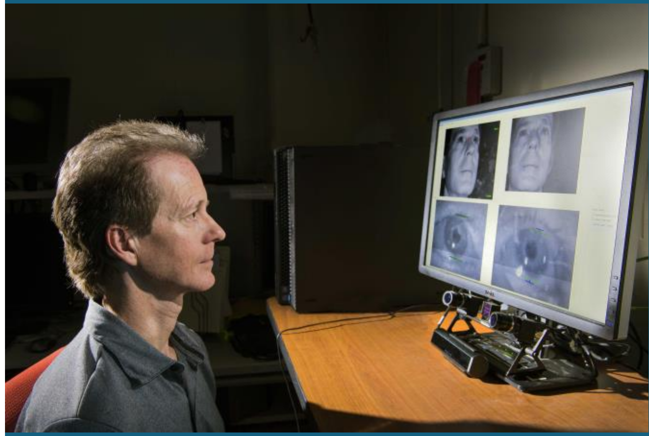
Seeing into the black box:

Using eye tracking in user-driven workflows to better understand decision-making processes



Kristin M. Divis (kmdivis@sandia.gov | 505.284.6561)

Max G. Chen, Laura A. McNamara, & J. Dan Morrow

Sandia National Laboratories

U.S. DEPARTMENT OF ENERGY

NNSA

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia LLC, a wholly owned subsidiary of Honeywell International Inc. for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

SAND 18-771612

Information Foraging Theory

- Human 'informavores'
- Valuable information isn't distributed evenly
- We look for *proximal semantic cues* to navigate to patches of interest
- Effective foraging strategies *maximize the value of information gained* relative to *effort finding the information*

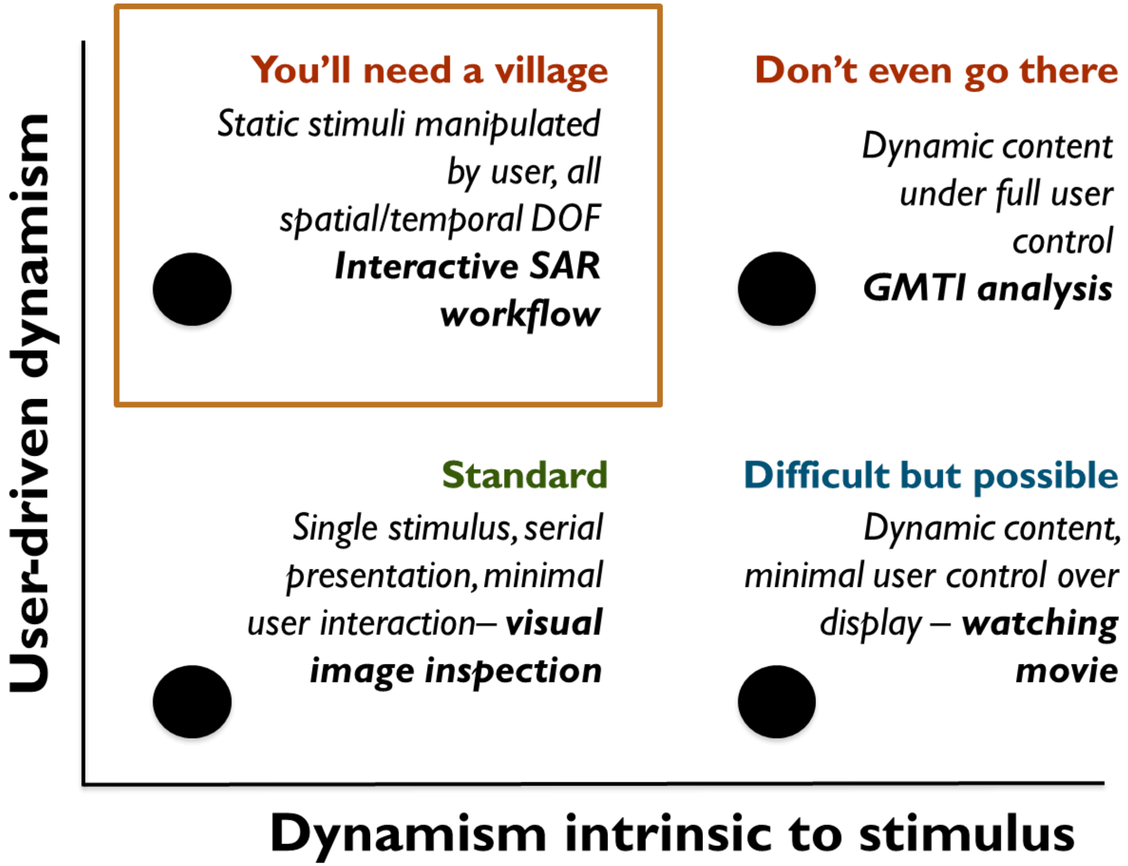
(see Pirolli & Card, 1999)

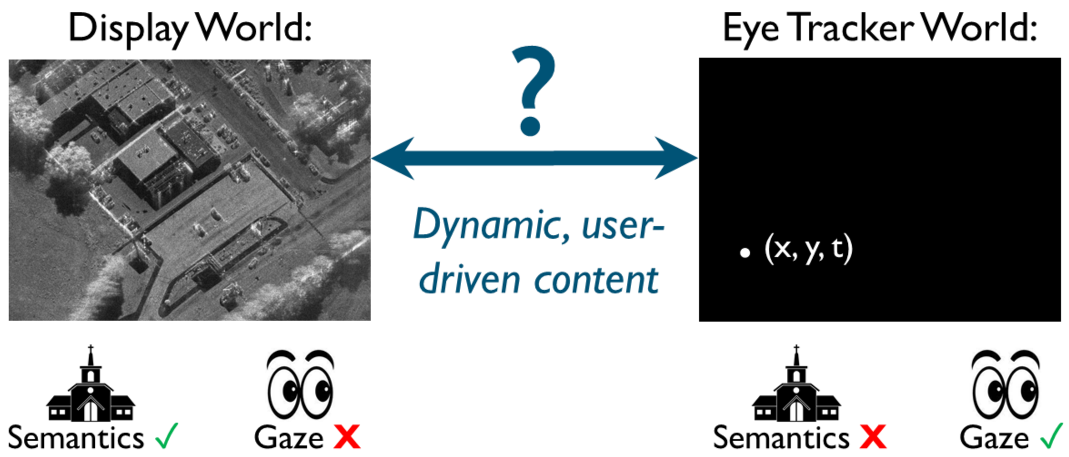
…so we "just" need to capture markers of analysts' search behavior?

- Design a simple signature detection task, invite SAR Imagery Analysts to find the targets (capturing interaction logs)
- Need to know *where* in an image the analyst looks (**gaze-contingent decision patterns**)

Great idea … but eye trackers couldn't support the workflows we wanted to study

The underlying problem:





What do we need?

Ingredients for a Simple Gaze-Informed Foraging Model:

Dynamically changing content →

Gaze tracked against display surface →

User operations (mouse, keyboard) →

Application/content log files →

VECTOR OF EVENTS

integrated into a single timeline with minimal latency

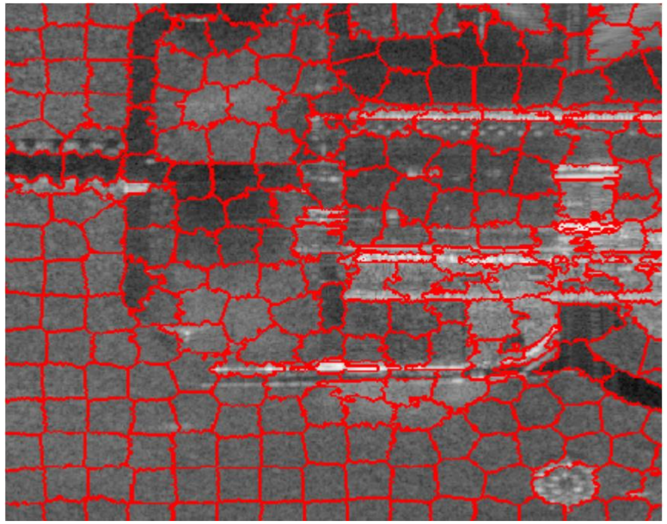
T_0 → T_n

Step 1: Data "Sandbox" Creation

Synthetic Aperture Radar (SAR) images with rich metadata

- Each pixel tagged with info such as geospatial location
- Superpixel and megapixel algorithms established

4 constrained visual search task (zoomed in; click and drag to pan) to establish ground truth



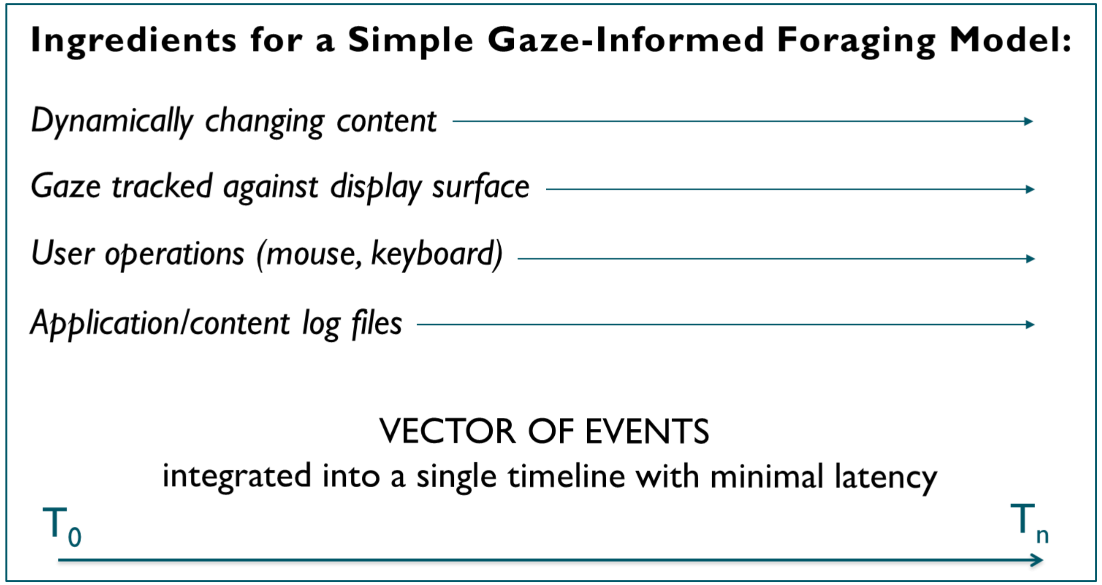
Superpixel Example
(Moya et al., 2014)

Step 2: Sequence of Events

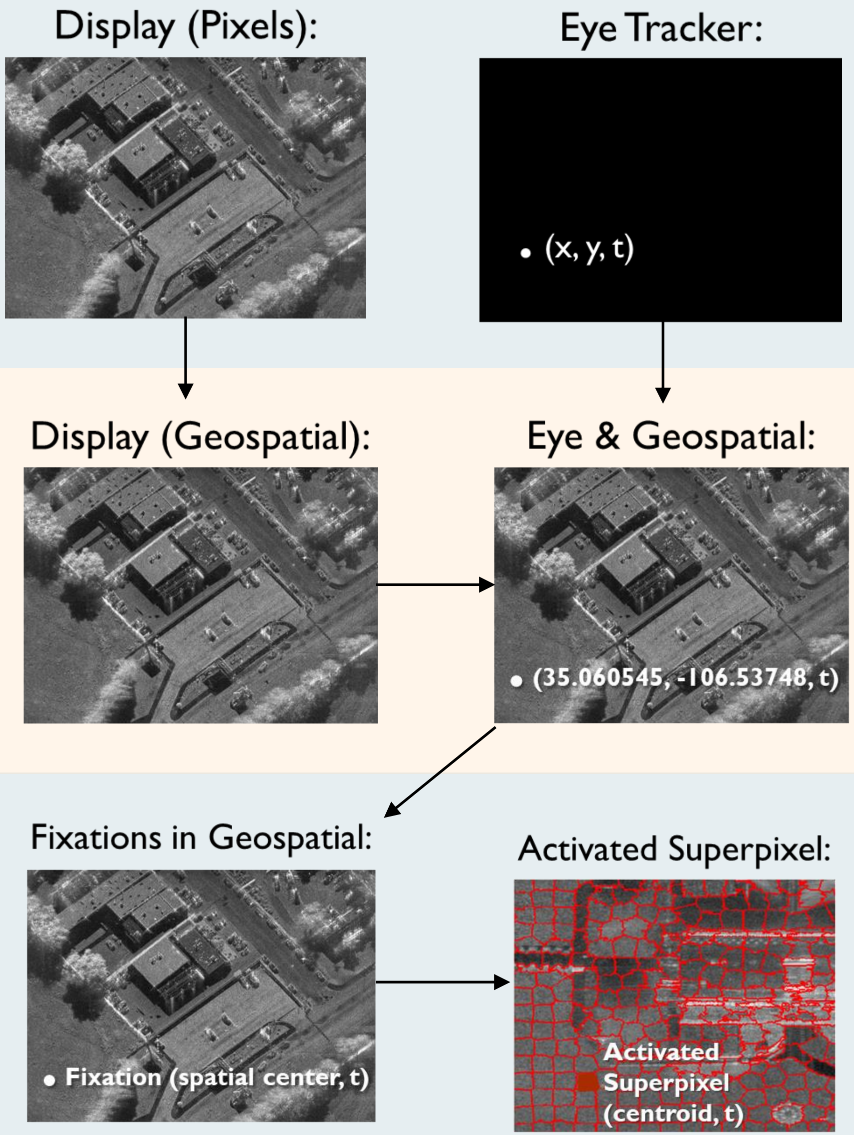
Map data streams into a **sequence** of events.

Requires:

- Integration of data streams
- Dimensionality reduction
- Temporal merging of data streams into a single sequence



Gaze data: Integration and dimensionality reduction



Starting place:

- Raw eye movement data (x,y,t) at every sample (60 Hz) in screen space
- Image coordinates of upper left of display

Simple transformations:

- Transform image coordinates to underlying geospatial coordinates (encoded in each pixel of SAR image)
- Transform raw eye movement data in screen space to eye movement data in geospatial coordinates

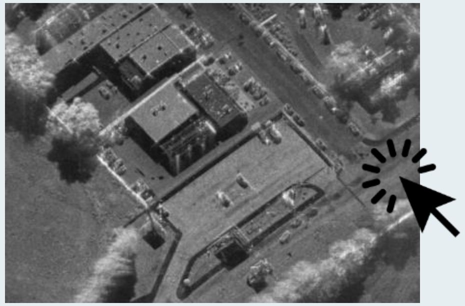
Dimensionality Reduction:

- Move from raw eye movement data to characterized gaze patterns (i.e., fixations)
- Move from fixations in geospatial coordinates to gaze-contingent activated superpixels

Log files: Integration and dimensionality reduction

Two additional data streams of interest: **mouse (panning behavior)** and keyboard* ("ground truth")

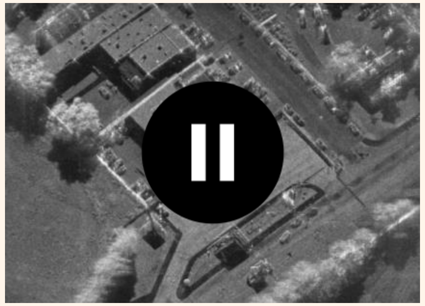
Raw data:



Coordinates of upper left corner of screen as participant pans (click + drag) to explore zoomed image

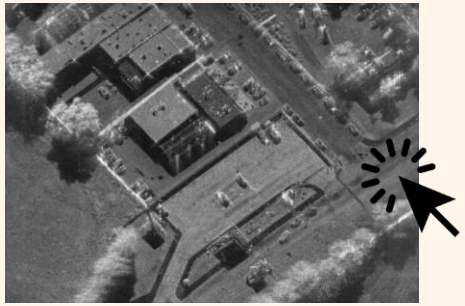
Discretize into runs & pauses:

Pause (not moving):



- Short (<1500 ms)
- Long (>1500 ms)

Run (moving):



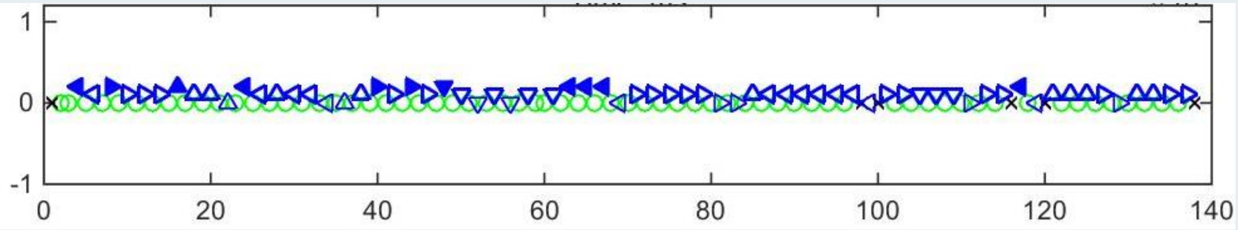
- Magnitude: Total distance (3)
- Angle: angle from start to end (4)
- Length ratio: Linearity or "curviness" (3)*

…and now we have **two critical data streams reduced in dimensionality** to a manageable space!

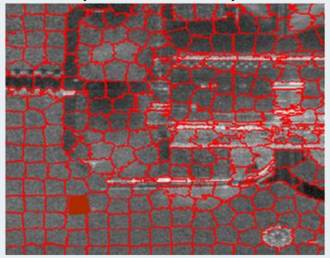
Merge data streams into single sequence

Discretized panning & gaze data:

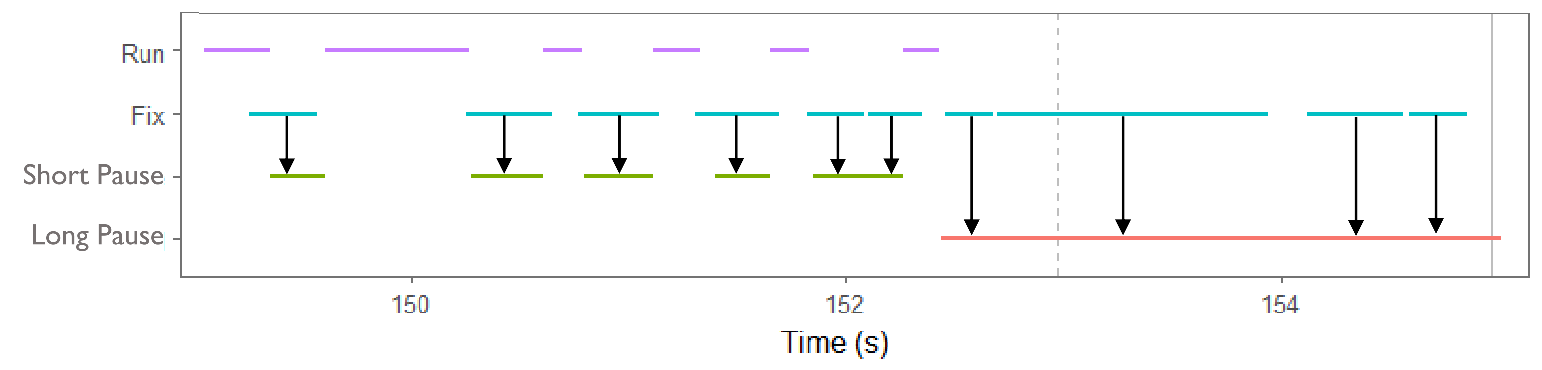
Runs (magnitude, angle, length ratio), short pauses, long pauses)



Activated superpixels (fixations)



Temporal alignment of panning & gaze data:



Merge into single sequence of tokens (0-255), replacing pauses with fixations:

67 - 7 - 53 - 136 - 1 - 42 - 1 - 188 - 7 - 78 - 203 - 91 - 116 - 1 - 26 - 8

Step 3: Analyze sequences

… we now have two critical data streams **reduced in dimensionality** and **integrated into a single sequence** of events!

- Treat sequences as a **vocabulary** of analyst interactions (e.g., text analytics, n-grams)
- Identify decision-making patterns and cues to interactions of interest (e.g., found a target or will find a target soon)
  - Markers of "good" or "bad" performance
  - Characterize individual analyst patterns of behavior
  - Characterize behavior associated with a particular task, regardless of analyst

Now what?

Refine, expand, and validate/verify our process

- End-to-end thinline prototype
- Narrow in on best features to include (e.g., better ways to discretize the data streams?)
- Move all the way to *meaningful*, *semantic* content
    - We've made it to superpixels, but what about "that's a car"?
    - Algorithms developed at Sandia can help with this
- Formally pull in ground truth data
- Validate and verify on new data set
    - Initial analyses focused on Task 1 (still have all of Tasks 2-4)
    - Collect additional data (same tasks, new tasks, expert population, beyond SAR, etc.)

Questions?

Kristin M. Divis | Sandia National Labs – New Mexico | kmdivis@sandia.gov | 505.284.6561