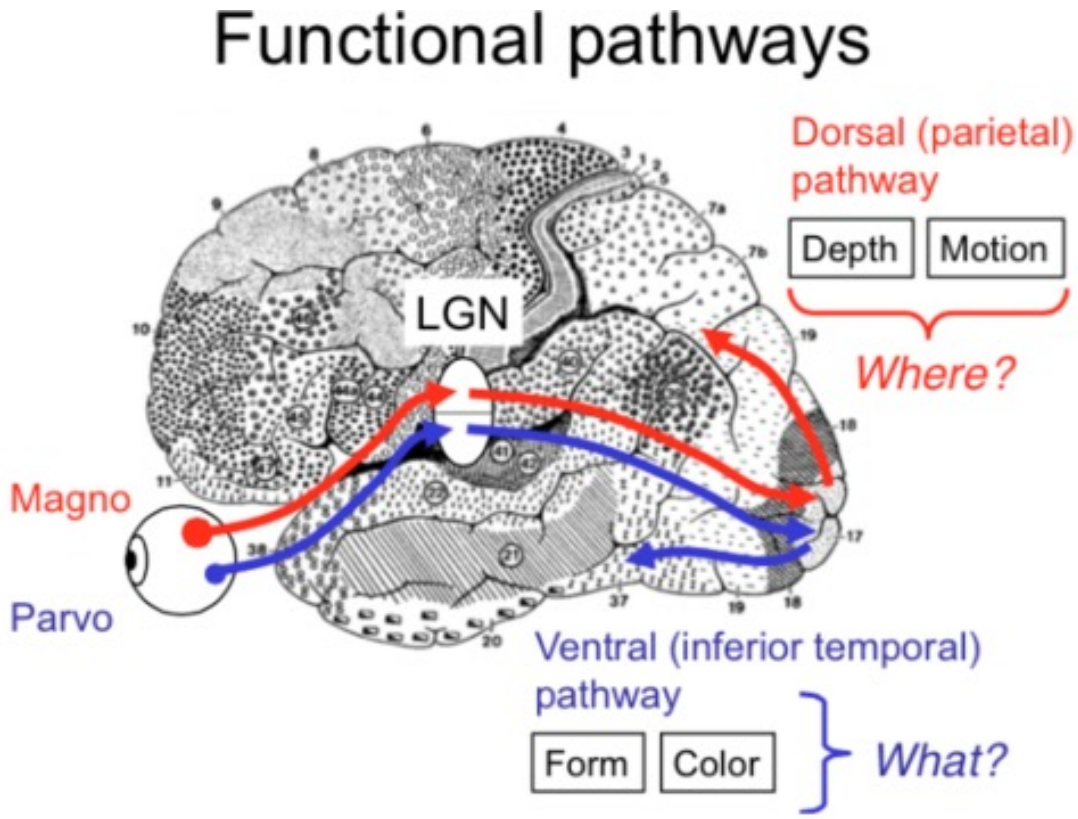

$e^{-\frac{t}{\tau_3}}$
t
τ_3
Layer 3
Clustering
N_3 features


Layer 3
N_3 features
Deeper
Times activated
Feature #
Times activated
Feature #
Classifier

Deep temporal learning with adaptive temporal feedback: Temporal Surfaces

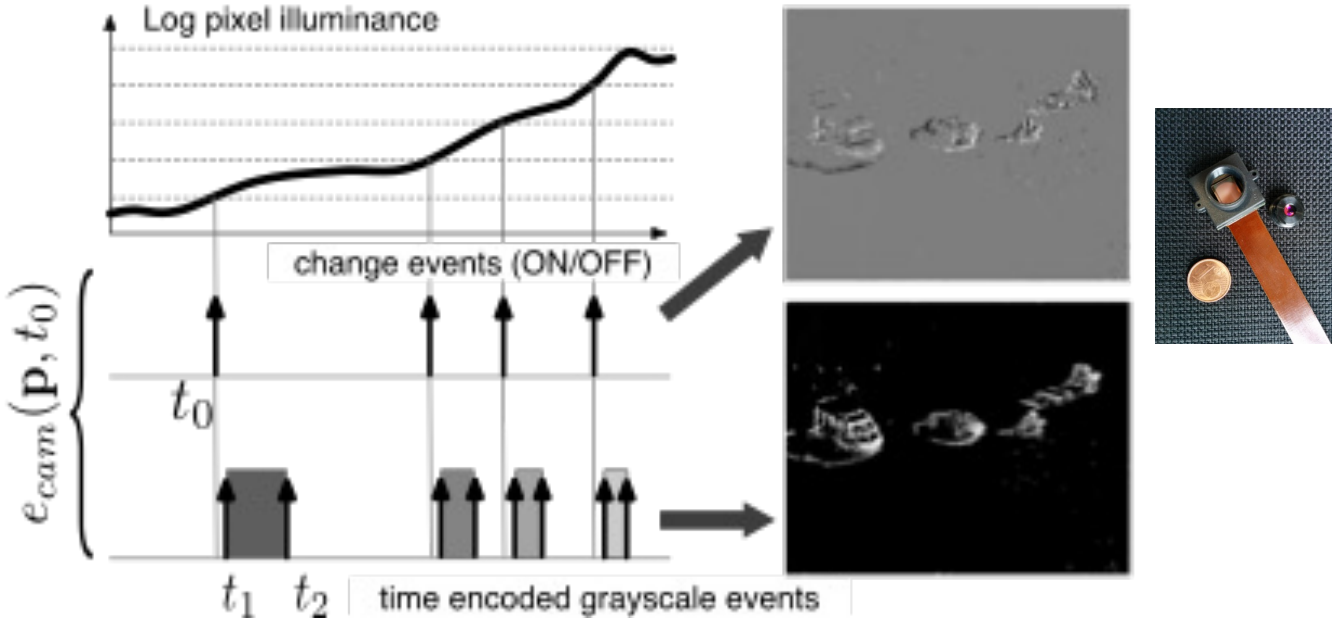


Structure and Statistics: Two sides of the same coin?



A QVGA 143 dB Dynamic Range Frame-Free PWM Image Sensor With Lossless Pixel-Level Video Compression and Time-Domain CDS

Christoph Posch, *Member, IEEE*, Daniel Matolin, and Rainer Wohlgenannt



Temporal events and **absolute light measurement**

Frames vs Events

conventional
frame-based camera

event-based camera

Asynchronous Event-Based Fourier Analysis

Quentin Sabatier, Sio-Hoi Ieng, and Ryad Benosman

***Abstract*— This paper introduces a method to compute the FFT of a visual scene at a high temporal precision of around 1-μs output from an asynchronous event-based camera. Event-based cameras allow to go beyond the widespread and ingrained belief**

the frame rate is set to, it will always be wrong because there is no relation whatsoever between dynamics present in a scene and the chosen frame rate, over-sampling and/or under-sampling occur, and moreover both usually happen

INVITED PAPER

Asynchronous Neuromorphic Event-Driven Image Filtering

Sparse coding of spatio–temporal signals lowers computational cost and raises the efficiency of visual processing.

By Sio-Hoï Ieng, Christoph Posch, *Senior Member IEEE*, and Ryad Benosman

ORIGINAL RESEARCH article
Front. Neurosci., 12 June 2018
Sec. Neuromorphic Engineering
Volume 12 - 2018 | https://doi.org/10.3389/fnins.2018.00373

Complexity Analysis of Iterative Basis Transformations Applied to Event-Based Signals

Sio-Hoi Ieng[1*], Eero Lehtonen[2] and Ryad Benosman[1]

Conclusions

- **Images** and **Events** contain **complementary information**
- Structure is what allows single shot learning
- Emphasizes the presence of interactions between different time scales
- This is a **novel paradigm for Machine Learning** allowed by the temporal properties of event sensors
- Images are useful but as a slower process for disambiguation: Fast and Slow pathways