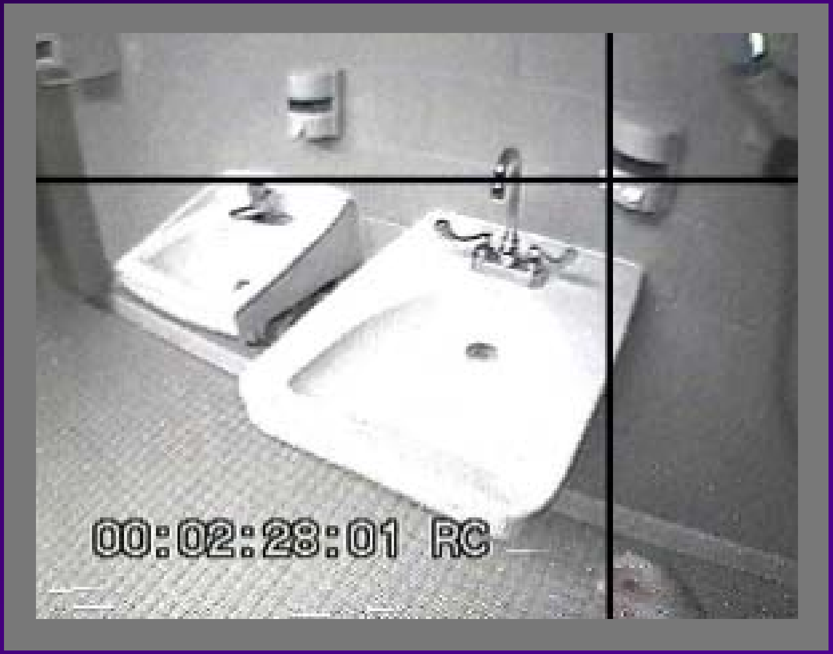
Visual Perception in Familiar, Complex Tasks



Jeff B. Pelz, Roxanne Canosa, Jason Babcock, & Eric Knappenberger

Visual Perception Laboratory
Carlson Center for Imaging Science
Rochester Institute of Technology

Students & other collaborators

Students:

Roxanne Canosa	(Ph.D. Imaging Science)
Jason Babcock	(MS Color Science)
Eric Knappenberger	(MS Imaging Science)

Collaborators:

Mary Hayhoe	UR Cognitive Science
Dana Ballard	UR Computer Science

Visual Perception in Familiar, Complex Tasks

Goals:

- To better understand:
  Visual Perception and Cognition

- To inform design of:
  AI and Computer Vision Systems

Strategic Vision

A better understanding of the strategies and attentional mechanisms underlying visual perception and cognition in human observers can inform us on valuable approaches to computer-based perception.

Motivation: Cognitive Science

> The mechanisms underlying visual perception in humans are common to those needed to implement successful artificial vision systems.

Visual Perception and Cognition

- Sensorial Experience
  - High-level Visual Perception
    - Attentional Mechanisms
      - Eye Movements

Motivation: Computer Science

The mechanisms underlying visual perception in humans are common to those needed to implement successful artificial vision systems.

Artificial Intelligence

- Computer Vision
  - Active Vision
    - Attentional Mechanisms
      - Eye Movements

Motivation

As computing power and system sophistication grow, basic computer vision in constrained environments has become more tractable. But the value of computer vision systems will be shown in their ability to perform higher level actions in complex settings.

Limited Computational Resources

Even in the face of Moore's Law, computers will not have sufficient power in the foreseeable future to solve the "vision problem" of *image understanding* by brute force.

Challenges

Computer-based perception faces the same fundamental challenge that human perception did during evolution:

limited computational resources

Inspiration - “Active Vision”

Active vision was the first step. Unlike traditional approaches to computer vision, *active vision* systems focused on extracting information from the scene rather than brute-force processing of static, 2D images.

Inspiration - “Active Vision”

Visual routines were an important component of the approach. These pre-defined routines are scheduled and run to extract information when and where it is needed.

Goal - "Strategic Vision"

"Strategic Vision" will focus on *high-level, top-down strategies* for extracting information from complex environments.

Goal - “Strategic Vision”

A goal of our research is to study human behavior in natural, complex tasks because we are unlikely to identify these strategies in typical laboratory tasks that are usually used to study vision.

Limited Computational Resources

In humans, “computational resources” means limited *neural resources*: even if the entire cortex were devoted to vision, there are not enough neurons to process and represent the full visual field at high acuity.

Limited Neural Resources

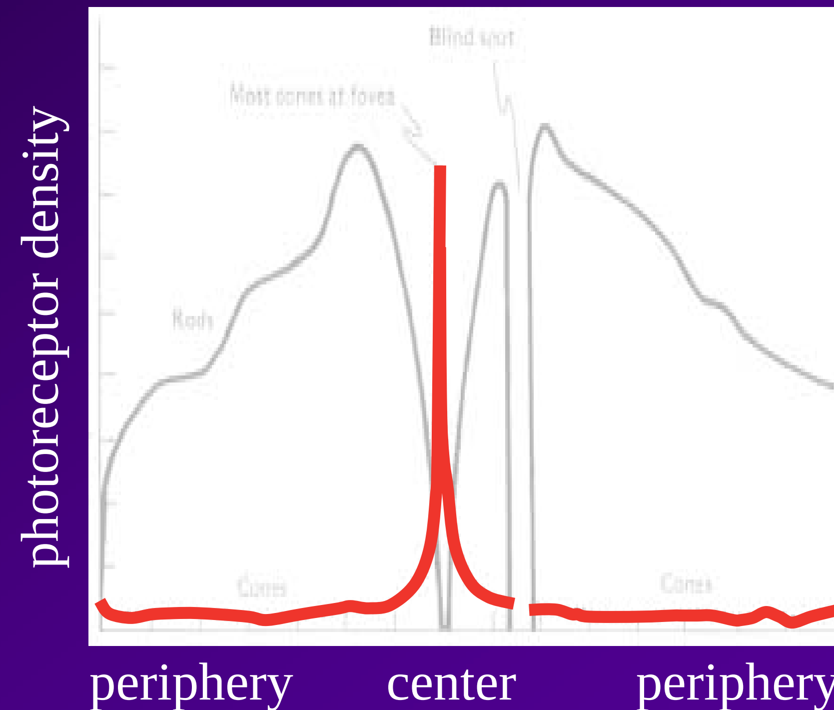
The solutions favored by evolution confronted the problem in three ways;

1. Anisotropic sampling of the scene
2. Serial execution
3. Limited internal representations

1. Anisotropic sampling of the scene

Retinal design:

The "foveal compromise" design employs very high spatial acuity in a small central region (the fovea) coupled with a large field-of-view surround with limited acuity.



Demonstration of the "Foveal Compromise"

Stare at the "+" below.

Without moving your eyes, read the text presented in the next slide:

+

Anisotropic Sampling: Foveal Compromise

\+ If you can read this you must be cheating

Anisotropic Sampling: Foveal Compromise

Despite the conscious percept of a large, high-acuity field-of-view, only a small fraction of the field is represented with sufficient fidelity for tasks requiring even moderate acuity.

Limited Neural Resources

The solutions favored by evolution confronted the problem in three ways;

1. Anisotropic sampling of the scene
2. Serial execution
3. Limited internal representations

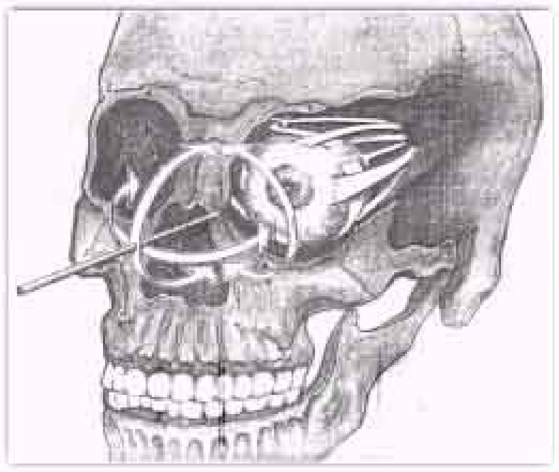
Serial Execution

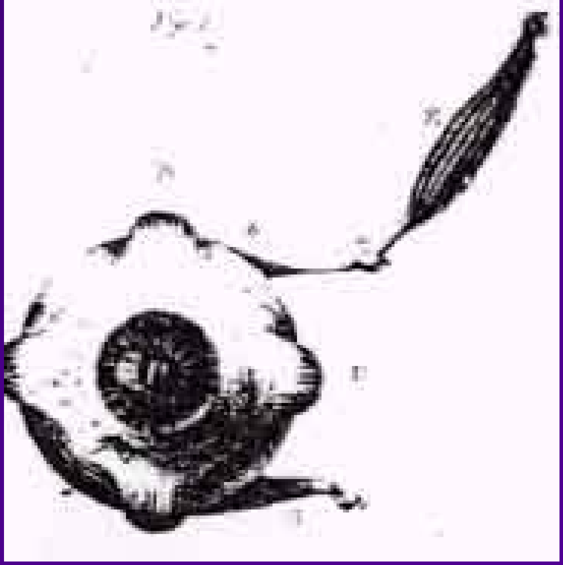
The limited acuity periphery must be sampled by the high-acuity fovea. This sampling imposes a serial flow of information, with successive views.

Serial Execution: Eye Movements

Extraocular Muscles:

Three pairs of extra-ocular muscles provide a rich suite of eye movements that can rapidly move the point of gaze to sample the scene or environment with the high-acuity central fovea.





Background: Eye Movement Types

Smooth pursuit/
Optokinetic response

Vestibular-ocular response

Vergence

} Image stabilization during object and/or observer motion

Saccades

Image *de*stabilization - used to shift gaze to new locations.

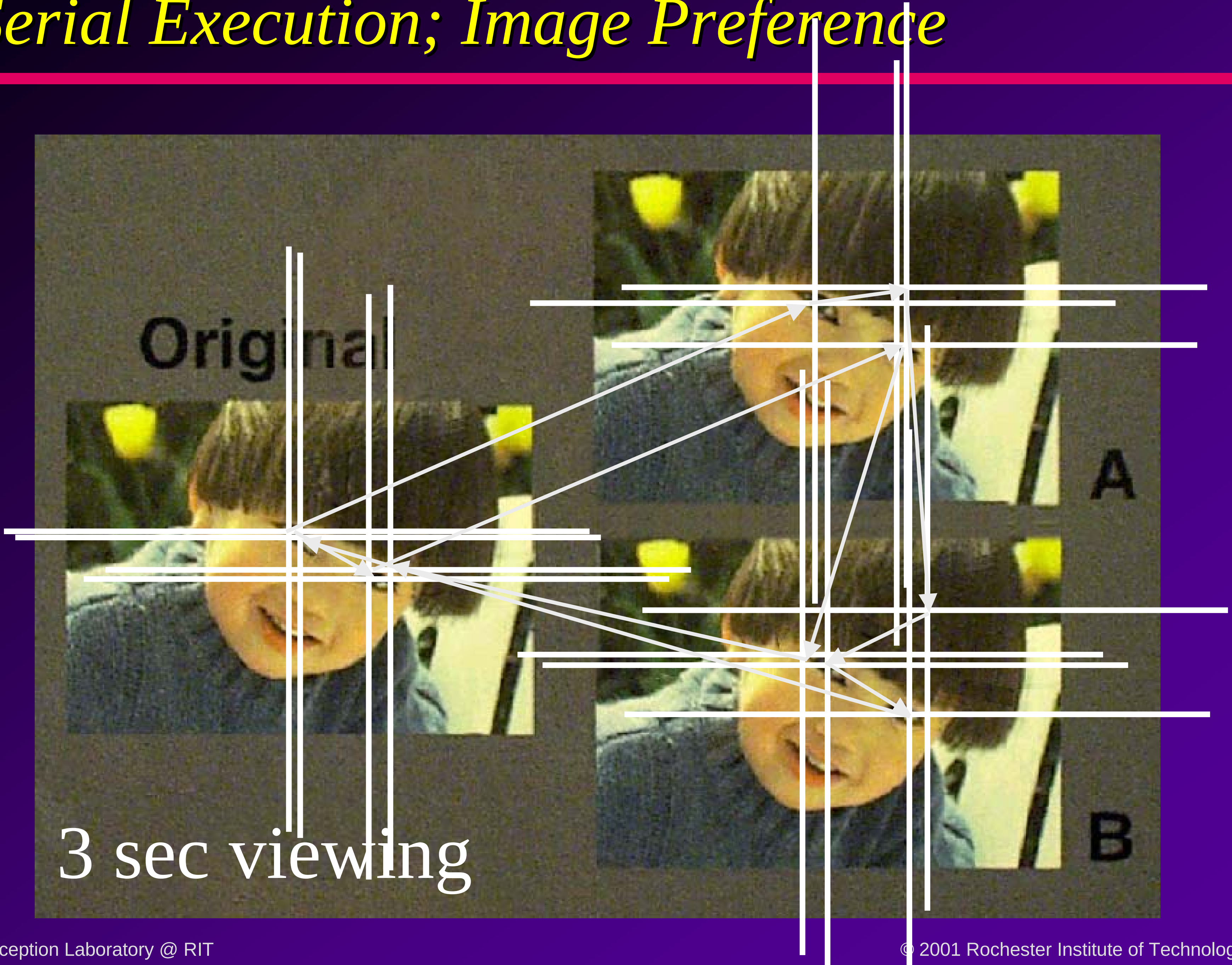
Background: Eye Movement Types

Saccadic Eye Movements:

Rapid, ballistic eye movements that move the eye from point to point in the scene

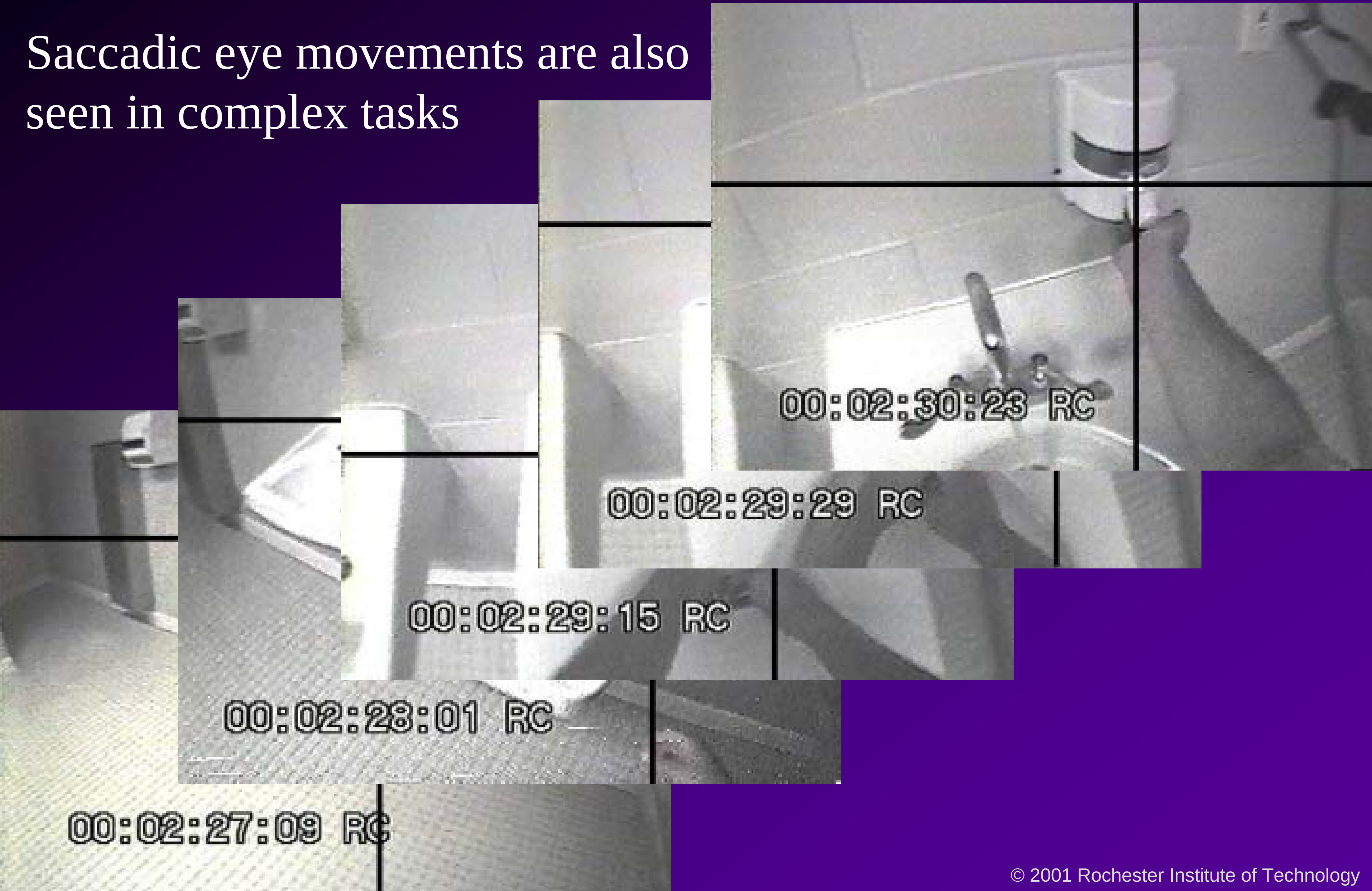
Separated by *fixations* during which the retinal image is stabilized

Serial Execution; Image Preference



Serial Execution: Complex Tasks

Saccadic eye movements are also seen in complex tasks



Limited Neural Resources

The solutions favored by evolution confronted the problem in three ways;

1. Anisotropic sampling of the scene
2. Serial execution
3. Limited internal representations

Integration of Successive Fixations

How are successive views integrated?

What fidelity is the integrated representation?

Integration of Successive Fixations

[fixate the cross]

Integration of Successive Fixations

Integration of Successive Fixations

Perhaps fixations are integrated in an internal representation

Integration of Successive Fixations

Perhaps fixations are integrated in an internal representation

Integration of Successive Fixations

Perhaps fixations are integrated in an internal representation

Integration of Successive Fixations

Perhaps fixations are integrated in an internal representation

Integration of Successive Fixations

Perhaps fixations are integrated in an internal representation

Integration of Successive Fixations

Perhaps fixations are integrated in an internal representation

Integration of Successive Fixations

Perhaps fixations are integrated in an internal representation

Limited Representation: “Change blindness”

If successive fixations are used to build up a high-fidelity internal representation, then it should be easy to detect even small differences between two images.

Try to identify the difference between *Image* **A** *& B*

Image A

Try to identify the difference between *Image A & B*

Try to identify the difference between *Image A & **B***

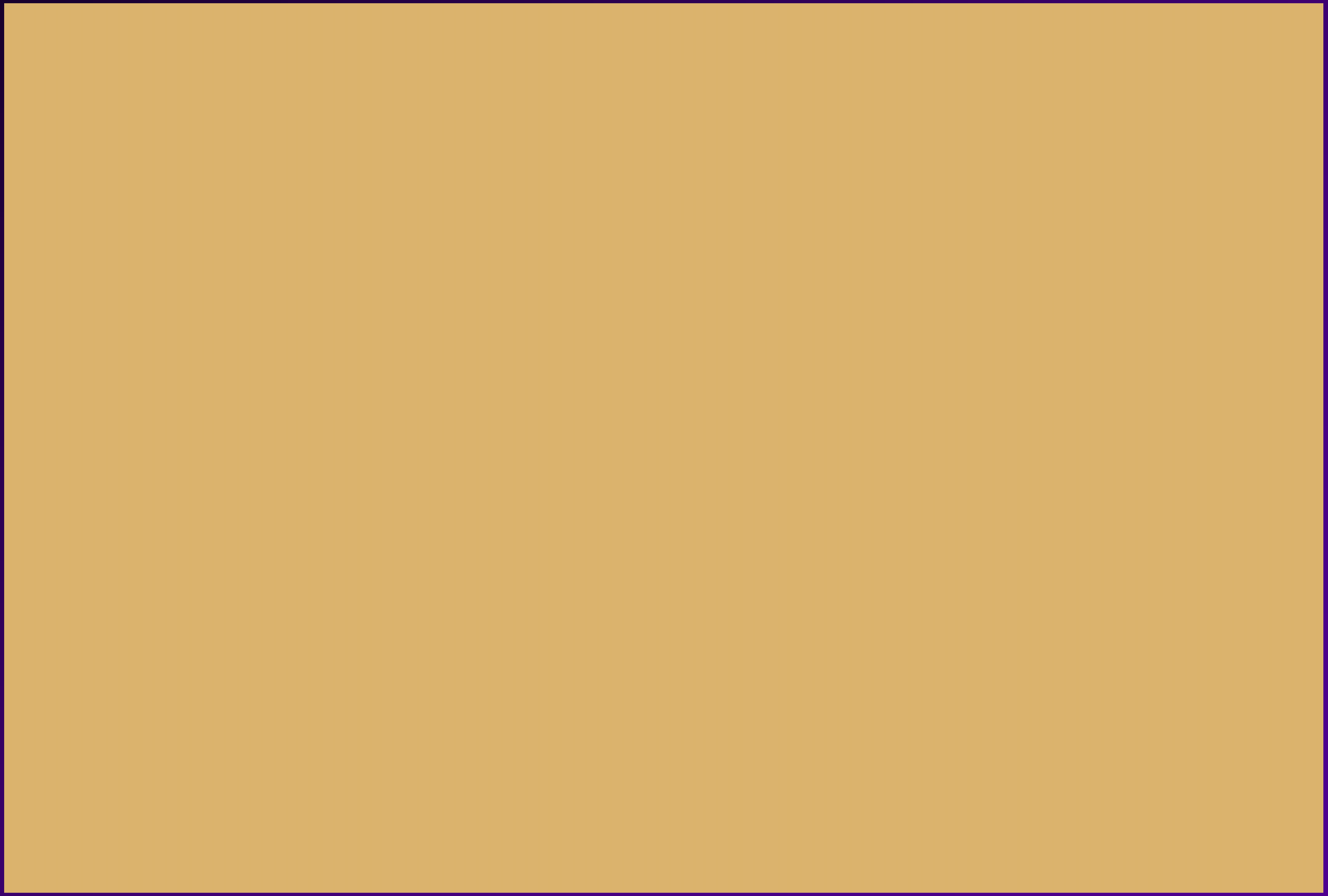
Image B

Try to identify the difference between *Image A & B*

Try to identify the difference between *Image* **A** *& B*

Image A

Try to identify the difference between *Image A & B*



Try to identify the difference between *Image A* & ***B***

Image B

Try to identify the difference between *Image* A *& B*

Image A

Try to identify the difference between *Image A & **B***

Image B

The question:

Can we identify oculomotor strategies that observers use to ease the computational and memory load on observers as they perceive the real world?

The question:

To answer that question, we have to monitor eye movements in the real world, as people perform real extended tasks.

One problem is the hardware:

Measuring eye movements

Many eyetracking systems require that head movements (and other natural behaviors) be restricted.

Scleral eye-coils

Dual Purkinje eyetracker

RIT Wearable Eyetracker

The RIT wearable eyetracker is a self-contained unit that allows monitoring subjects' eye movements during natural tasks.

RIT Wearable Eyetracker

The headgear holds CMOS cameras and IR source. Controller and video recording devices are held in a backpack.

Perceptual strategies

Beyond the mechanics of how the eyes move during real tasks, we are interested in strategies that may support the conscious perception that is continuous temporally as well as spatially.

Perceptual strategies

When subjects' eye movements are monitored as they perform familiar complex tasks, novel sequences of eye movements were seen that demonstrate strategies that simplify the perceptual load of these tasks.

Perceptual strategies

In laboratory tasks, subjects usually fixate objects of immediate relevance. When we monitor subjects performing complex tasks, we also observe fixations on objects that were relevant only to *future* interactions.

The next slide shows a series of fixations as a subject approached a sink and washed his hands:

“Look-ahead” Fixations

t = -700 msec

t = 0 msec

t = 1500 msec

t = 2000 msec

t = 2600 msec

t = 2800 msec

Perceptual Strategies: Where, when, why?

The "look-ahead" fixations are made in advance of "guiding" fixations used before interacting with objects in the environment.

look-ahead fixation

guiding fixation

interaction

Conclusion

Humans employ strategies to ease the computational and memory loads inherent in complex tasks. *Look-ahead* fixations reveal one such strategy:

opportunistic execution of information-gathering routines to pre-fetch information needed for future subtasks.

Future Work

Implementing this form of opportunistic execution in artificial vision systems can test the hypothesis that strategic visual routines observed in humans can benefit computer-based perceptual systems.