

---

# Advanced Topics in Scalable Learning

CSE 6392 Lecture 1 Administration &  
Introduction

Junzhou Huang, Ph.D.

Department of Computer Science and Engineering



# Administration

---

- **Course CSE6392**

- What: Advanced Topics in Scalable Learning
- When: Friday 1:00 ~ 3:50pm
- Where: NH 111
- Who: Junzhou Huang (Office ERB 650) [jzhuang@uta.edu](mailto:jzhuang@uta.edu)
- Office Hour: FRIDAY 3:50 ~ 6:00pm and/or appointments
- Webpage: <http://ranger.uta.edu/~huang/teaching/CSE6392.htm>  
(Please check this page regularly)

- **Lecturer**

- PhD in CS from Rutgers, the State University of New Jersey
- Research areas: machine learning , computer vision, medical image analysis and bioinformatics

- **GTA**

- Saiyang Na (Office ERB 105B), [sxn3892@mavs.uta.edu](mailto:sxn3892@mavs.uta.edu)
- Office hours: Friday 10:00am ~ 12:00pm and/or appointments



# Study materials

---

- **Prerequisites**

- Algorithm and Data Structures (CSE2320)
- Introduction to Computers & Programming (CSE 1310 )
- What this really means:
  - You know at least one programming language.
  - Elementary knowledge of Linear Algebra
  - Elementary knowledge of Data Structure
  - Elementary knowledge of Algorithms

- **More (Not Necessary)**

- Machine learning
- Computer Vision
- Data Mining
- Image Processing



# Study Materials

---

- **Text book**
  - None Necessary
  - Cutting-Edge Topics
- **Related Textbooks**
  - "[The elements of statistical learning : data mining, inference, and prediction, 2nd Edition](#)", by Hastie, T., Tibshirani, R., Friedman J. Springer, 2009
  - "[Computer Vision -- A modern approach](#)", by David Forsyth and Jean Ponce, Prentice Hall, 2002
  - “Deep Learning”, by Ian Goodfellow, Yoshua Bengio, Aaron Courville, MIT Press, 2016
- **Acknowledgments**
  - Material from textbook sites or other class sites
  - Lots of material available on the web (via google search, wikipedia)
  - Papers from proceedings of ICML, NeurIPS, ICLR, CVPR, ICCV, KDD



# Assignments

---

- **Paper Selection**

- Each group has two members at most.
- Each group will select at least one paper from the following paper list and then be scheduled to present their selected papers in our class.
- You can choose any papers from the paper lists in the class
- Please talk to the lecturer if you prefer to select a paper out of the list
- The selected paper has to be confirmed by the second week
- GTA will set up the paper selection sheet
- Different groups will present different papers

- **Start early !!!**



# Grading

---

- **Distribution**

- 30% Paper Presentation
  - 30% Slide Preparation
  - 30% Questions & Answering
  - 10% Class Participation
- 
- 100%

- **Attention**

- No midterm or final exam for this course.
- Please read the selected paper and prepare the final presentation as early as possible
- This is research seminar course. Asking questions and discussion are highly encouraged



# Information

---

- **Course Webpage**
  - Check the web page regularly
  - Announcements and lecture notes will be posted there.
- **Grade Appeal**
  - Please refer to the UTA Catalog for the detailed guide of grade appeals.
- **Drop Policy**
  - The university withdrawal policy will be strictly adhered to.
- **Others**
  - Accommodating students with disabilities
  - Student Support Services
  - Etc.



# Questions

---





# Course Overview

---

- **What is it?**
  - Advanced Topics in Scalable Learning
  - Specifically, *Deep Graph Learning, Large Language Models*
- **Why is a CSE course?!?**
  - Hardware & Software
- **Will I really ever use this stuff again?**
  - Important knowledge for a CSE student
  - You may not become a professional guys in this field but you need know what it is, which will help you to follow this rapidly changing world.
  - GOOGLE, FACEBOOK, YOUTUBE, MICROSOFT, KINECT
- **How to succeed in this course?**
  - Attend the class and follow the slides
  - Read the related studying materials
  - Asking questions



# Why are you in this class?

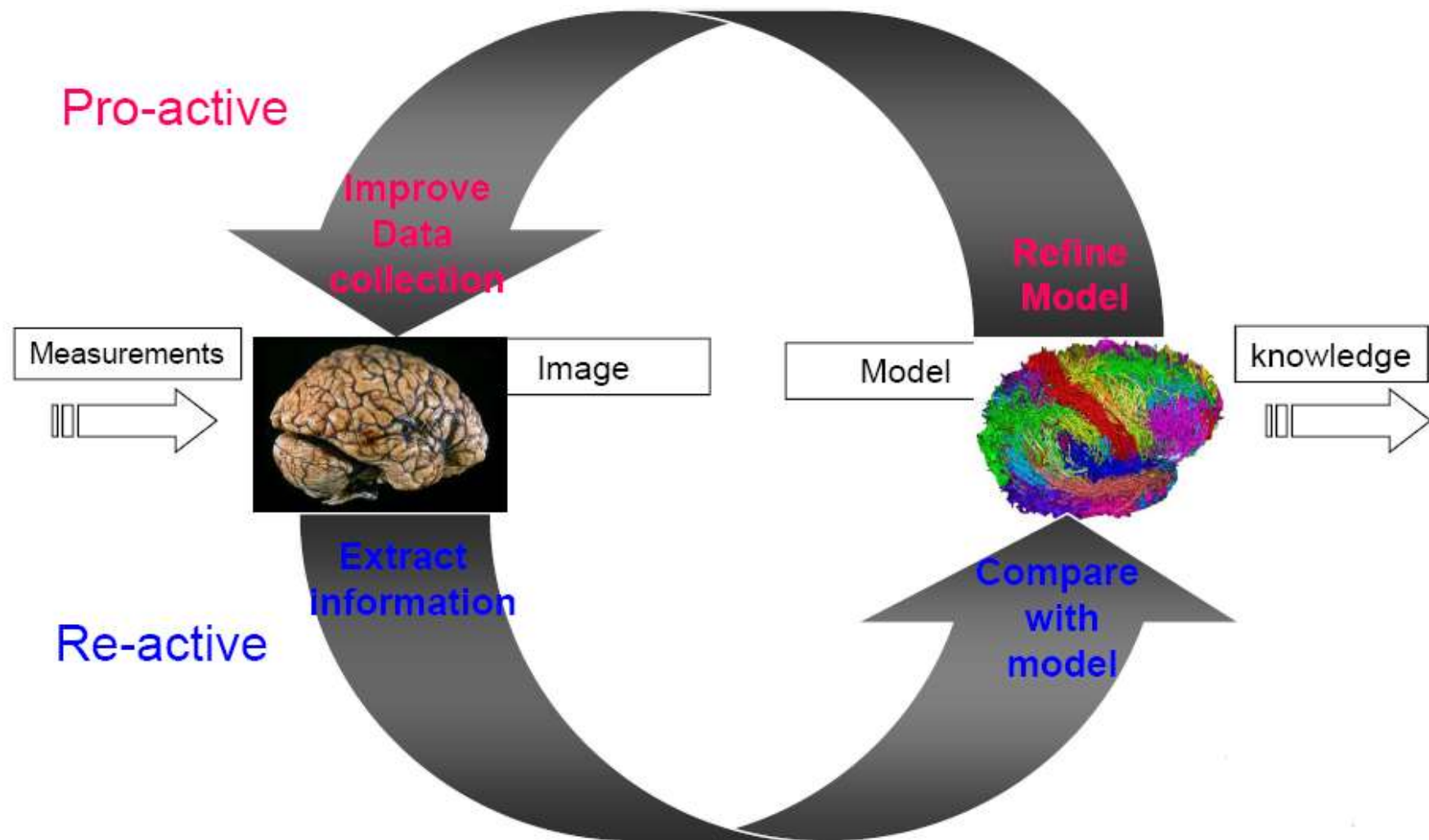
---

- **Something interesting about you**
  - Why you picked your major?
  - Job and Market
- **Learn background of advanced topics**
  - Sparse learning, transfer learning, deep graph learning,
  - CNN, Transformer, Vision Transformer, ...
  - AlphaGo, AlphaFold, ChatGPT...
- **Learn the problem and techniques**
  - Learn where the problems come from
  - Learn what the related techniques to solve them
- **Prepare for understanding recent innovations**
  - Fast optimization for big data analysis
  - Image/Text/Video generation, Survival prediction, Weather forecasting, etc
  - AlphaGo, AlphaFold, ChatGPT



# What is the focus?

Using computational tools to maximize information for knowledge gain



# What is the goals?

---

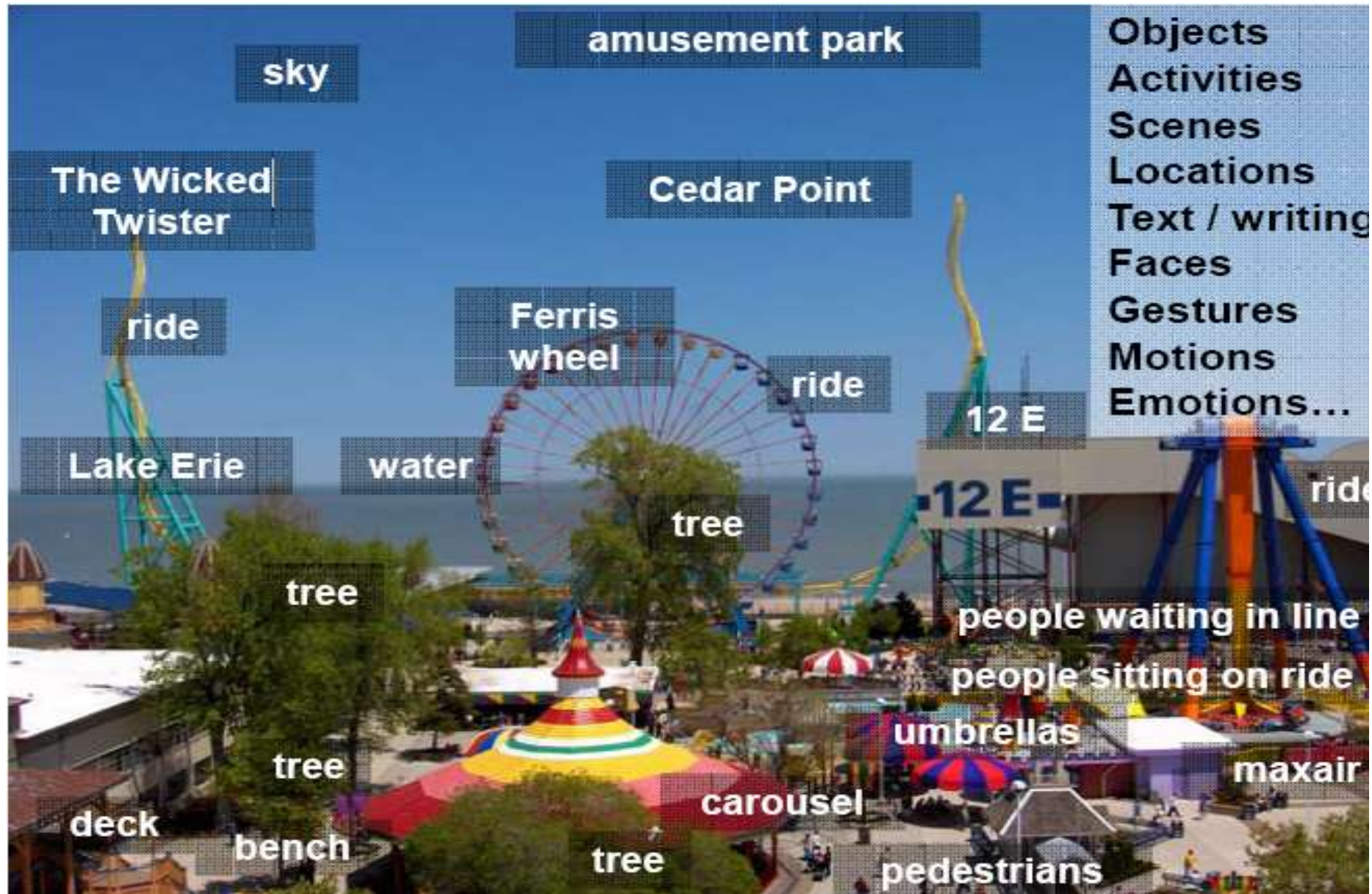
- **Automatic understanding of text, images and video (data)**
  - Computing properties of the 3D world – from visual data (measurement)
  - Algorithms and representations to allow a machine to recognize objects, people, scenes, and activities. (perception and interpretation)
  - Algorithms to mine, search, and interact with visual data (search and organization)



# Measurement

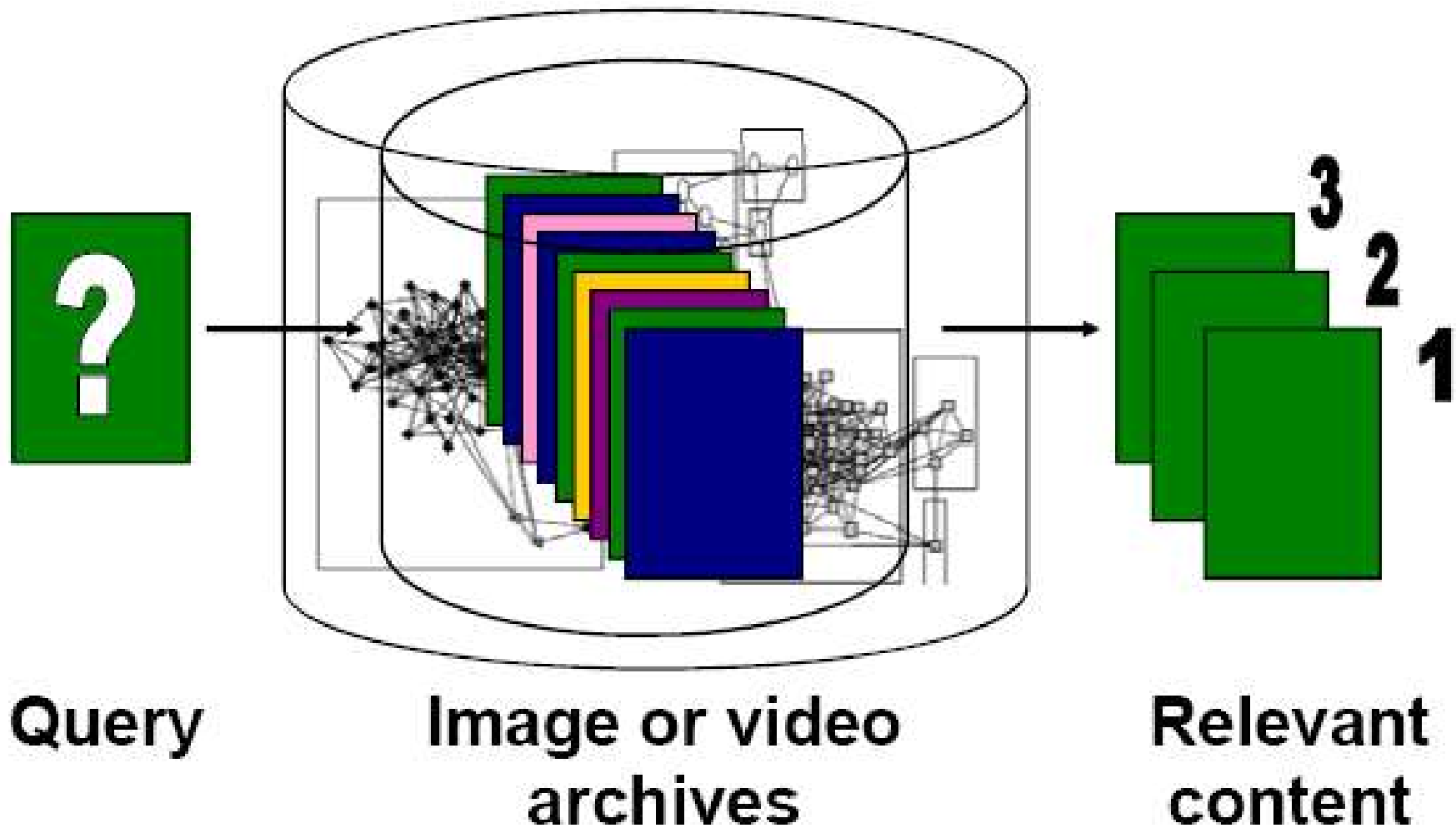


# Perception and Interpretation



# Searching and Recognition

---



# Searching and Recognition

---



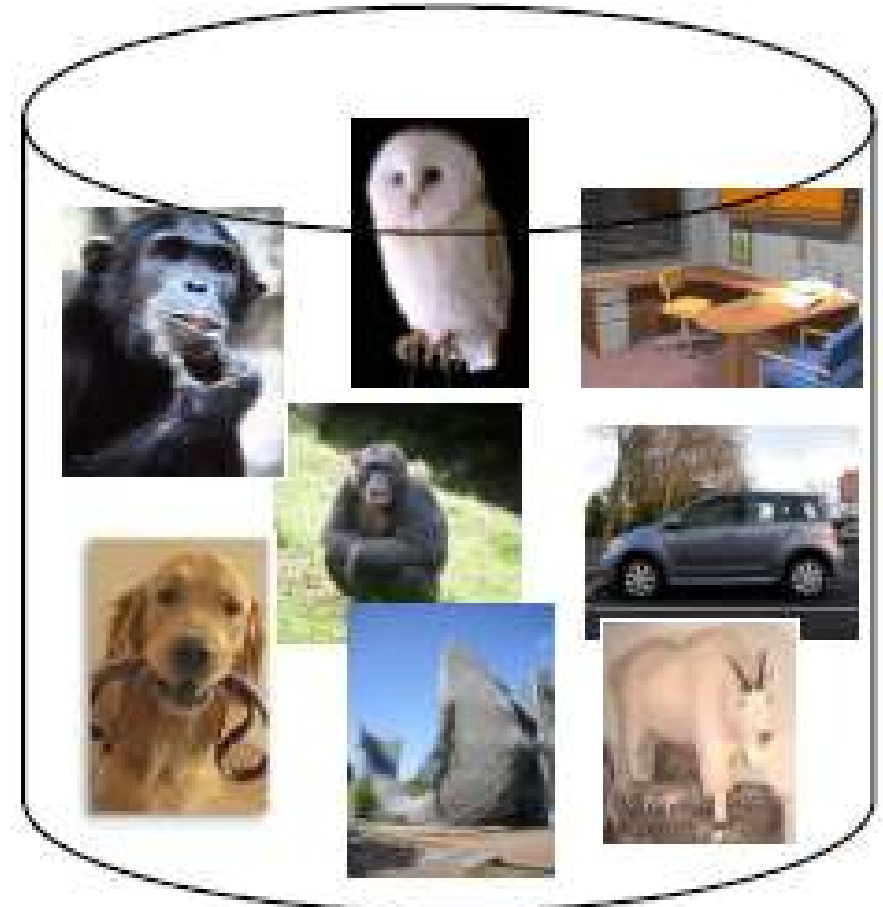
- **A fundamental part of perception**
  - Robots,
  - Autonomous agents,
  - intelligent system
  - For example: Kinect game system
- **Organize and give access to visual content**
  - Connect to information
  - Detect trends and themes
  - Make prediction and avoid risk



# Motivation

---

Object categorization

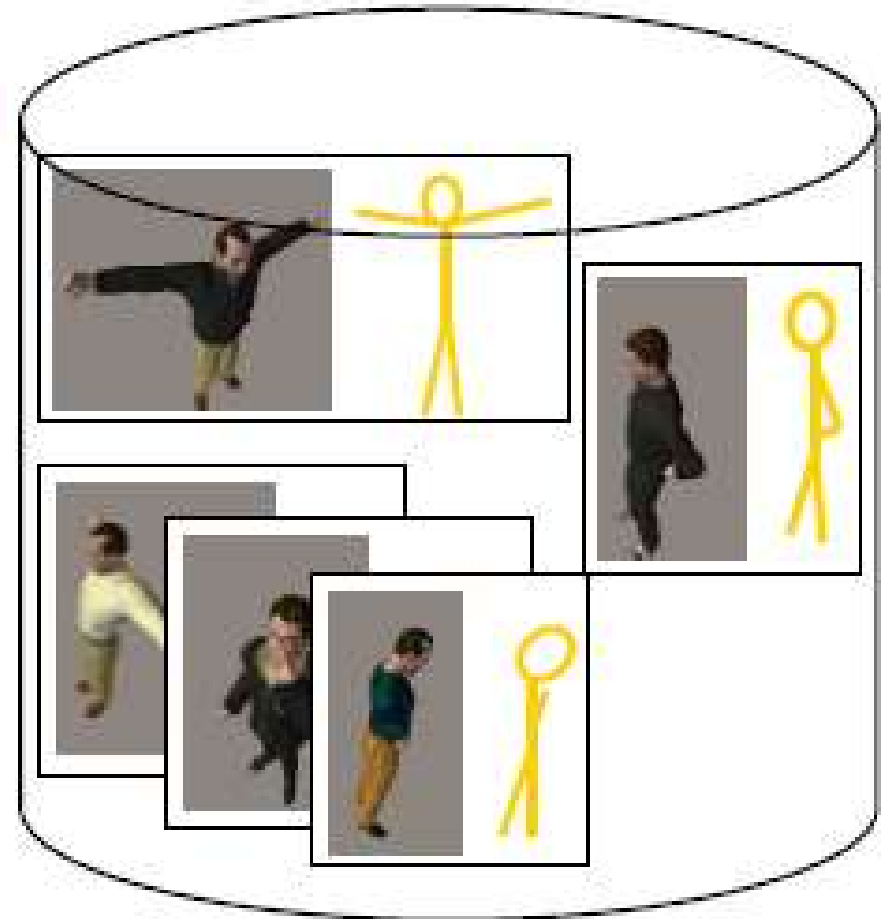
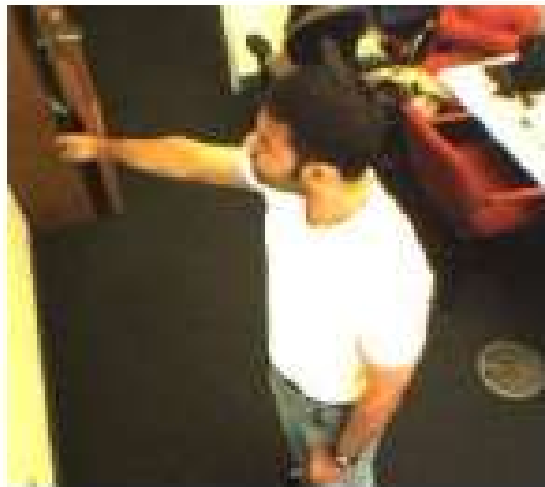


[http://www.cs.utexas.edu/~grauman/slides/jain\\_et\\_al\\_cvpr2008.ppt](http://www.cs.utexas.edu/~grauman/slides/jain_et_al_cvpr2008.ppt)

# Motivation

---

## Example-based pose estimation

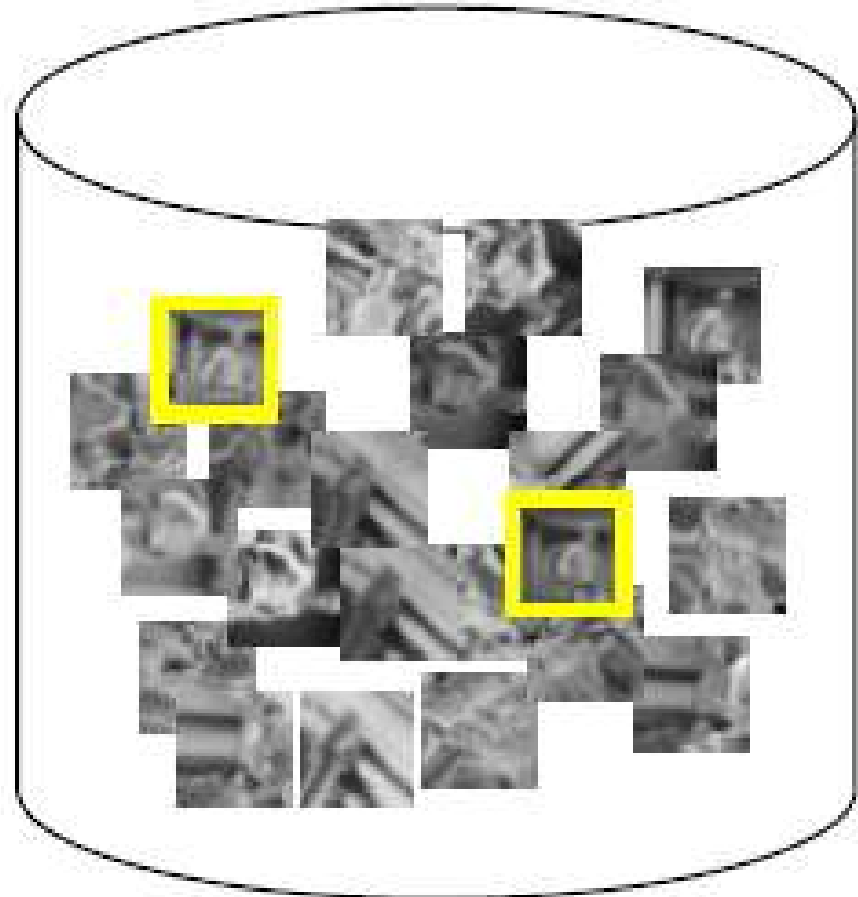
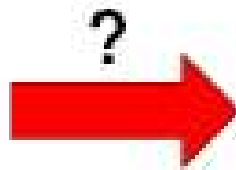


[http://www.cs.utexas.edu/~grauman/slides/jain\\_et\\_al\\_cvpr2008.ppt](http://www.cs.utexas.edu/~grauman/slides/jain_et_al_cvpr2008.ppt)

# Motivation

---

## Structure from Motion



[http://www.cs.utexas.edu/~grauman/slides/jain\\_et\\_al\\_cvpr2008.ppt](http://www.cs.utexas.edu/~grauman/slides/jain_et_al_cvpr2008.ppt)

---

# Scalable Searching via Learning?



# Big Data in the Wild



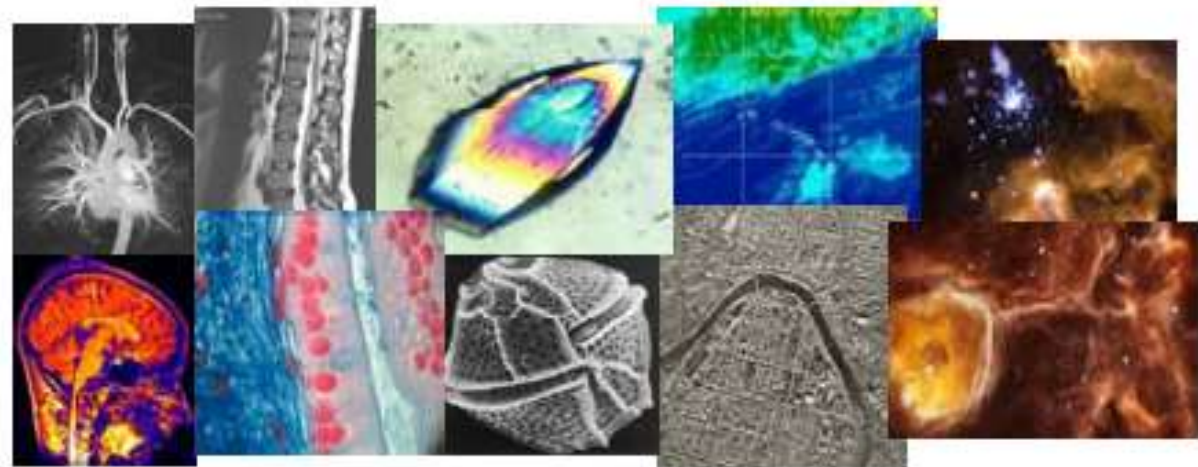
Personal photo albums



Movies, news, sports



Surveillance and security



Medical and scientific images



# Big Data in the Wild



**flickr** 350 mil. photos,  
1 mil. added daily

**Google** 1.6 bil. images indexed  
as of summer 2005



**IMDb** 916,271 titles

**YouTube** 10 mil. videos, 65,000 added daily



# How machine detect objects?

---



<http://www.darpa.mil/grandchallenge/gallery.asp>



# Mobile Computing



Situated search  
Yeh et al., MIT



kooaba



MSR Lincoln





# Searching in Shopping

The screenshot shows the Like.com website interface. At the top, there's a navigation bar with categories like 'ALL SHOES', 'MENS APPAREL', 'WOMENS APPAREL', 'KIDS', 'ACCESSORIES', 'JEWELRY & WATCHES', 'HOLIDAY', and 'TOYS & HOME'. A search bar contains 'Women's Shoes'. Below the search bar are three filter sections: 'Refine by Style' (with icons for Pumps, Sandals, Flats, etc.), 'Refine by Color' (with color swatches for Red, Black, Beige, etc.), and 'Refine by Brand' (with icons for Dolce, Jodi, etc.).

On the left side, there's a section titled 'Why is Like.com Different?' with the text: 'Like is a visual shopping engine that lets you find items by color, shape and pattern. Click on [How to Use Like.com](#) to get started.' Below this is a 'Your Search Item' section with a checkbox and the text: 'Which part of the image do you like? Draw a box on the item to find your search instances.' A large image of a red high-heeled shoe is shown with a search box overlaid on it.

The main content area shows 'Search Results' for 'Women's Shoes'. It includes a 'Sort By' dropdown set to 'Relevance', a 'Price' filter, and a 'Change Your View' section with icons for grid, list, and map views. The results are displayed in a grid format. Each item includes a small image of the shoe, a title, a price, and a 'Shop at Like.com' button. The items listed are:

- Natural Comfort - LV58**: \$99.95. Description: 'A easy closed pump with a pillow-like footbed to keep your feet happy. Leather or patent leather upper, air spaced memory foam footbed, covered heel, leather sole. Colors: Black, Navy, Details: Save to LikeList'.
- Cole Haan 'Carma Air' Patent Leather Open Toe Pump**: \$275.00. Description: 'Open toe dress pump, cushioned pump with a wrapped foot and a mini platform. Color(s): Black patent, dark-chocolate suede. Also patent, black python, natural python, beige leather. Brand: Cole Haan. Colors: Black, Details: Save to LikeList'.
- rtmp - Catdye**: \$89.95. Description: 'An easy on-the-eye pump features platform sole to make it easy on your feet too, patent leather uppers, alloyed studded toe. Cushioned footbed, covered heel, leather outsole, made in Italy. P. 01. More Details: Save to LikeList'.

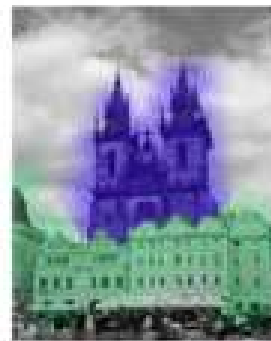
<http://www.cs.utexas.edu/~grauman/>



# Exploring community photo collections



Snaveley et al.



Simon & Seitz // [www.cs.utexas.edu/~grauman/](http://www.cs.utexas.edu/~grauman/)



# Pattern Discovery



**Objects** Sivic & Zisserman



**Categories** Lee & Grauman

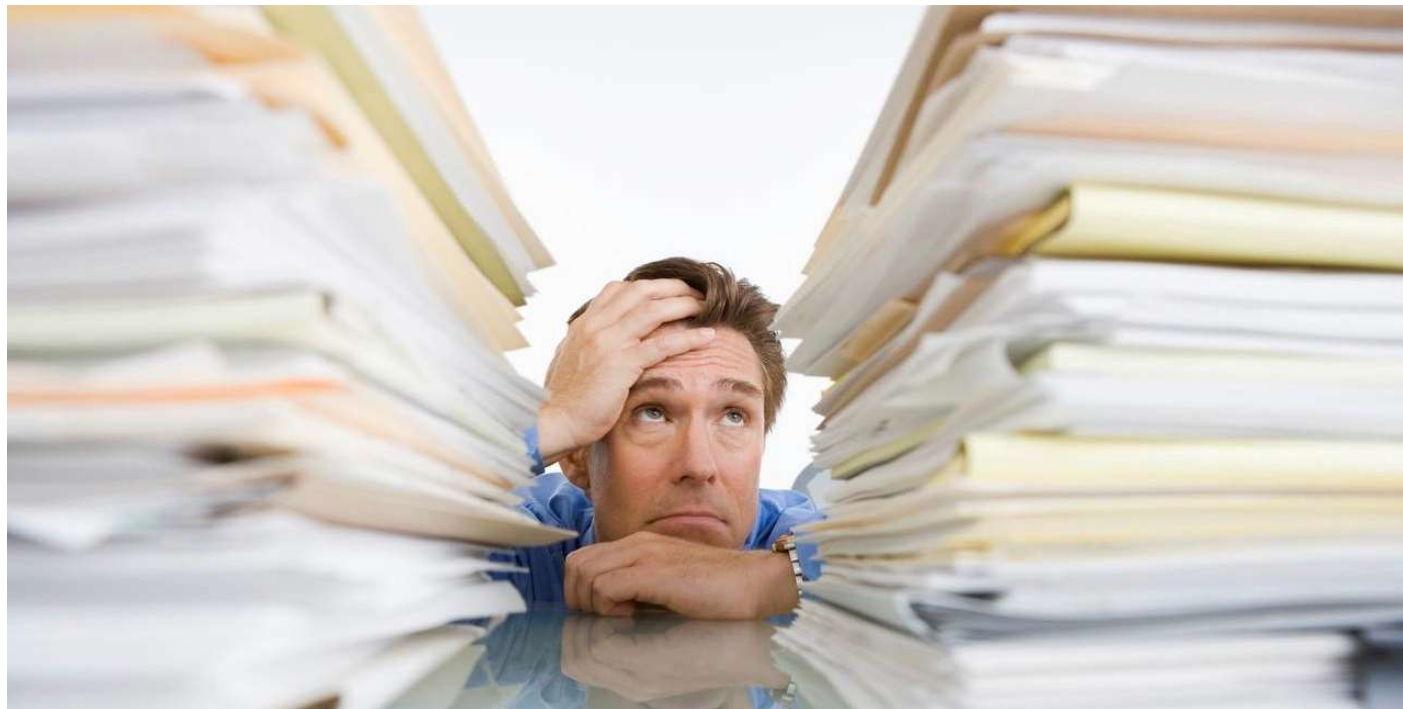


**Actions**

Wang et al.

---

# Why Challenge?

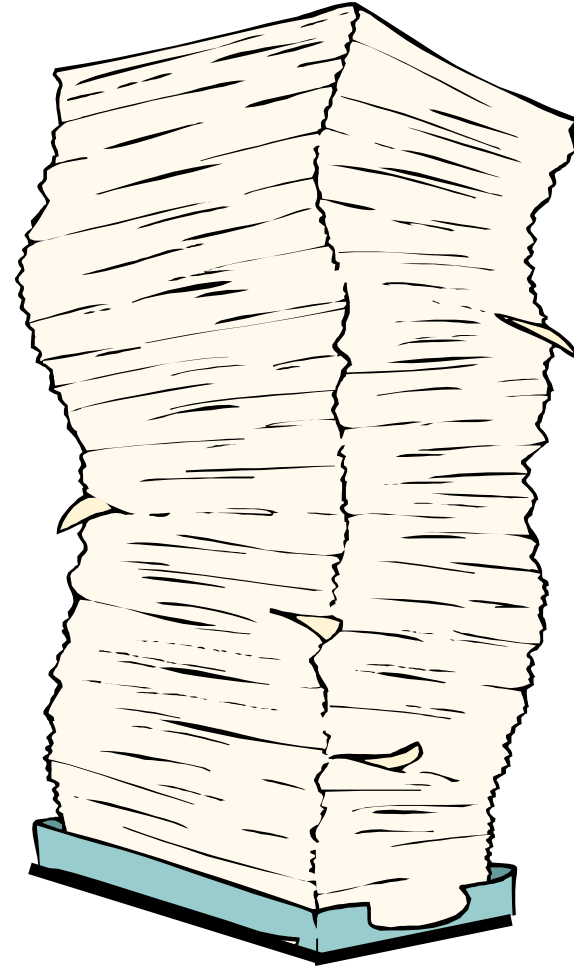


# Motivation: Scalable Searching

---

50 Thousand Images

4m



# Motivation: Scalable Searching

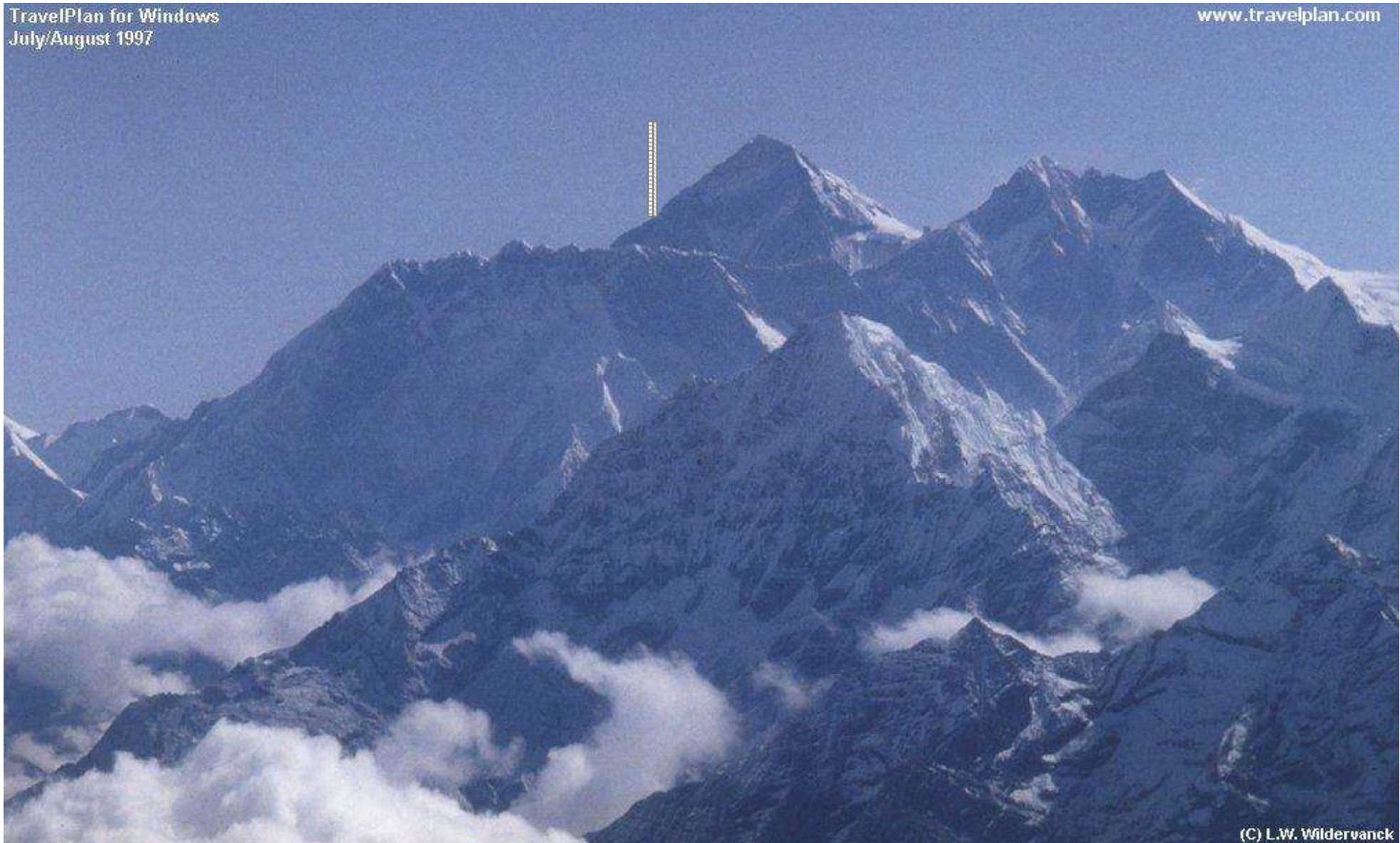


# Motivation:

---



# Motivation: Scalable Searching





# Scalable Searching and Optimization

---

10 Billion Images in 1 Seconds



# Summary of Challenges

---

- **Data Complexity**

- Billions of images indexed by Google Image Search
- Millions of videos every day
- Thousands to millions of pixels in an image
- 3,000-30,000 human recognizable object categories
- 30+ degrees of freedom in the pose of articulated objects (humans)
- 18 billion+ prints produced from digital camera images in 2004
- 295.5 million camera phones sold in 2005
- About half of the cerebral cortex in primates is devoted to processing visual information [Felleman and van Essen 1991]

- **Computational Complexity**

- How to search interested data in a blink?



---

# What We have?



# Text based Searching

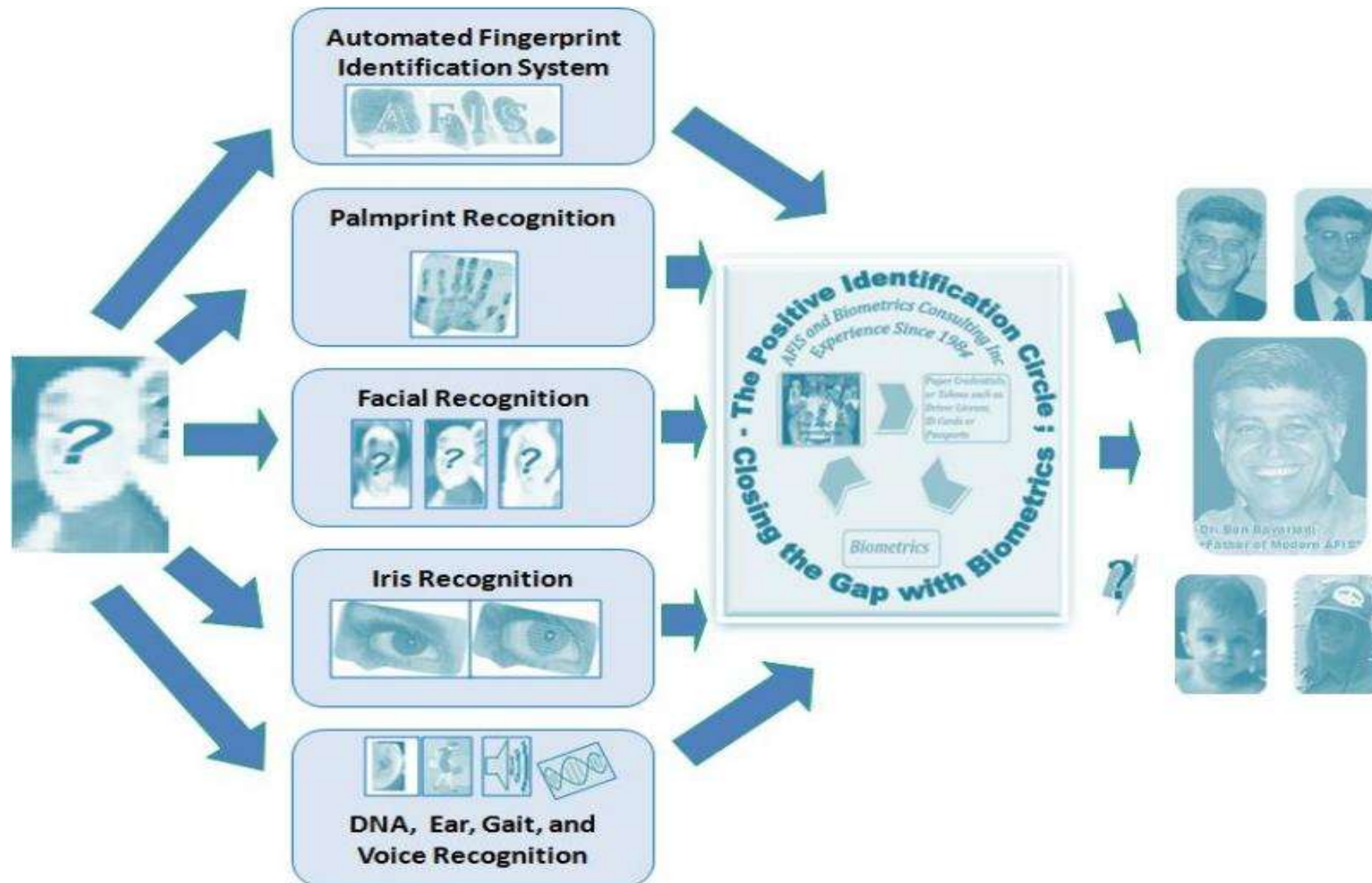
- Text-based multimedia search
  - File IDs, Keywords, Captions



# Searching and Recognition



# Searching and Recognition



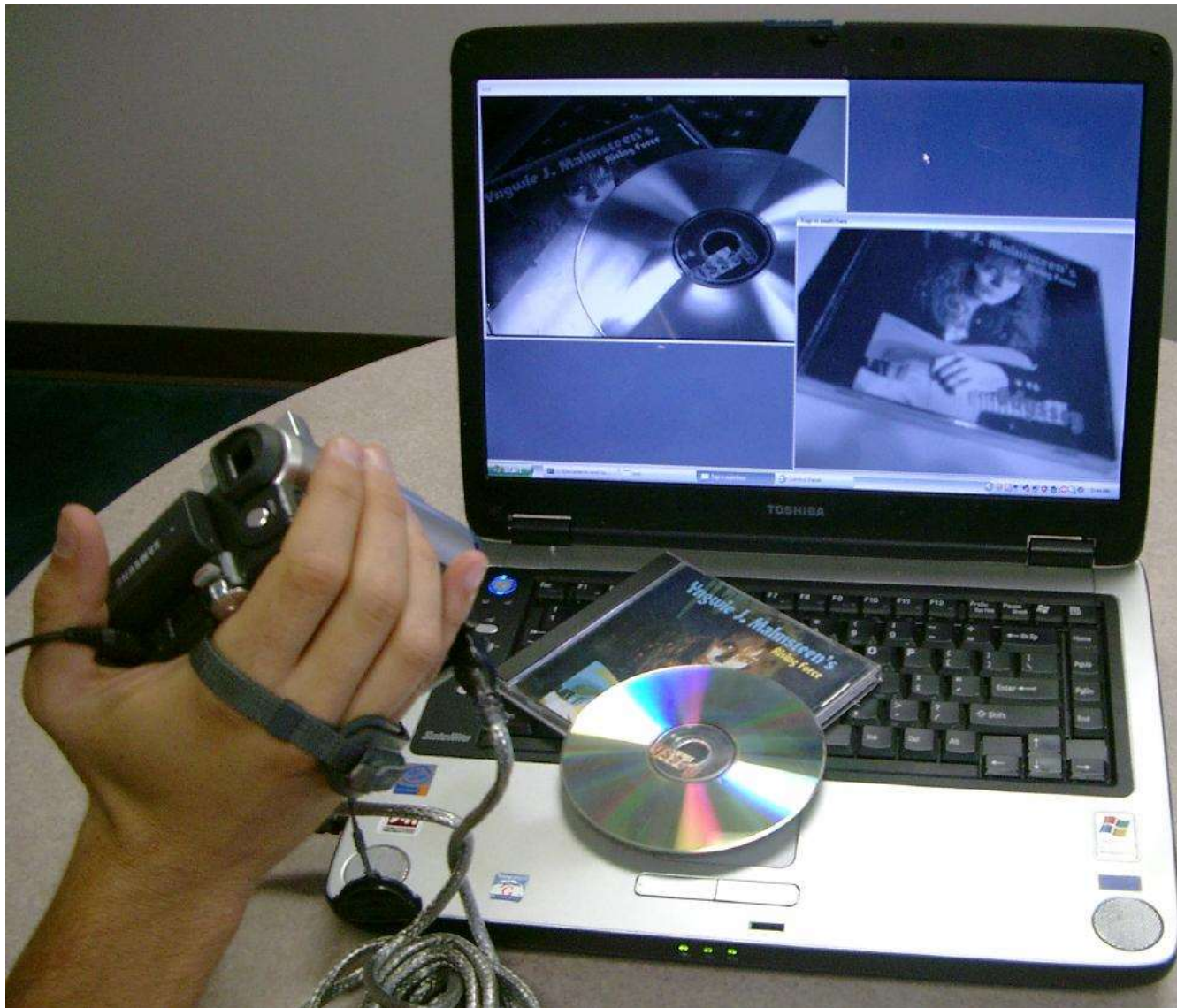
# Searching and Recognition

---



# Searching and Recognition

---





# Course Goals

---

- **Teaching Goals**

- Introduction to the mathematical, statistic and computational principles underlying modern learning, searching, imaging, and vision systems.
- Fundamentals of machine learning techniques as well as more general concepts required for them
- Popular algorithms/techniques in machine learning as well as applications of these learning algorithms/techniques to data mining, computer vision, medical image analysis, bioinformatics, et al.
- Application-driven and includes newly emerged topics in learning, imaging and vision.
- Selected hot topics relating to the emerging random theory and machine learning techniques.



# Course Expectation

---

- **What to expect from the course:**
  - Will cover key issues and concepts in class..
  - A mid-term exam and a final exam (None)
  - Paper Reading
- **What do I expect of you:**
  - Come to class
  - Read the papers (Listening to me or presenters is not good enough)
  - Work through the problems in the papers (not really homework... but it helps)
  - Ask questions (**IMPORTANT**)



# How to read and review research papers?

---

- **Brief (2-3 sentences) summary: what is the problem?**
- **What is the background the problem?**
- **What are the STOA methods to solve the problem?**
- **What is the proposed solution in this paper?**
- **What is the main contribution of the proposed solution?**
- **Strengths? Weaknesses?**
- **How convincing are the experiments?**
- **Suggestions to improve them?**
- **Extensions?**
- **Additional comments, unclear points**
- **Relationships between different papers**



# Presentation Guide

---

- **Procedure**
  - Read the selected papers that you are interested in
  - Prepare for a well talk about 60 minutes
  - Answer questions, ~15 minutes
- **What should be included**
  - Problem overview, motivation, background
  - STOAs
  - Algorithm explanation, technical details
  - Contribution? Novelty?
  - What kinds of experiment conducted?
  - Any similarity and dissimilarity between techniques in the papers
  - Weakness, why? Strong, why?
  - Any possible extension



---

# Quick Tour



---

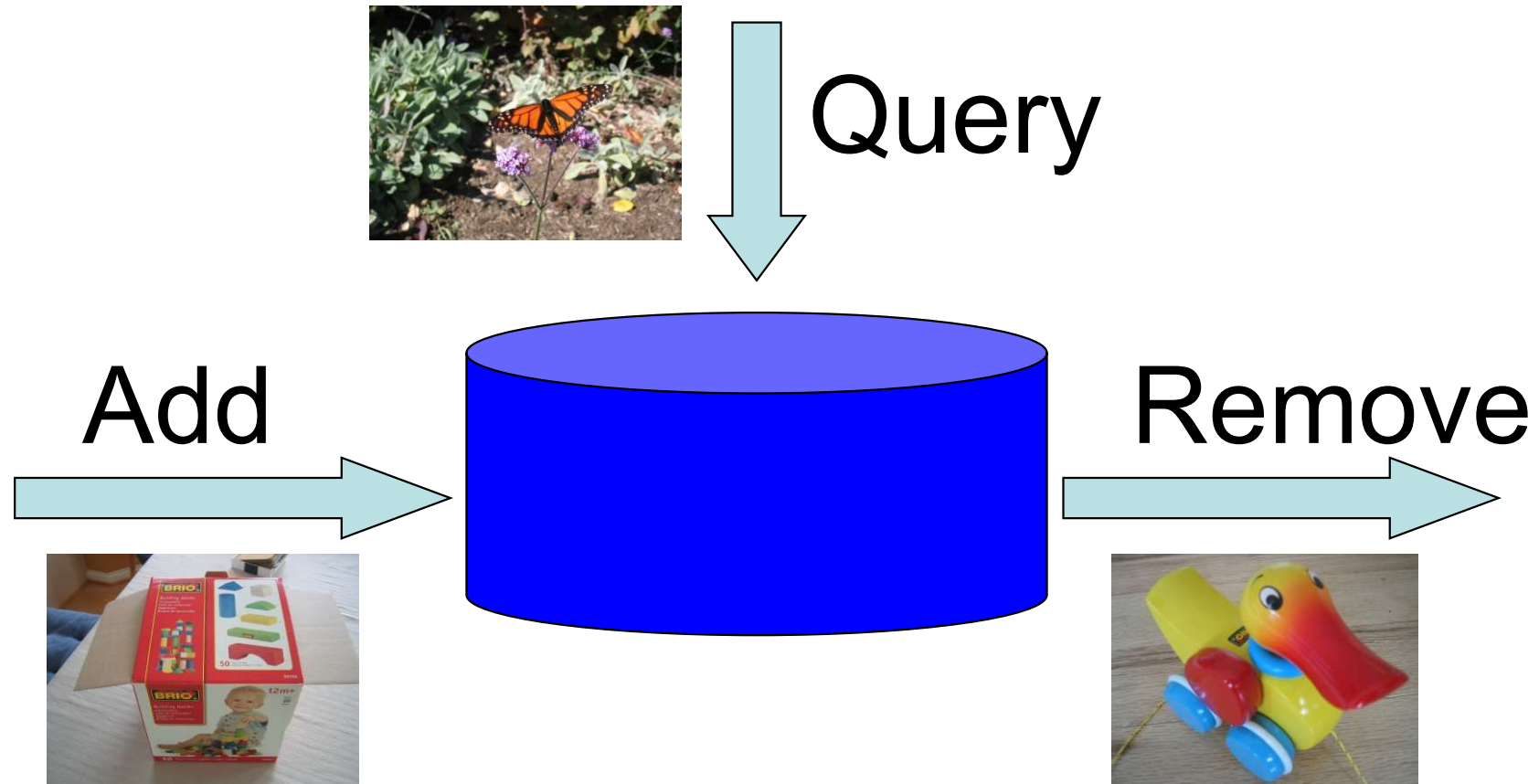
# Scalable Search and Recognition via Vocabulary Tree

<http://vis.uky.edu/~stewe/>



# Adding, Querying and Removing Images

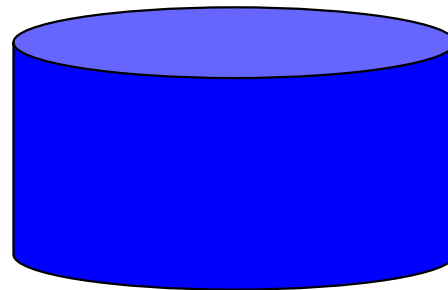
---



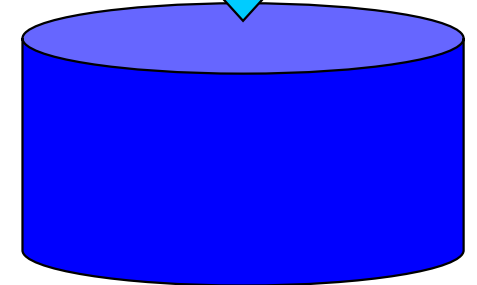
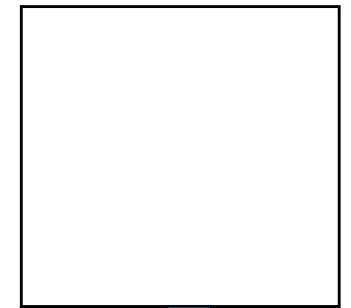
# Training and Addition are Separate

---

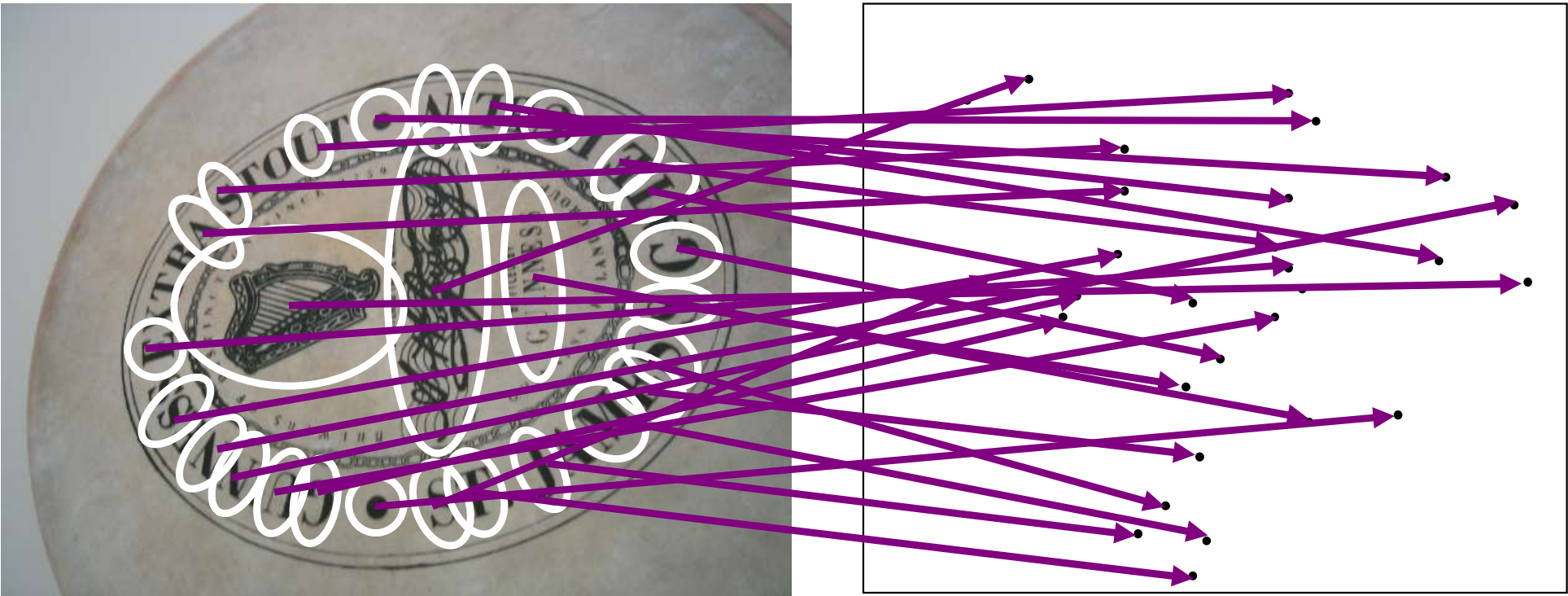
Common Approach

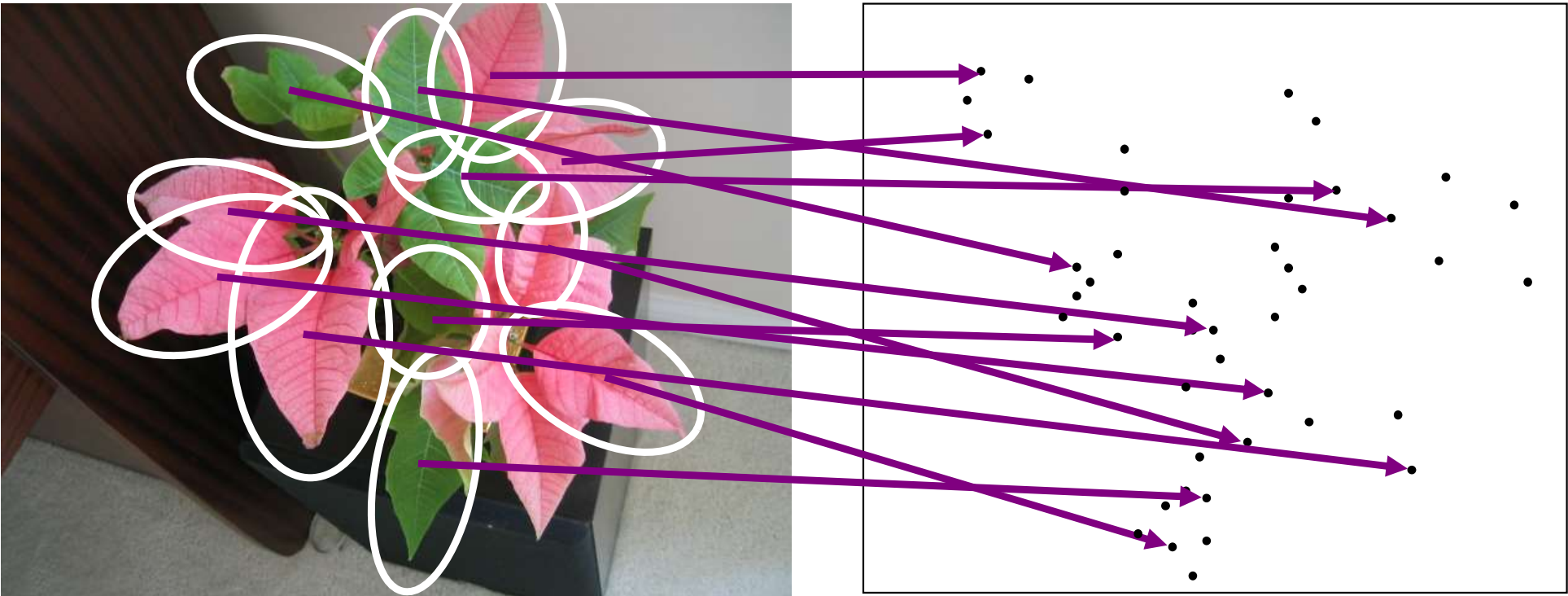


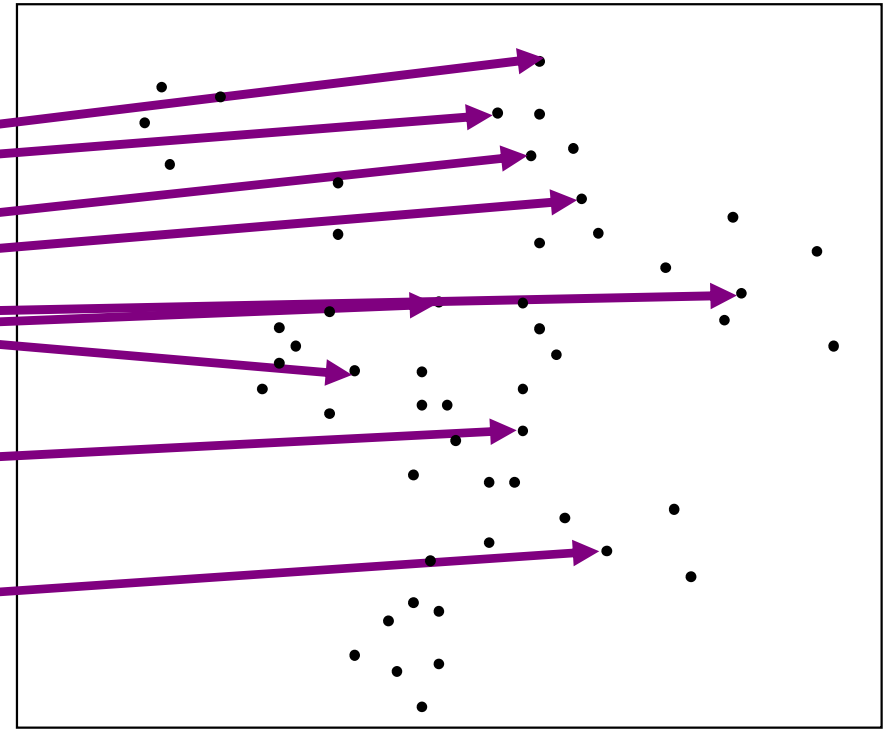
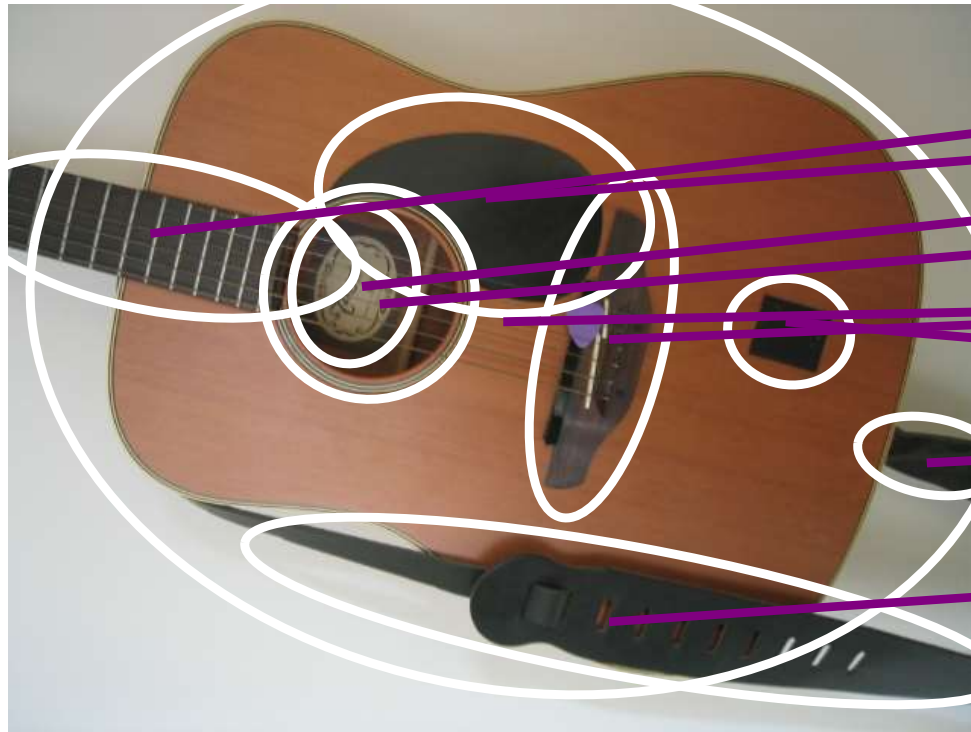
Scalable approach

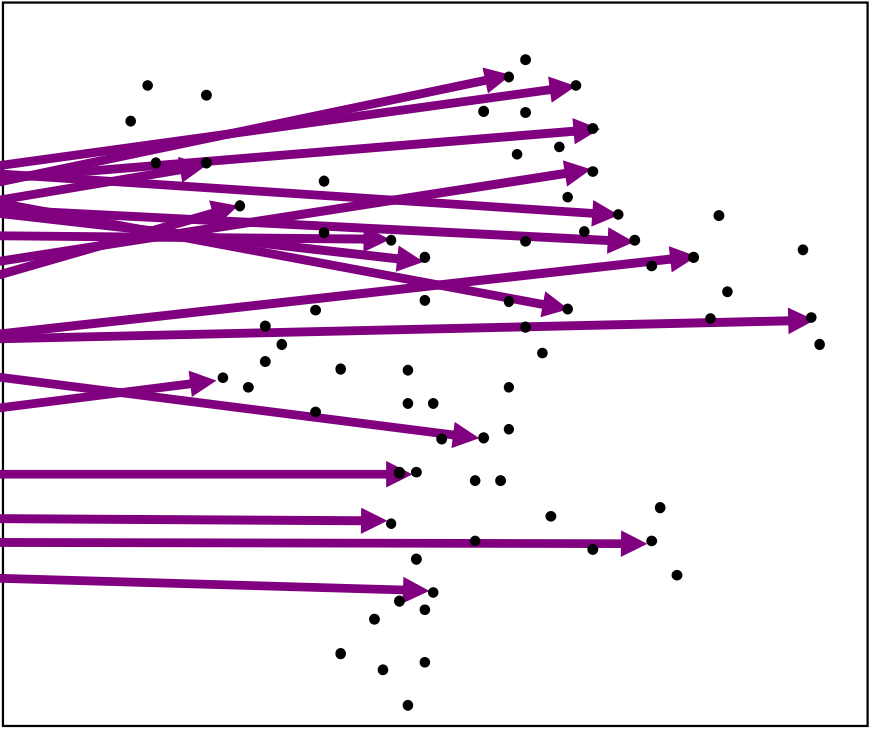
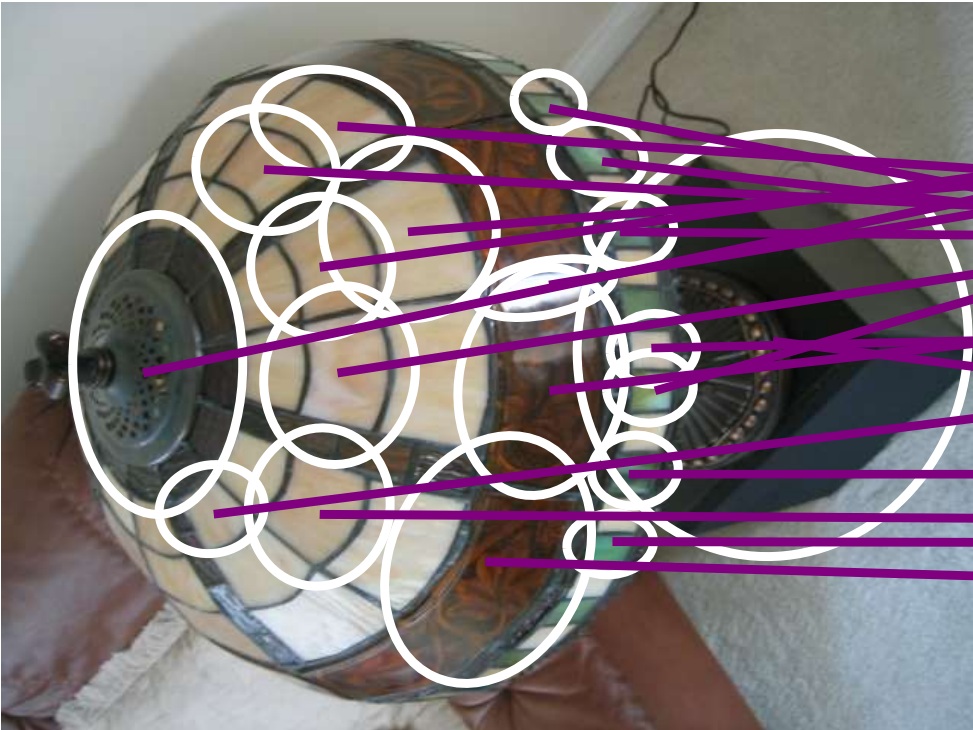


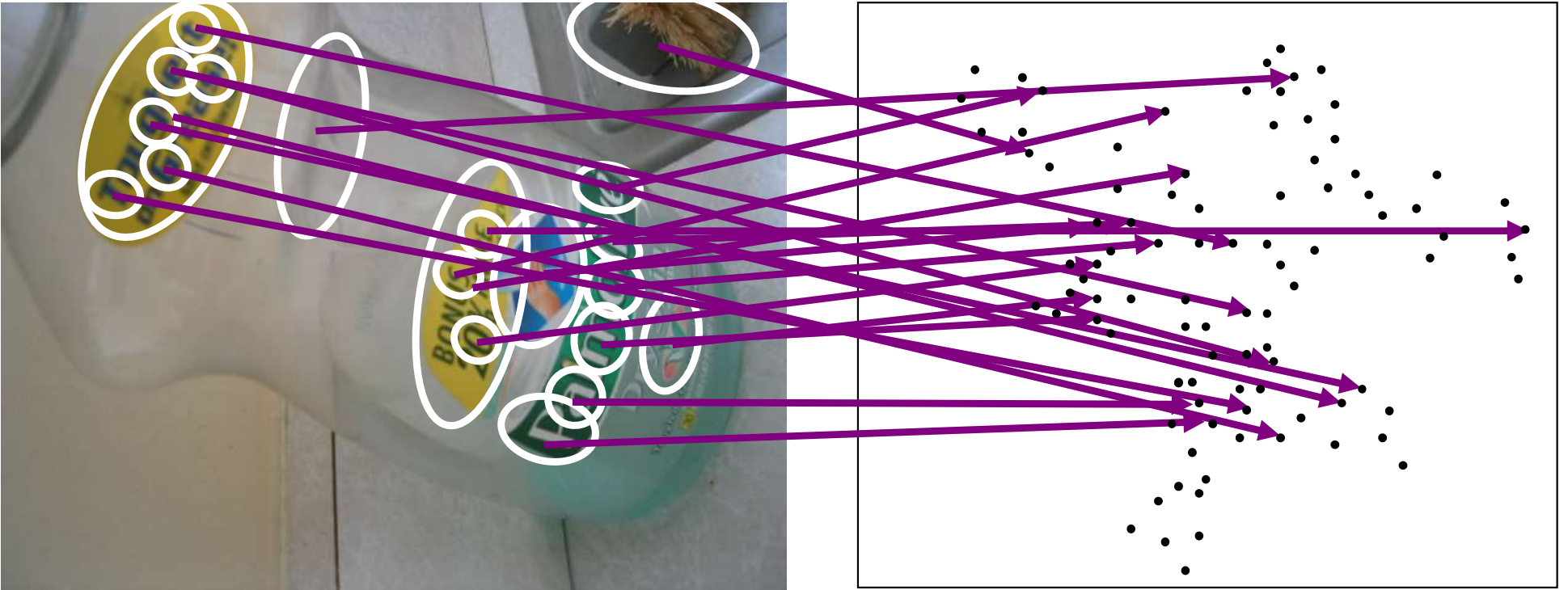


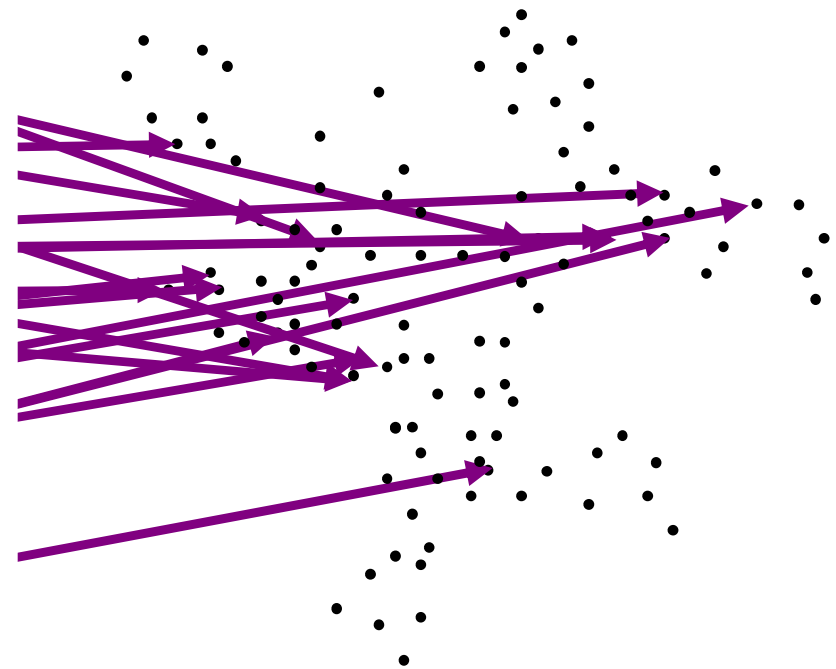


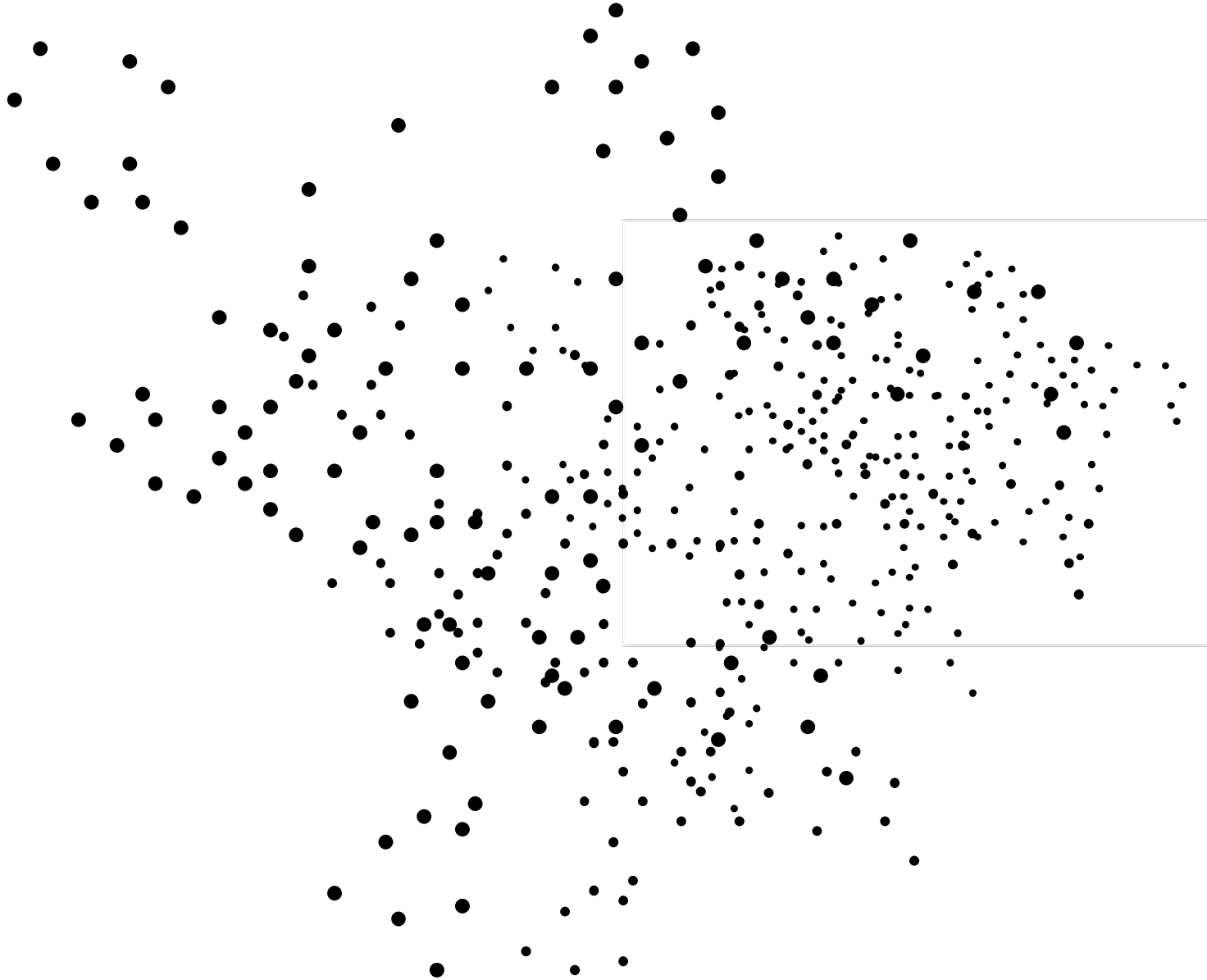


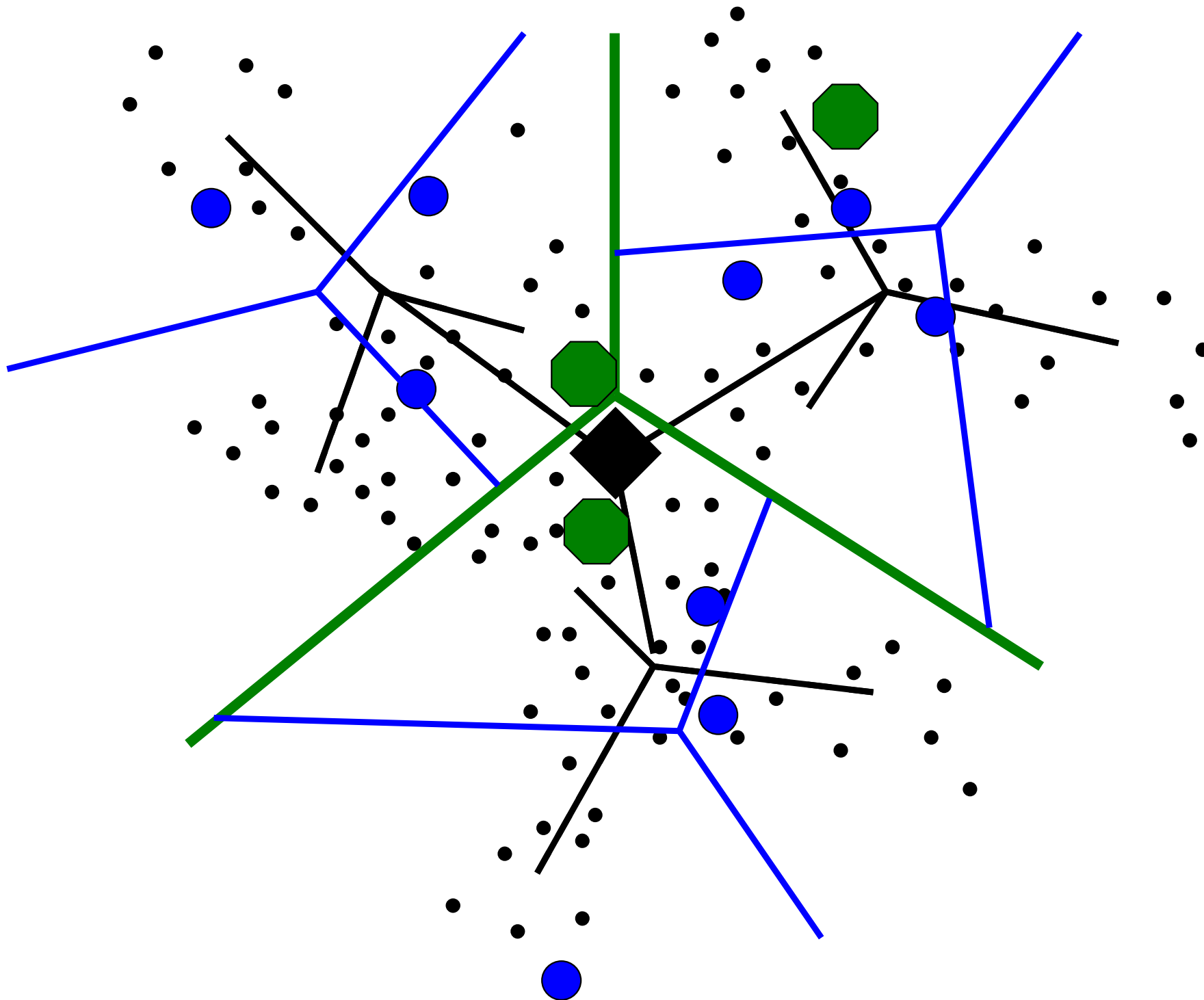




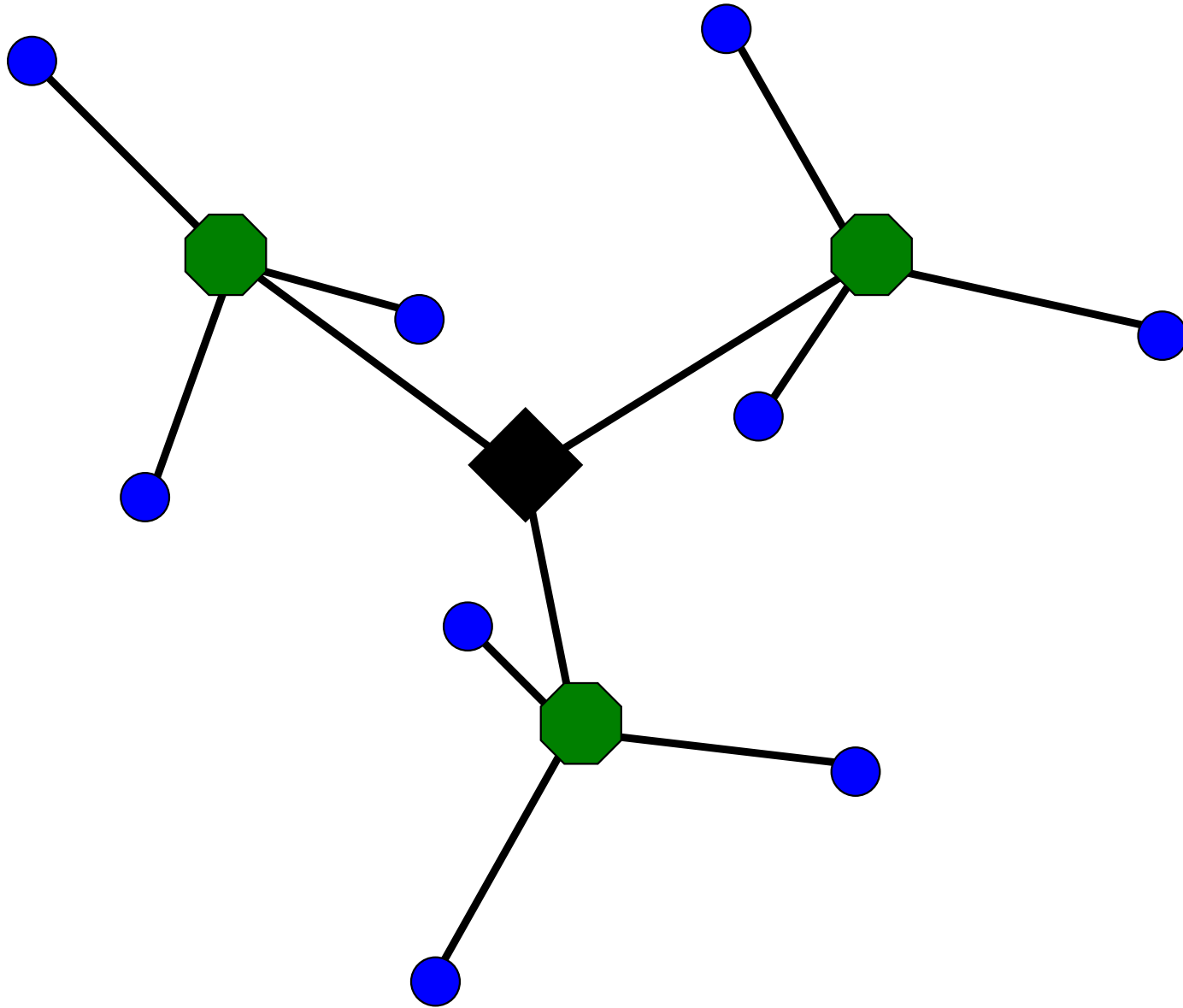


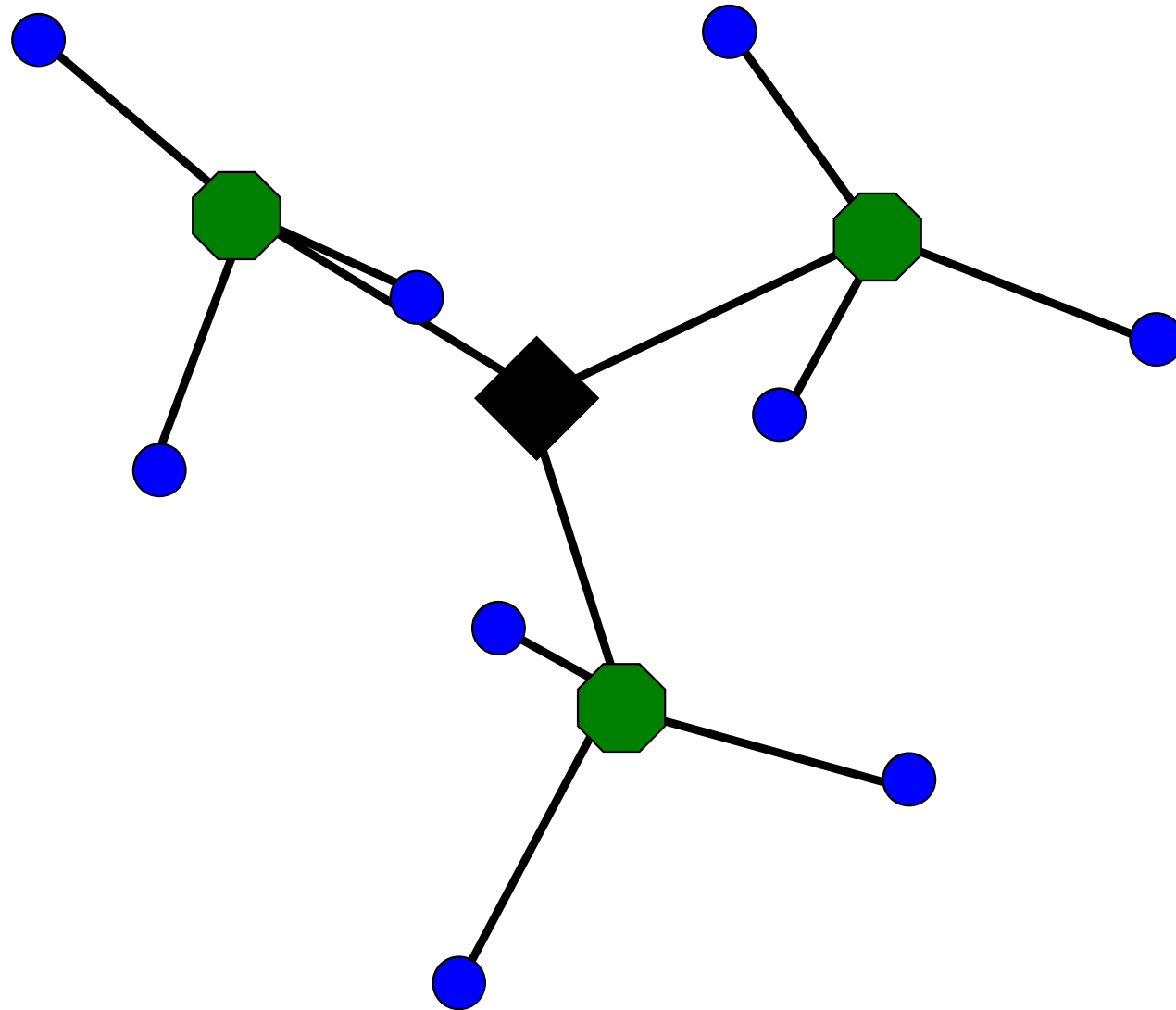


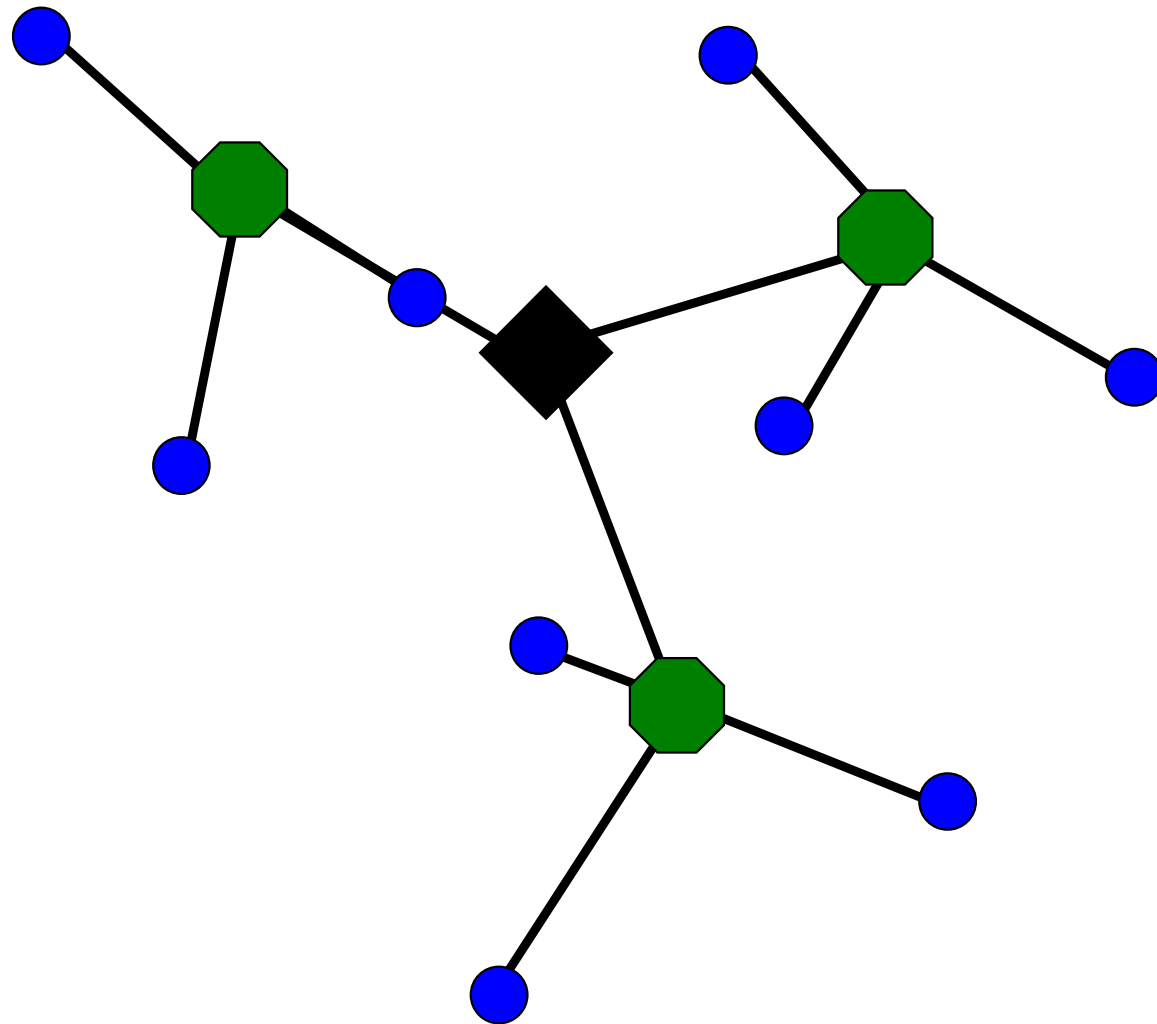


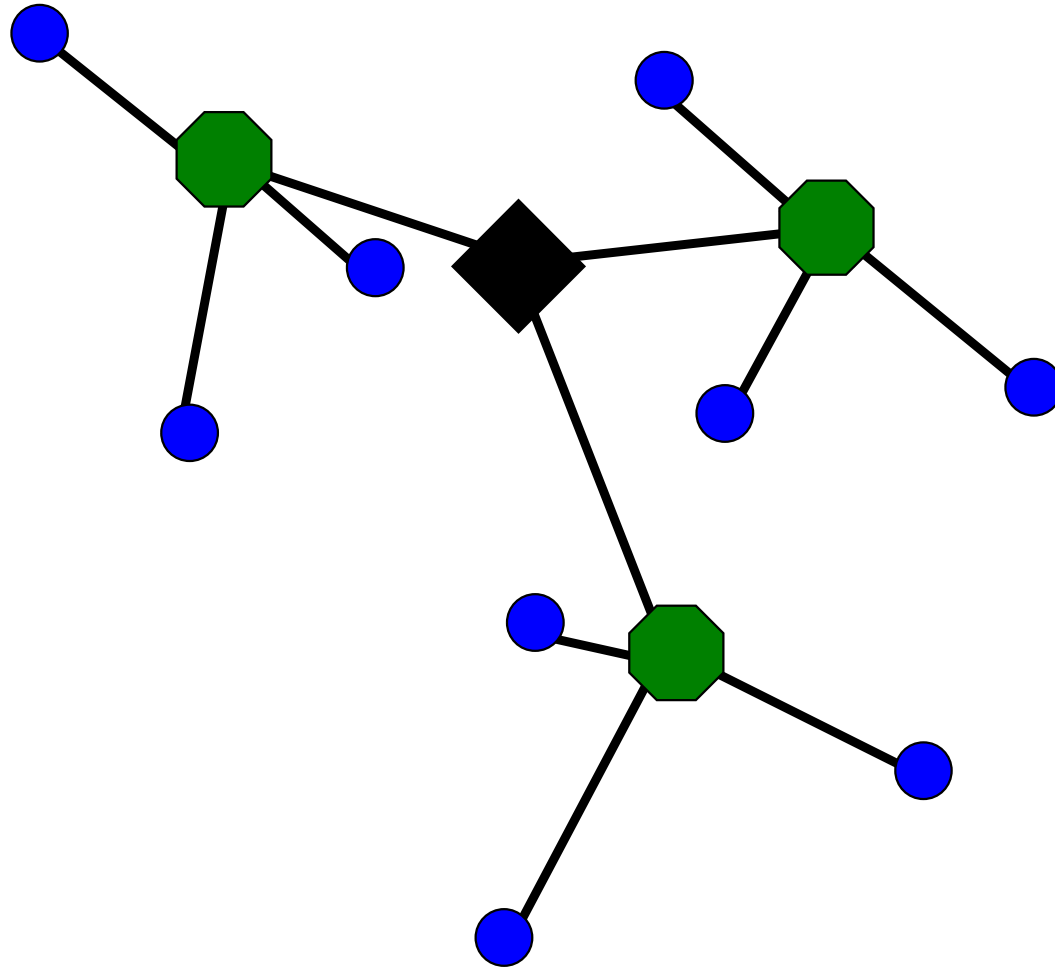


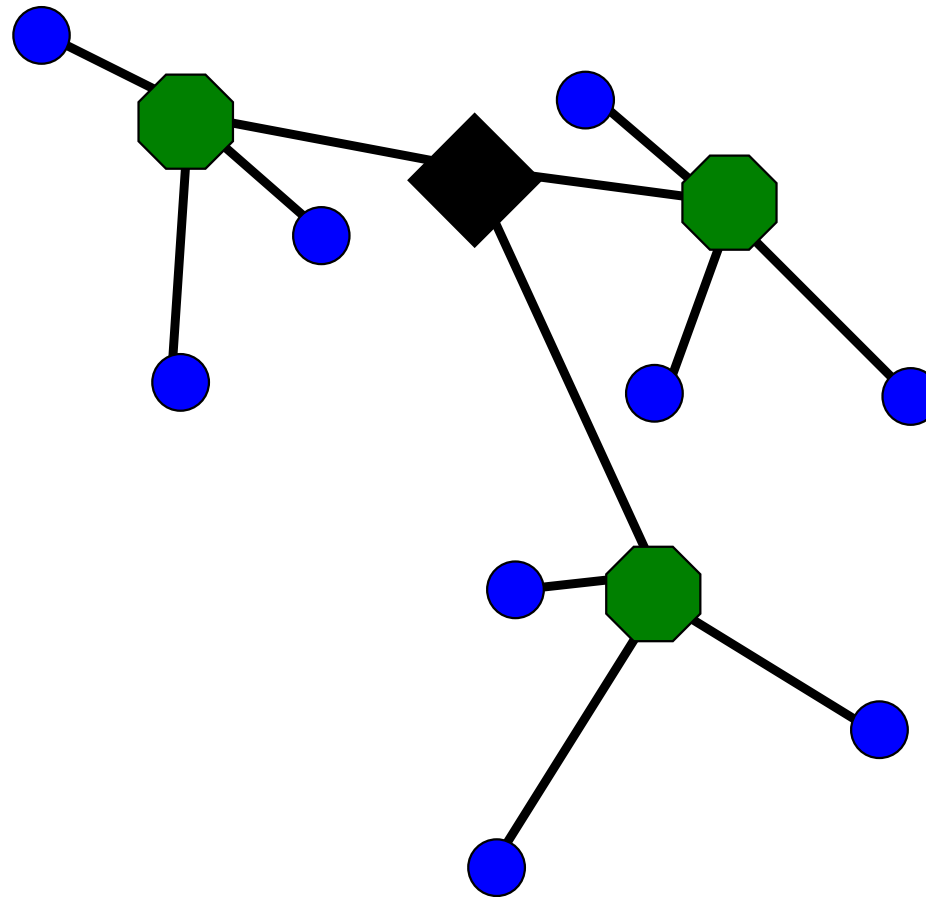


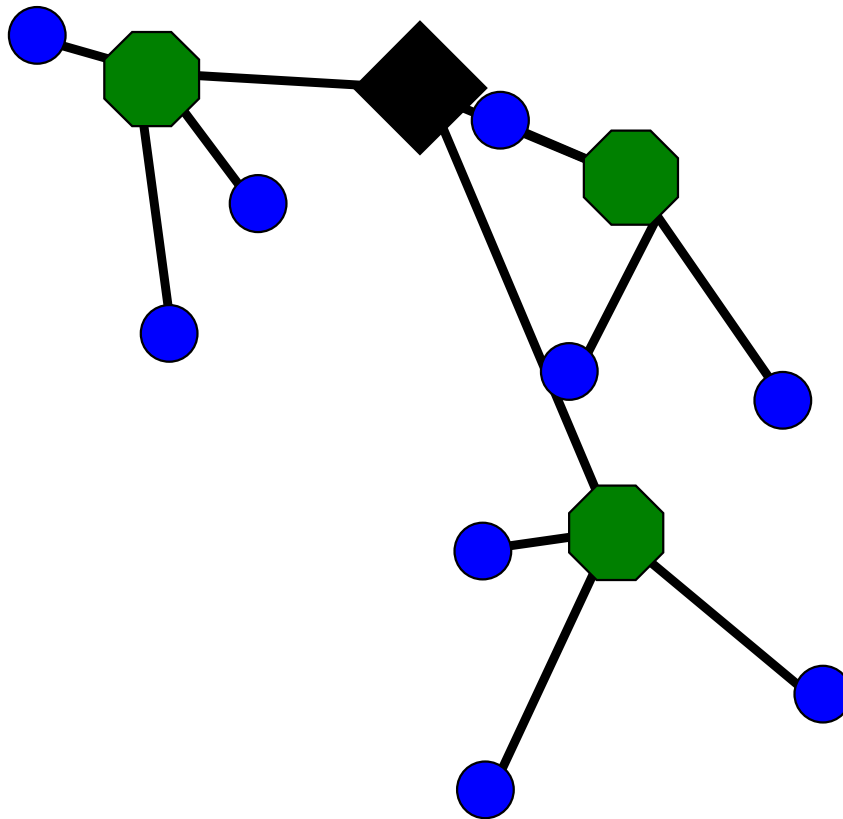


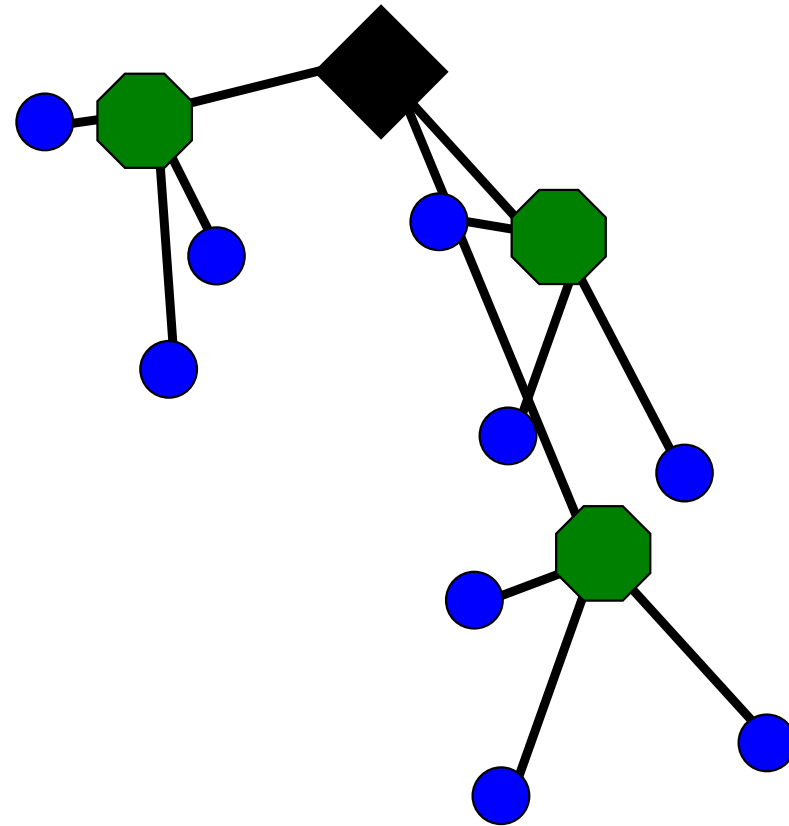


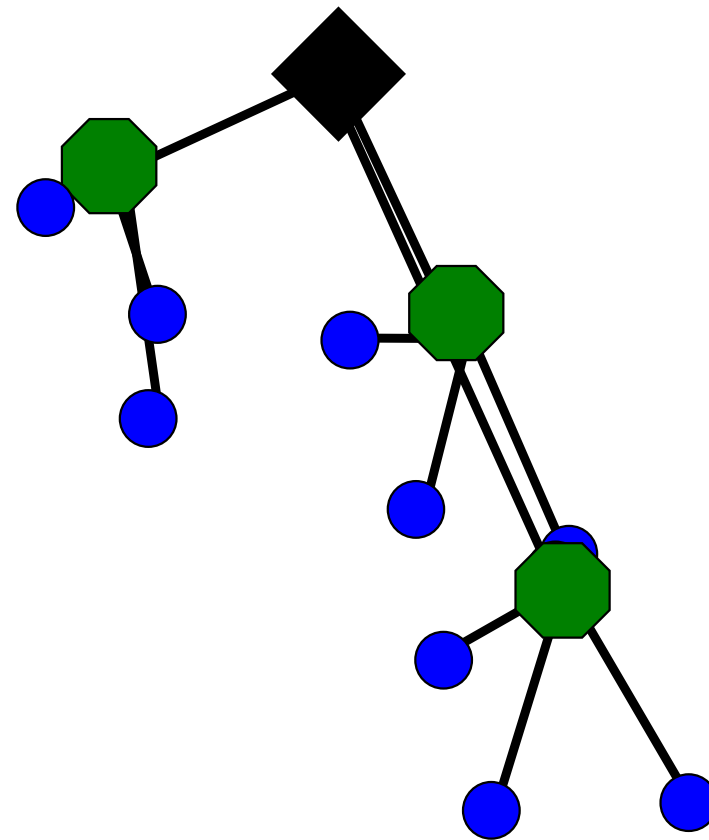




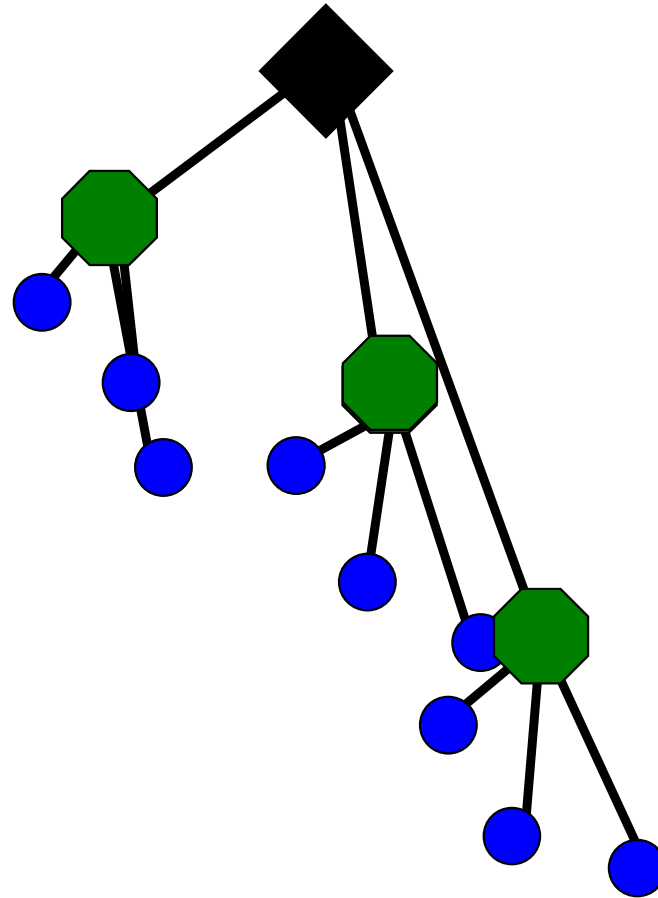


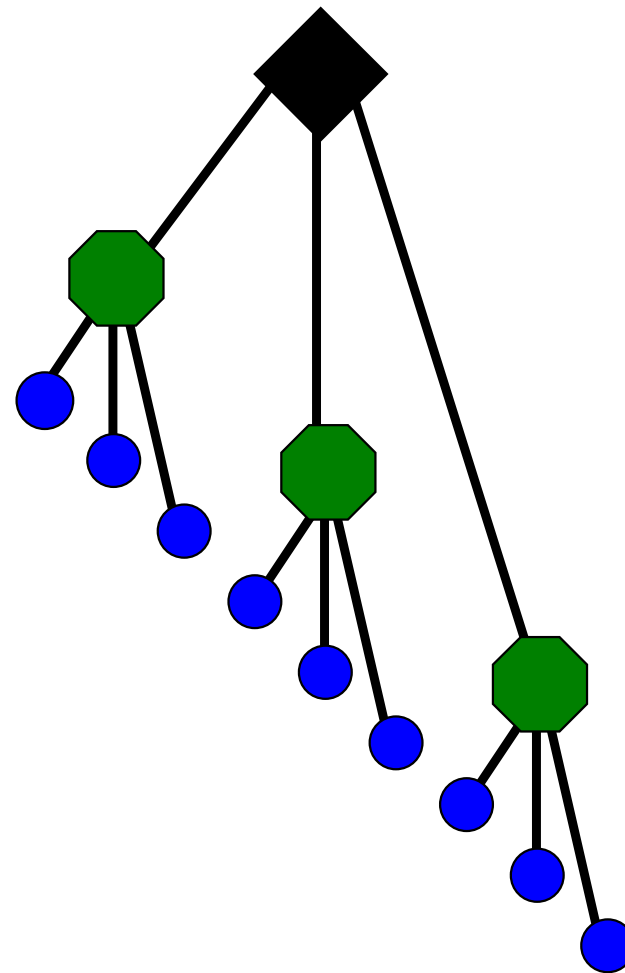


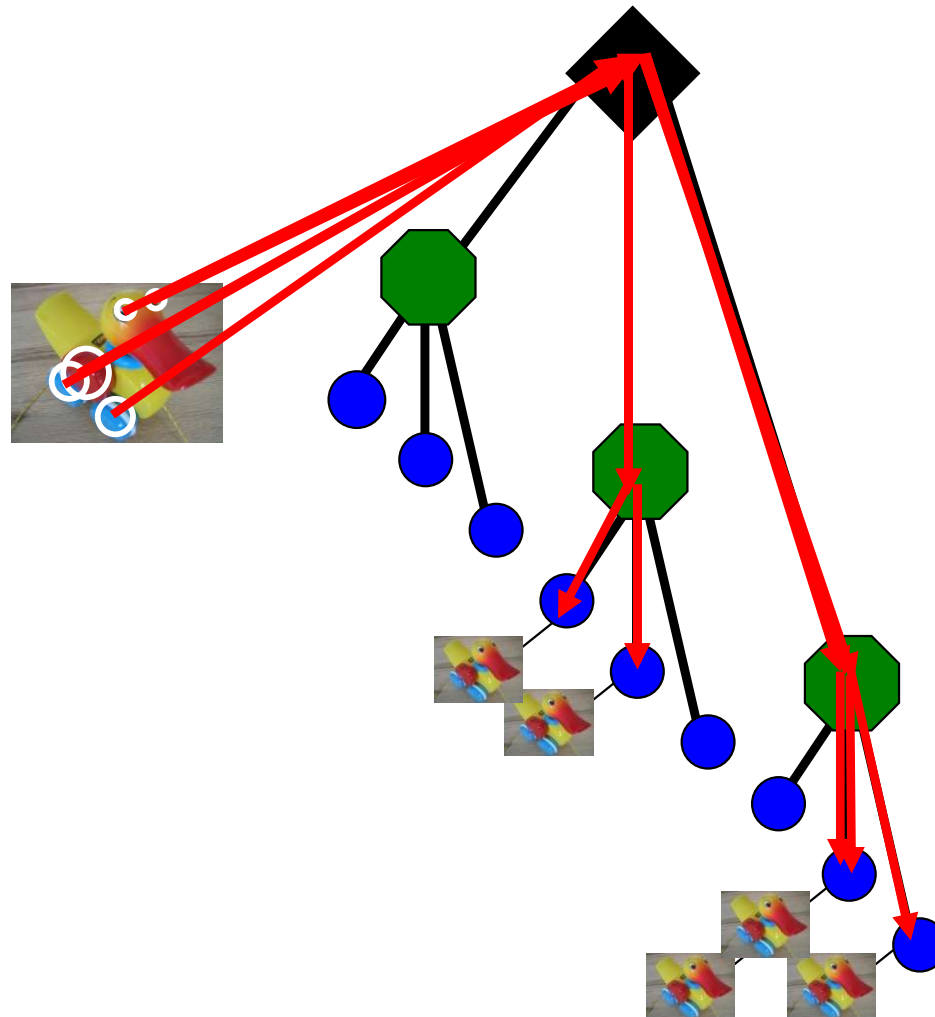


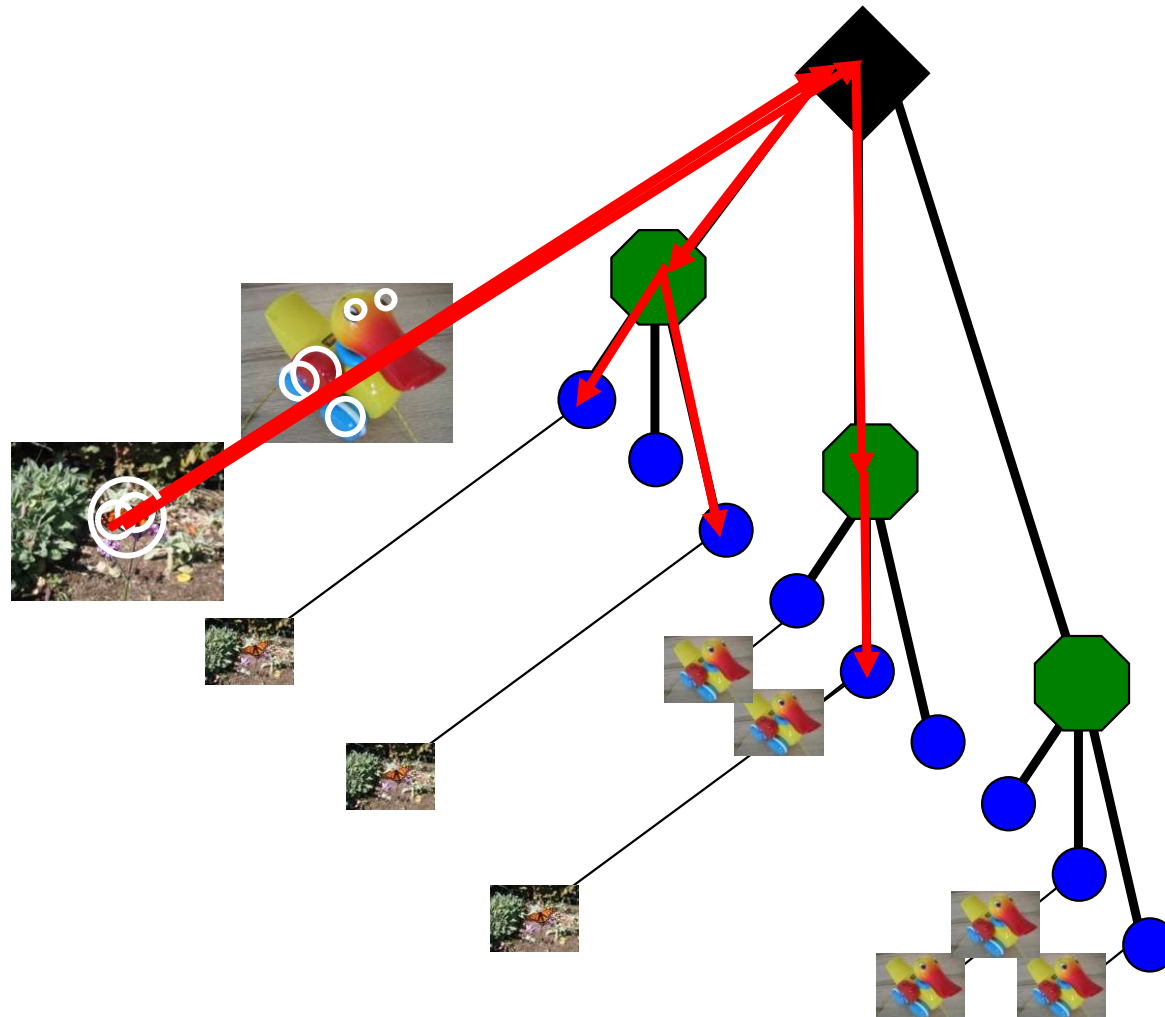


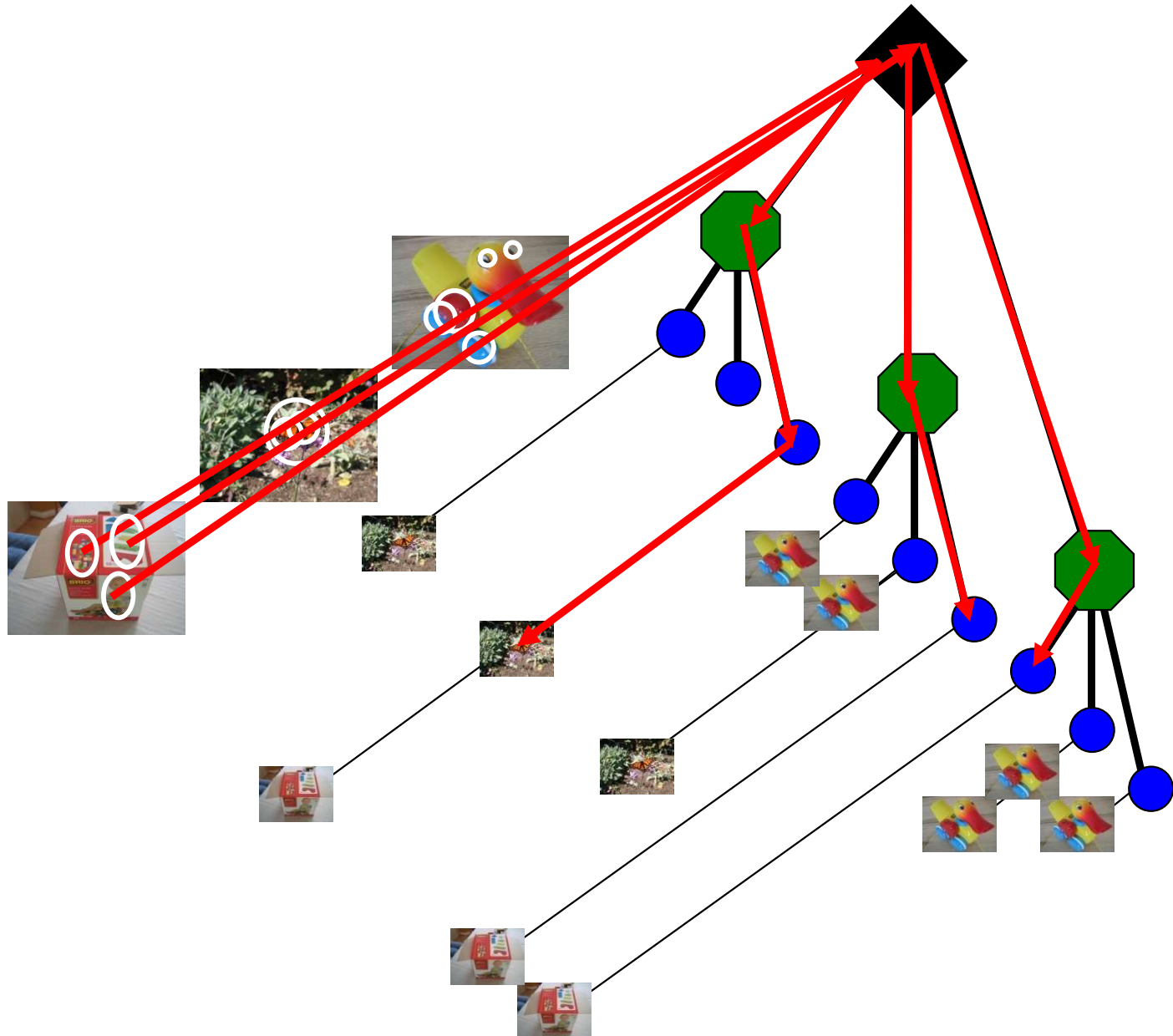


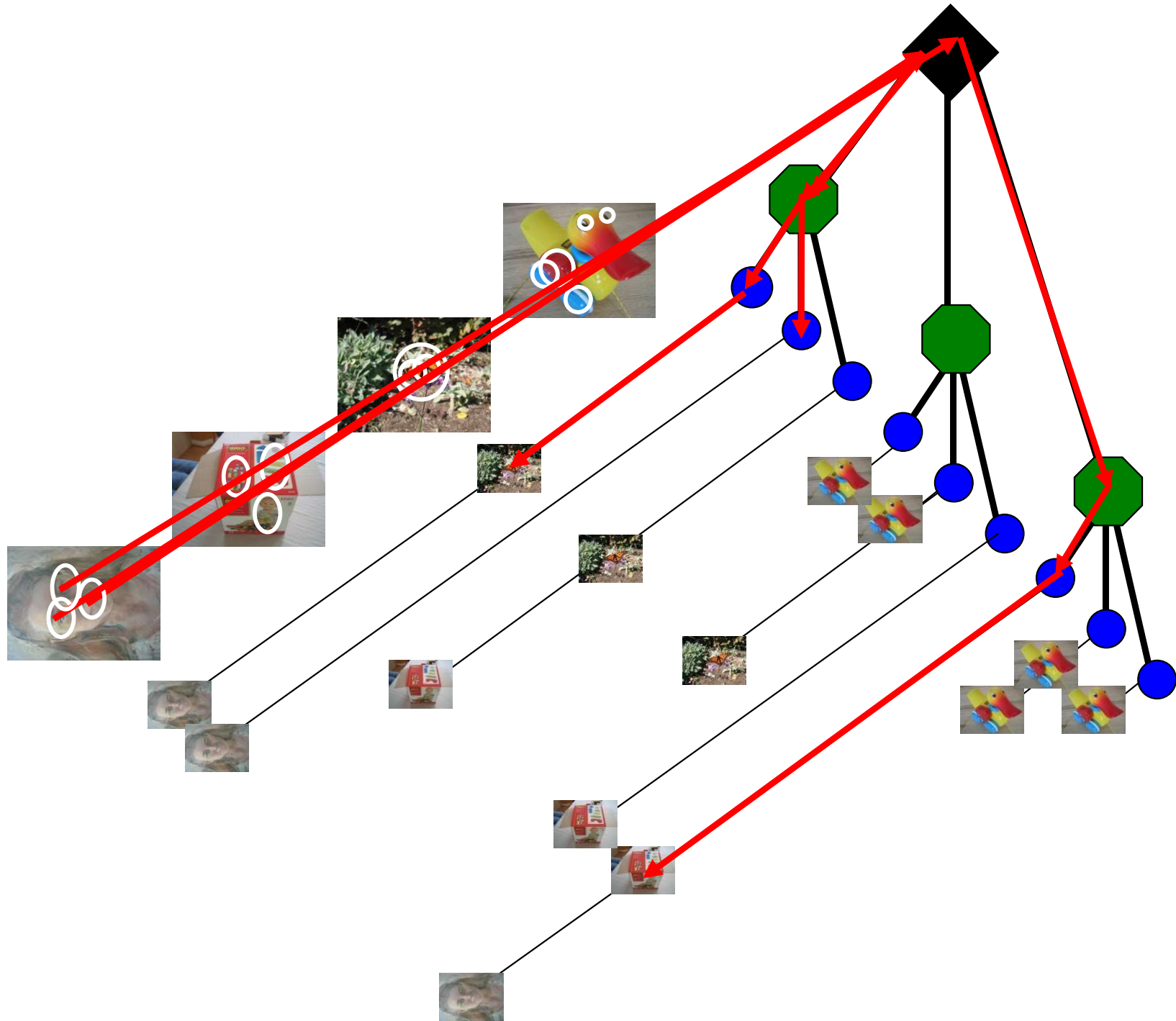


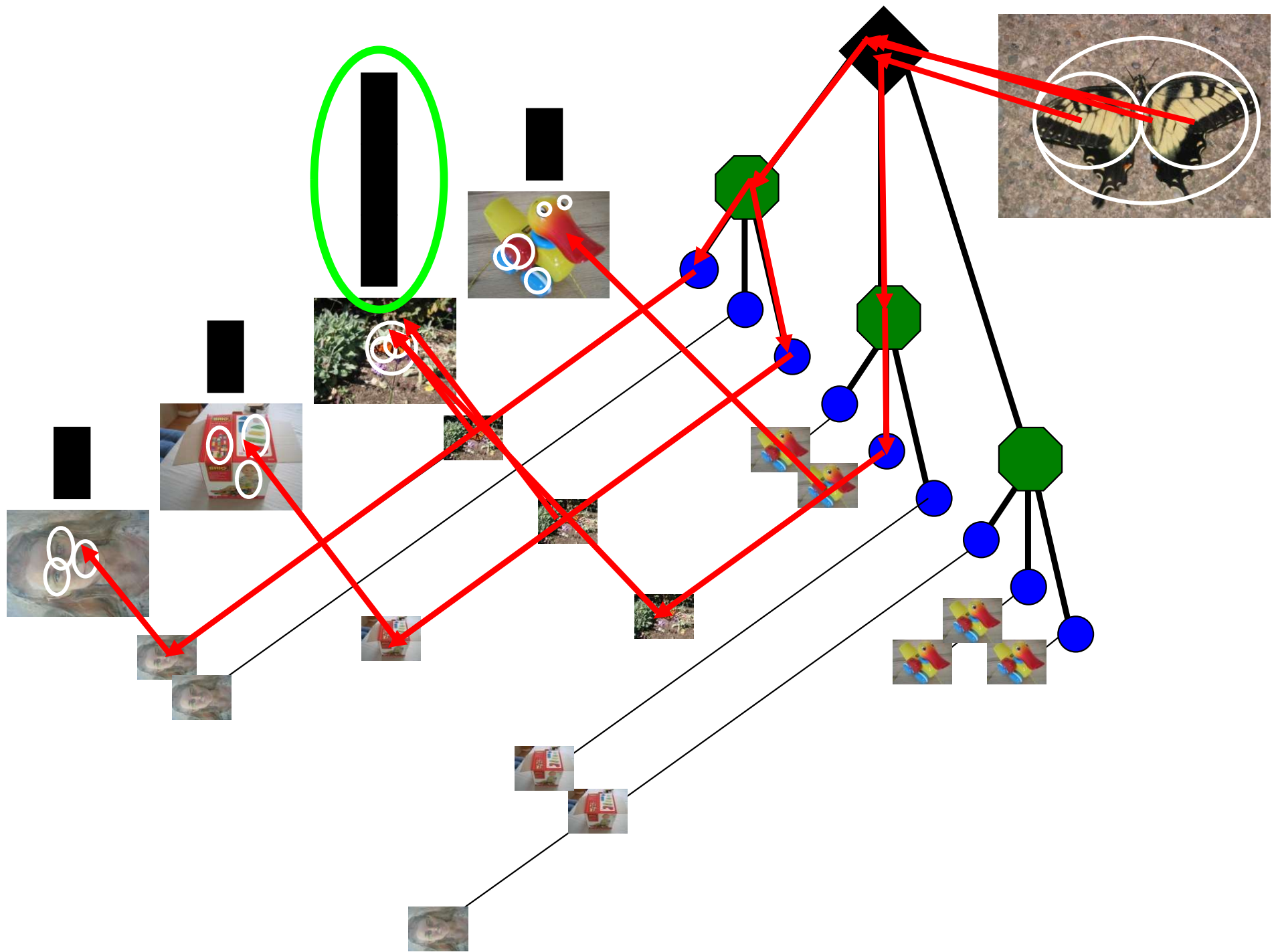




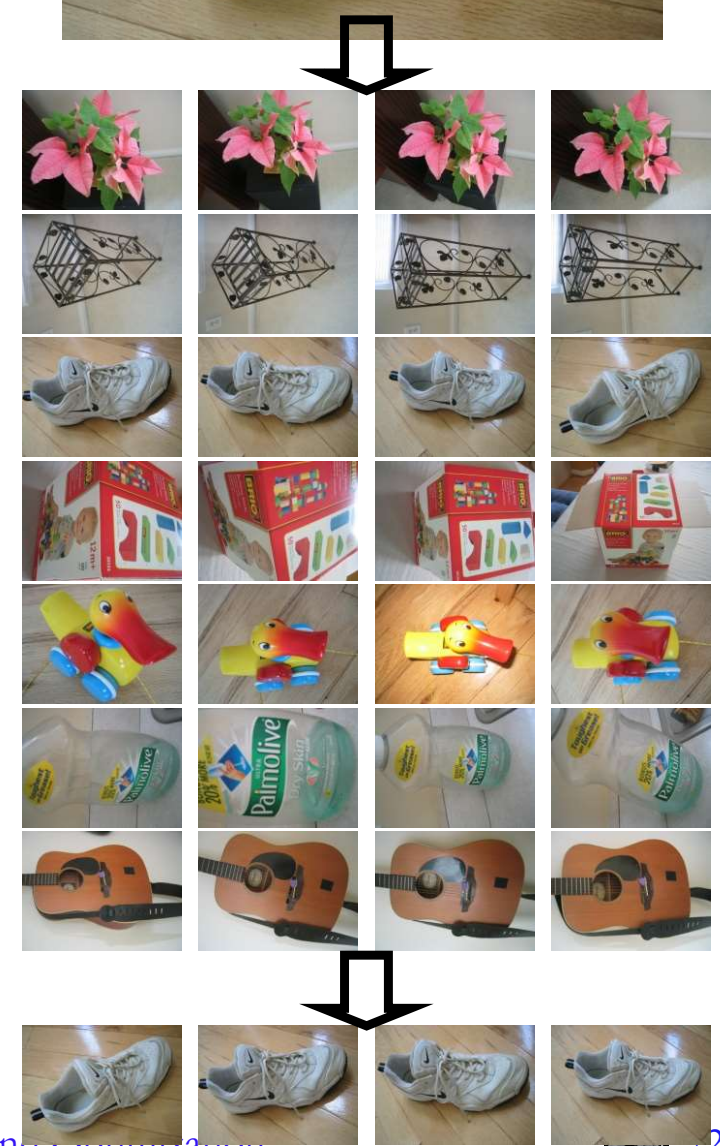
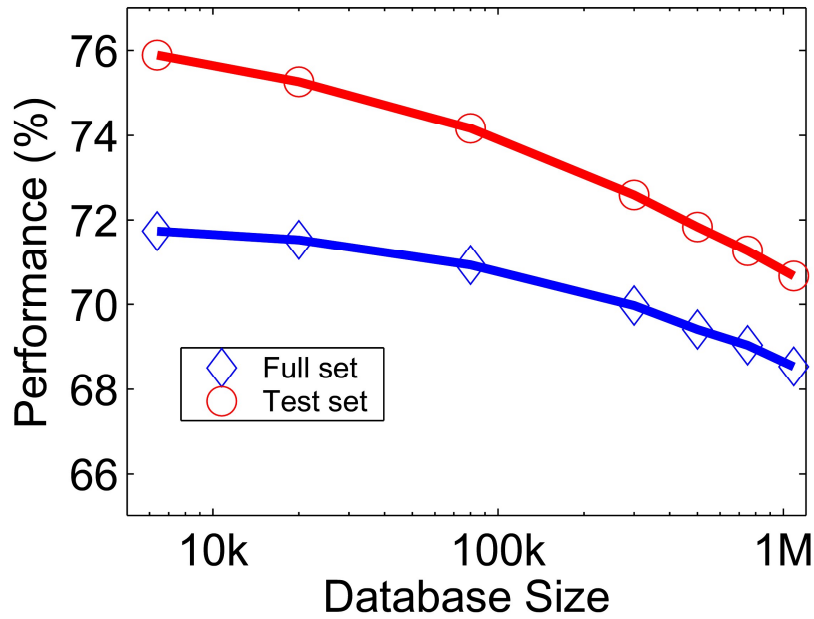








# Performance



## ImageSearch at the VizCentre

New query:  Browse... Send File

File is 500x320



Top n results of your query.



bourne/im1000043322.pgm bourne/im1000043323.pgm bourne/im1000043326.pgm bourne/im1000043327.pgm