



# From Images to Data: Scaling and Streamlining Research Workflows

Dirk Colbry  
Institute for Cyber-Enabled Research

MICHIGAN STATE  
UNIVERSITY



## Agenda

- Scientific Image Analysis
- Scaling issues
- Exploratory research
- Image Grammar

MICHIGAN STATE  
UNIVERSITY

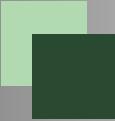




## Research Team

- Dirk Colbry
  - Institute for Cyber Enabled Discovery
- Fred Dyer
  - Zoology
- Ian Dworkin
  - Zoology
- Yang Wang
  - Mathematics
- Lifeng Wang
  - Statistics & Probability



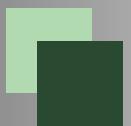


## Phenomics

Area of biology concerned with the measurement of phenomes — the physical and biochemical traits of organisms — as they change in response to genetic mutation and environmental influences.

Houle, D.; Govindaraju, D.R.; Omholt, S. (2010), "Phenomics: the next challenge", Nature Reviews Genetics 11 (12): 855–66

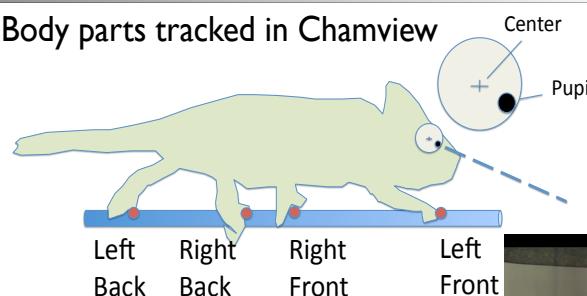
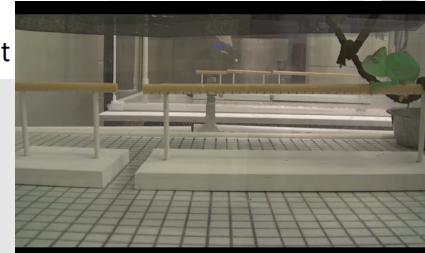


# Animal Behavior

Dr. Fred Dyer, MSU

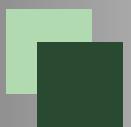
Body parts tracked in Chamview

MICHIGAN STATE UNIVERSITY

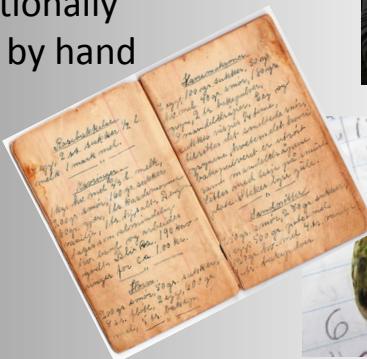
5





# Visual Science

- Long history in Biology
- Traditionally done by hand






MICHIGAN STATE UNIVERSITY



## Photography

- Changing science
  - Scientists are able to record video without knowing what they will see
  - Cameras may see something the scientists missed
  - Different scientists can view the same data with different scientific questions in mind







MICHIGAN STATE  
UNIVERSITY



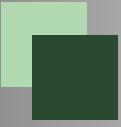
<http://skyview.gsfc.nasa.gov/>

- Do you really need a telescope to ask your research question?
- Query data saved from other experiments to save money and time.
- Enables researchers to ask larger questions



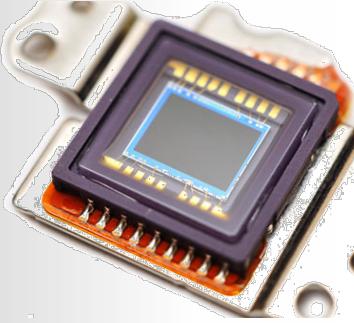


MICHIGAN STATE  
UNIVERSITY



## Cameras Everywhere

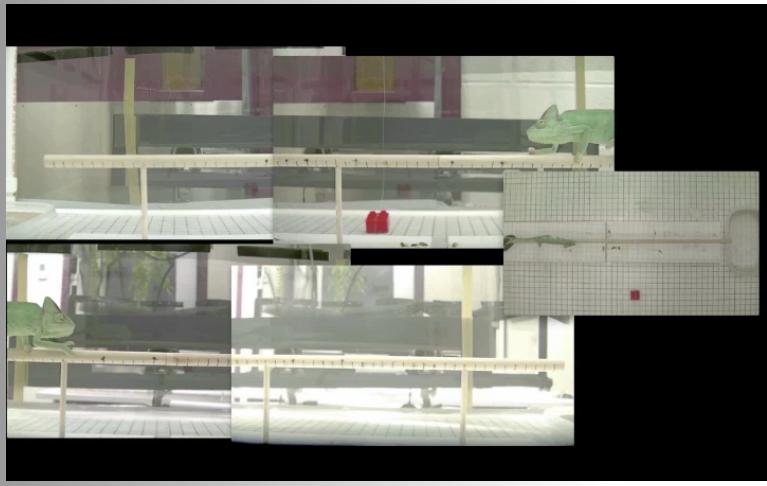
- Transforming how scientists gather data
- Very affordable
- Data is becoming very cheap to gather, so there is a lot more of it

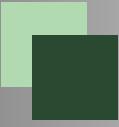


Charge-Coupled Device (CCD)



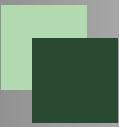
## Even small projects grow fast

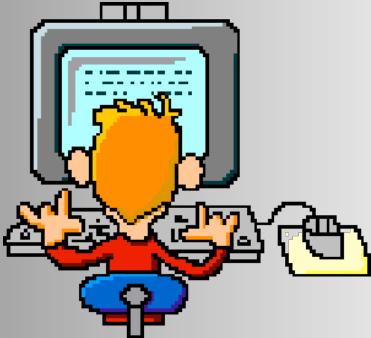


 Problems with Video/Image Data 

- Data is growing at a fast rate:
  - 11 years of YouTube video are uploaded every day ([http://www.youtube.com/t/press\\_statistics](http://www.youtube.com/t/press_statistics))
- Information is specific to the research problem.
- The noise to information ratio is very high in most image/video projects.
  - For example, in a single frame we may only need one measurement.

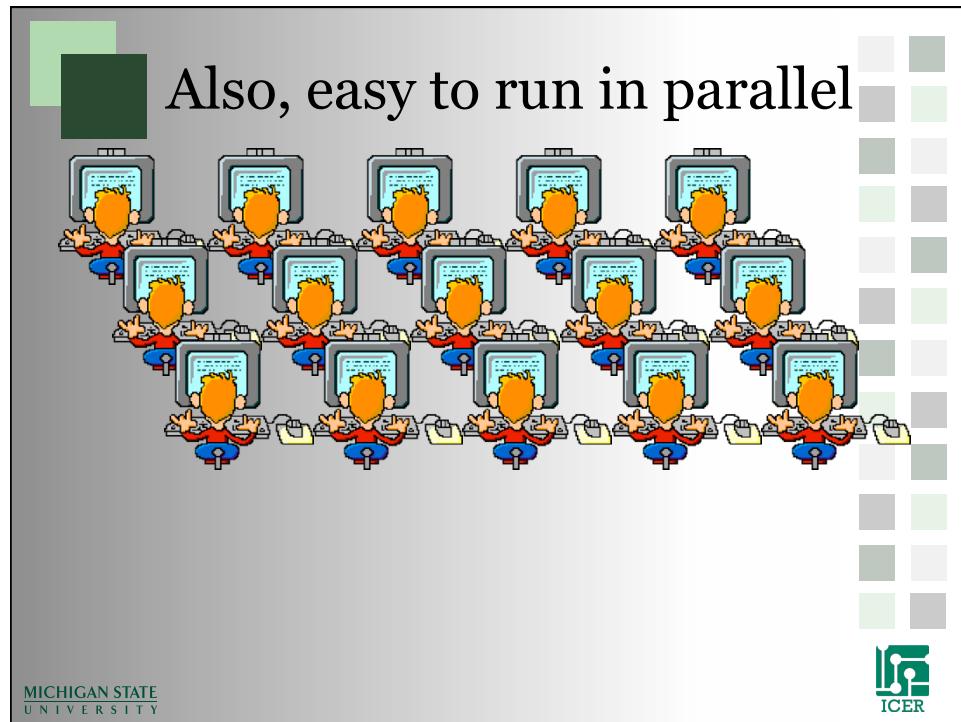
 

 How are digital images analyzed? 



Graduate students are cheap...  
Undergraduates are even cheaper!



ZOONIVERSE

MOON ZOO

GALAXY ZOO

Crater Survey

Play  
About  
Community  
Craters  
Boulder Walls  
My Moon Zoo  
Live!  
Explore Moon  
Glossary  
Tutorials  
FAQ  
Forum  
Blog  
Moonometer  
Log Out

Examples

HIDE ALL  
HIDE CURRENT  
SUBMIT

Illumination Crater Mound

MICHIGAN STATE UNIVERSITY

ICER

**Massively Parallel Image analysis**

**ZOONIVERSE** REAL SCIENCE ONLINE  
<https://www.zooniverse.org/>

**Climate**  
Model Earth's using wartime  
Help scientists make worldwide weather observations made by Navy ships.  
oldWeather

**Nature**  
Hear Whales communicate  
You can help researchers understand what whales are saying.

**Humanities**  
Study the lives of ancient Greeks  
The data gathered from Lives helps scholars learn about Oxyrhynchus.

**Space**  
How do galaxies form?  
NASA's Hubble Space Telescope archive provides hundreds of thousands of galaxy images.

**GALAXY ZOO**

**MICHIGAN STATE UNIVERSITY**

**ICER**

## Goal: Automate Image Analysis

- Let computers do the work for you
- Automatically identify information in images
- Return that information as data to the scientist
  - Pattern Recognition
  - Machine Vision
  - Image Processing

**MICHIGAN STATE UNIVERSITY**

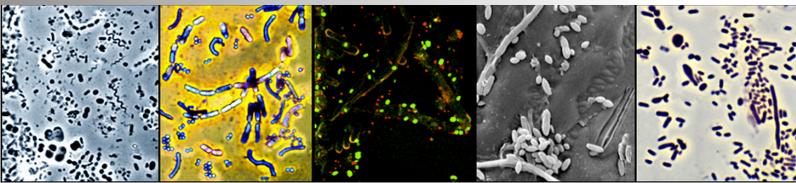
**ICER**

**Existing automated approaches**

<p><b>Task Specific</b></p> <ul style="list-style-type: none"> <li>• Program solves a specific problem for a specific type of input</li> <li>• Domain specific assumptions make it easy to automate image analysis</li> </ul> <p>Ex. CMEIAS, wingmachine</p>	<p><b>General</b></p> <ul style="list-style-type: none"> <li>• Tools make it easy to do global manipulation of images</li> <li>• Difficult to do anything specific to a problem</li> </ul> <p>Ex. Photoshop, ImageJ</p>
--	---




**CMEIAS – Center for Microbial Ecology Image Analysis System**  
Dr. Frank Dazzo, MSU

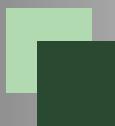


Comprehensive suite of bioimage informatics analysis software applications designed to strengthen quantitative, microscopy-based approaches for understanding microbial ecology, at spatial scales relevant to the individual microbes and their ecological niches.

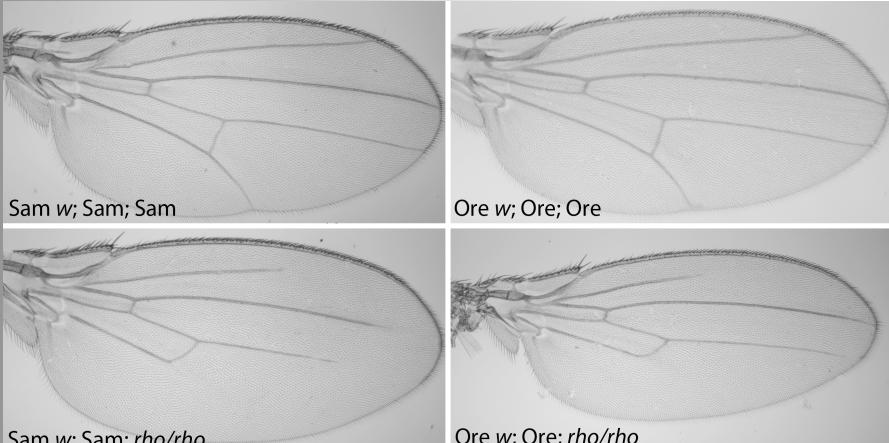


<http://cme.msu.edu/cmeias/>



 Wing Images

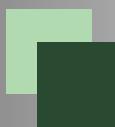
Dr. Ian Dworkin, MSU

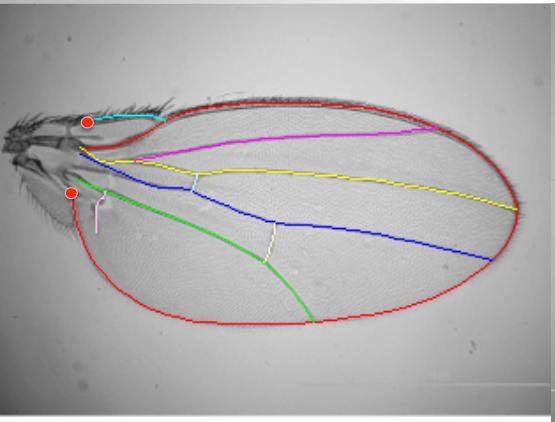


Sam w; Sam; Sam      Ore w; Ore; Ore  
 Sam w; Sam; rho/rho      Ore w; Ore; rho/rho





 Wingmachine



Wingmachine software developed by the Houle lab, FSU  
*Houle et al. 2003 BMC Evo. Biol. 3:25*








- Free, Open source
- Over 1,700 users
- More than 500 Plugins listed on their website
- Many, many features

MICHIGAN STATE UNIVERSITY




## Why is this hard?

- In order to automatically analyze images you need a new tool for every problem
  - Every domain is different
  - There is no one tool that will solve all problems
  - General tools (like ImageJ) need to be assembled to make automated tools
  - It takes time to develop domain specific tools

MICHIGAN STATE UNIVERSITY





## Our Goal

- Focus on the workflow of making tools, not on the individual tools themselves
- Make it easier to make new tools
- How?



## Idea: Make an Image Grammar

- Can we make a language that helps us build tools?
- What would this image grammar look like?
- How would it work?
- Would such a grammar make exploratory science easier?



## Syntax of images

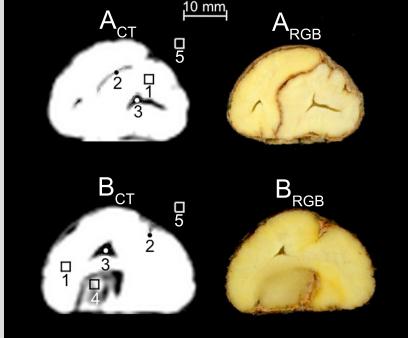
- How images are made and used
  - Pixels
  - Relationships between pixels
  - Temporal
  - Spatial
  - Etc.
- General algorithms that are available in packages like ImageJ
  - Image encasement
  - Cropping
  - Segmenting
  - Etc.

MICHIGAN STATE UNIVERSITY



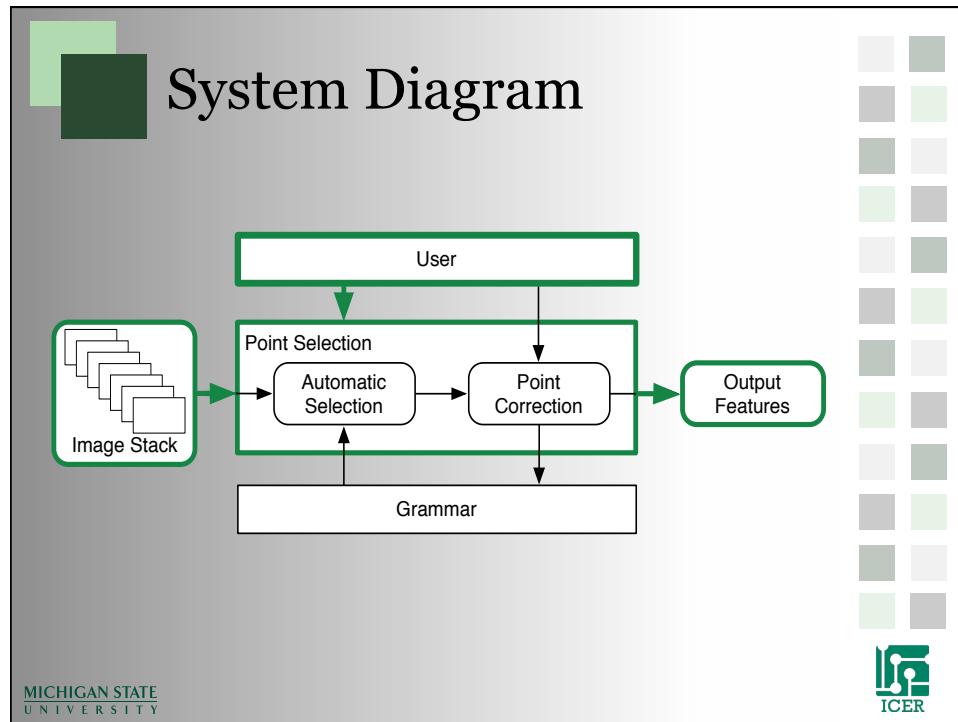
## Semantics

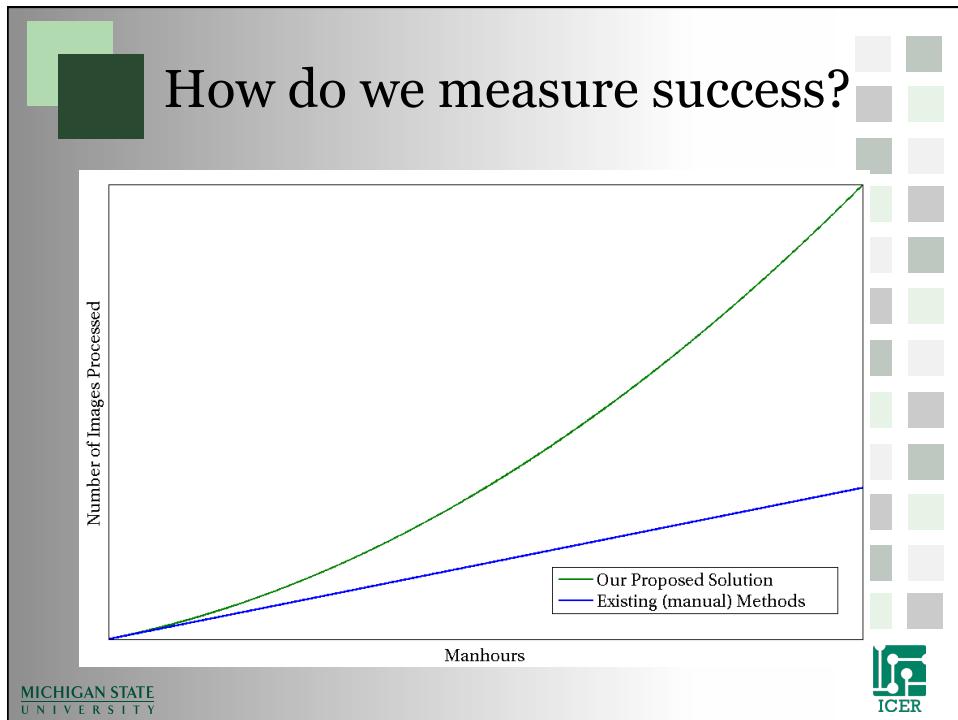
- What is it we are measuring and looking for?
  - Wings
  - Brains
  - Chameleons
  - Arm
  - Leg
  - Etc.



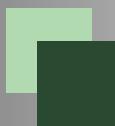
MICHIGAN STATE UNIVERSITY





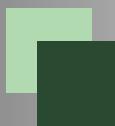


- ## Prototype System Objects
- **Preprocessors:** Enhance or segment images. Both the input and the output of a preprocessor is an image.
  - **ImageStack:** This is the main object that stores information about sets of images.
  - **Predictor:** Picks points within an image.
  - **Choosers:** Picks the best points from a list of point predictions.

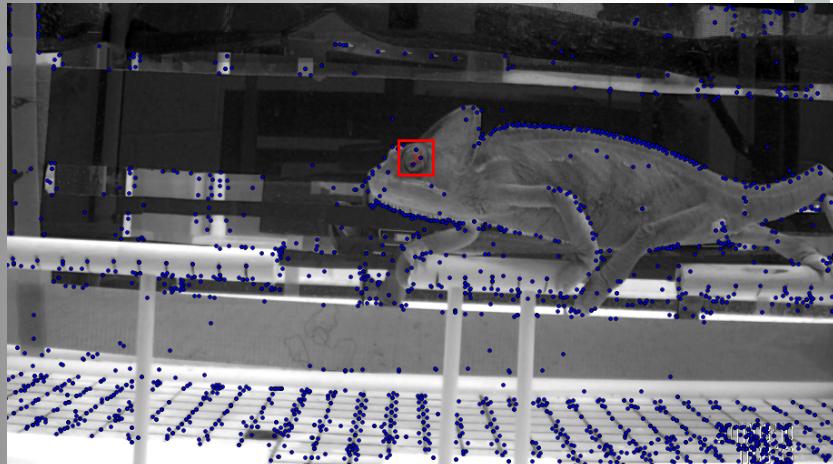


## Example

- Preliminary Point Pickers:
  - Kinetic
  - Template Matching
  - SIFT Features

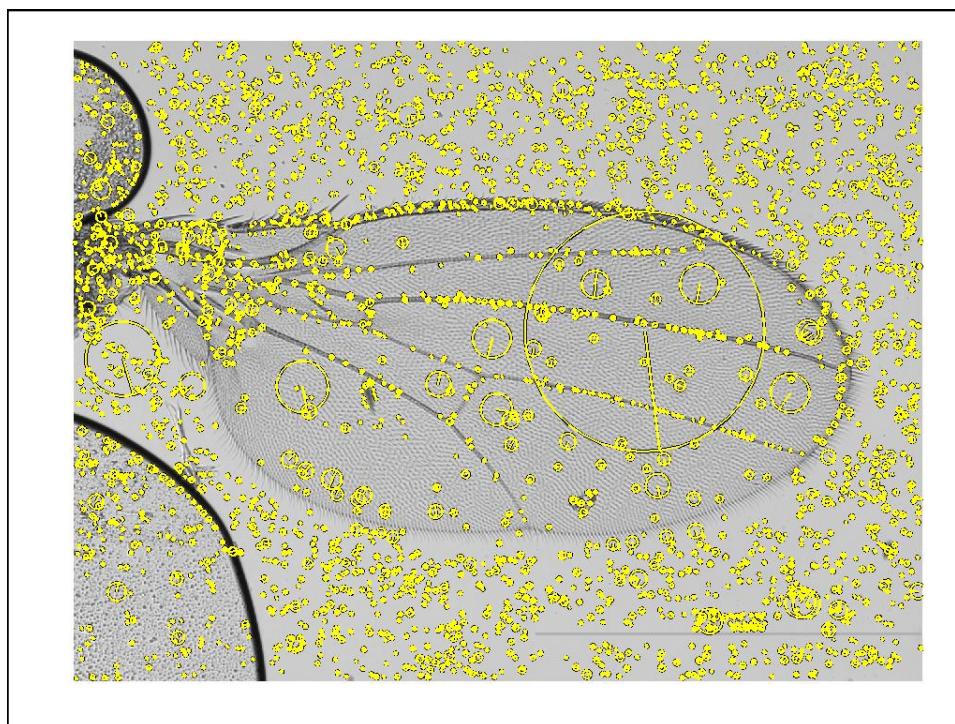
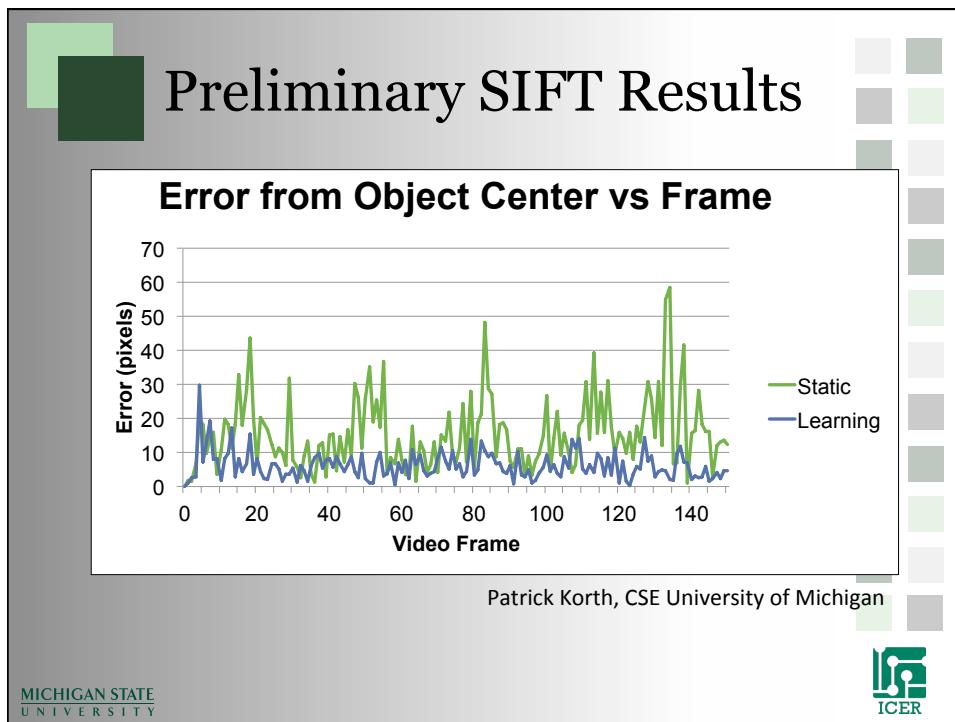


## Point tracking on Chameleons



Patrick Korth, CSE University of Michigan





## Future Challenges

- More pickers
- More Projects
- More than just points
- Automated search though the grammar space
- Automated tweaking of algorithm parameters
- Bigger datasets (Run on the HPCC)
- Intelligent feedback to the researchers

MICHIGAN STATE  
UNIVERSITY

