CloudCV:
Large-Scale Computer Vision on the Cloud

http://cloudcv.org/

Dhruv Batra

Virginia Tech

Lead: Machine Learning & Perception Group @ VT

CloudCV Team:

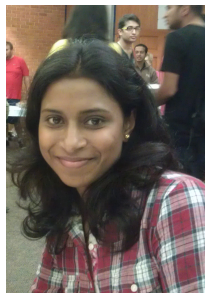
Harsh Agrawal

Clint Solomon

Neelima Chavali

Yash Goyal

Prakriti Banik

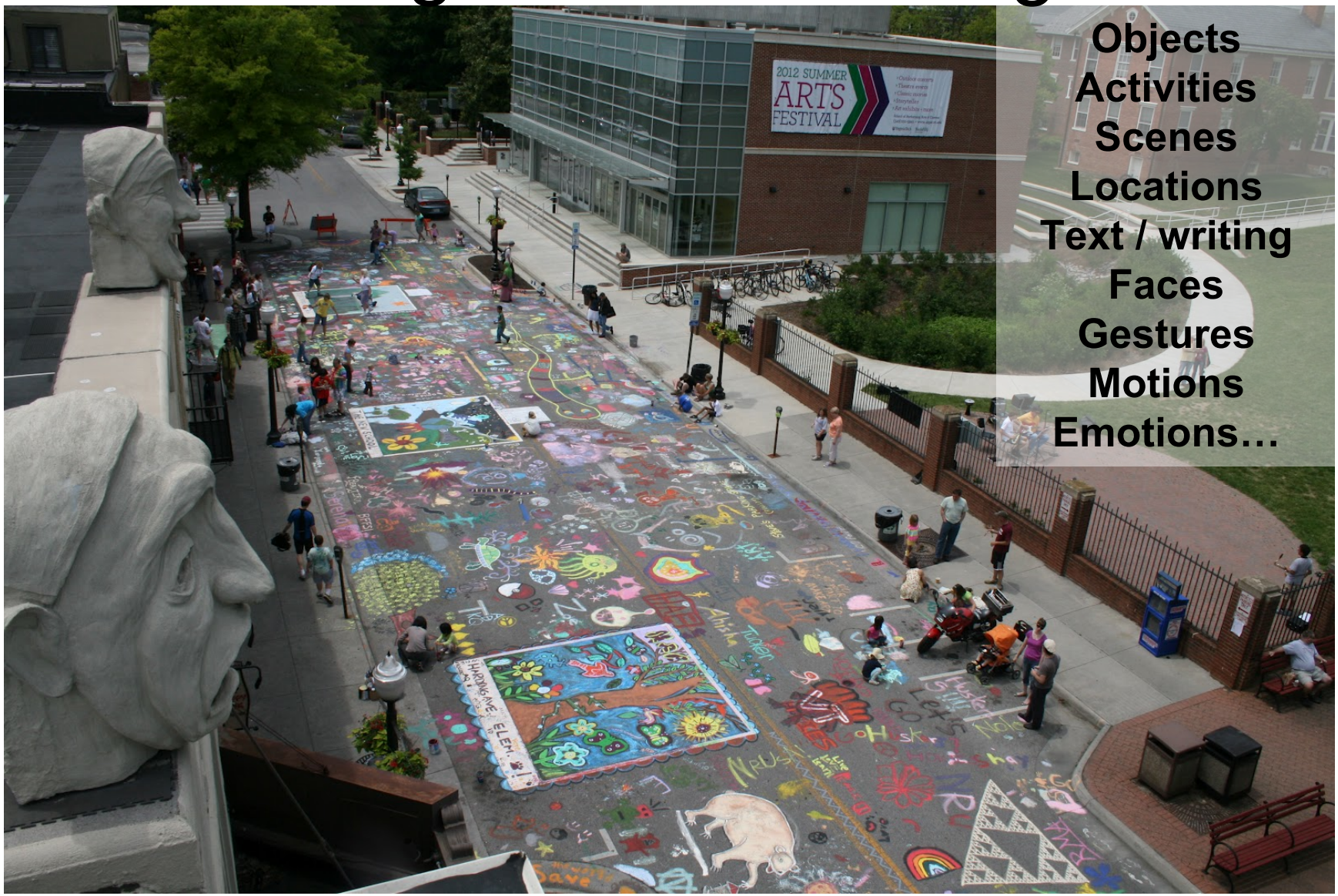


Outline

- Historical context about Computer Vision
- CloudCV
  - A mix of
    - Research in my group
    - Deployment and demos at cloudcv.org

Image from: http://kirkh.deviantart.com/art/BioMech-Eye-168367549

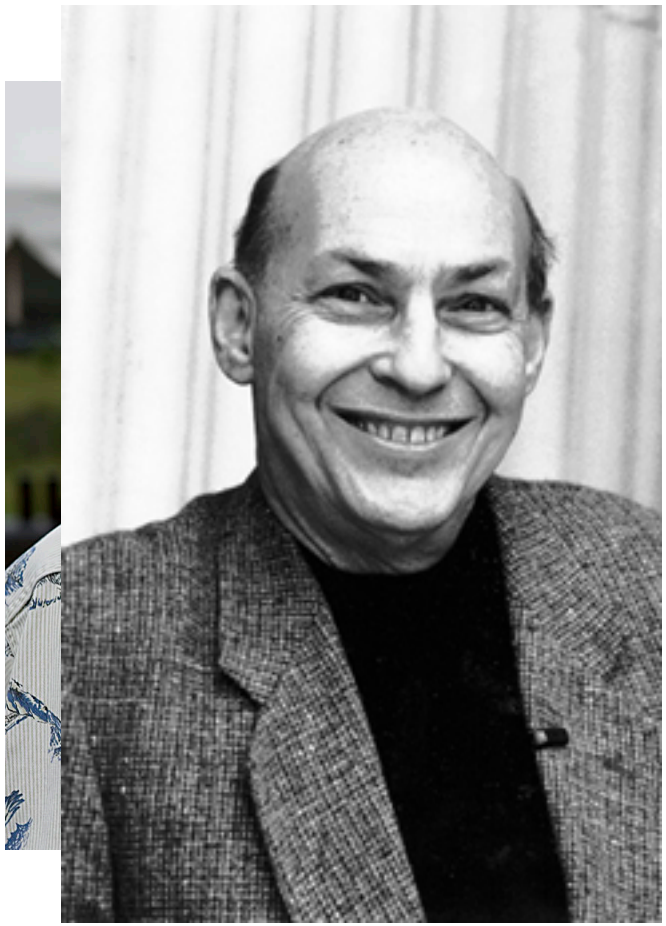
Image Understanding



"Color College Avenue", Blacksburg, VA, May 2012

Slide credit: Devi Parikh

Computer Vision

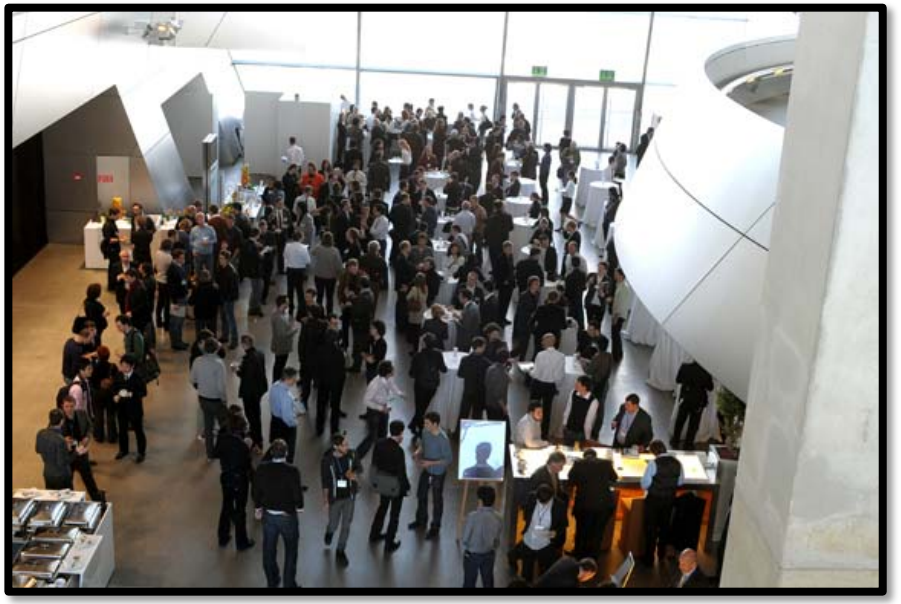


"spend the summer linking a camera to a computer and getting the computer to describe what it saw"

- Marvin Minsky (1966), MIT

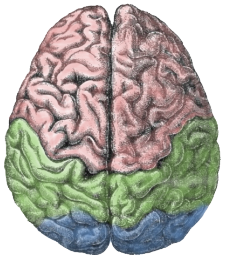
… 45 years later

Computer Vision

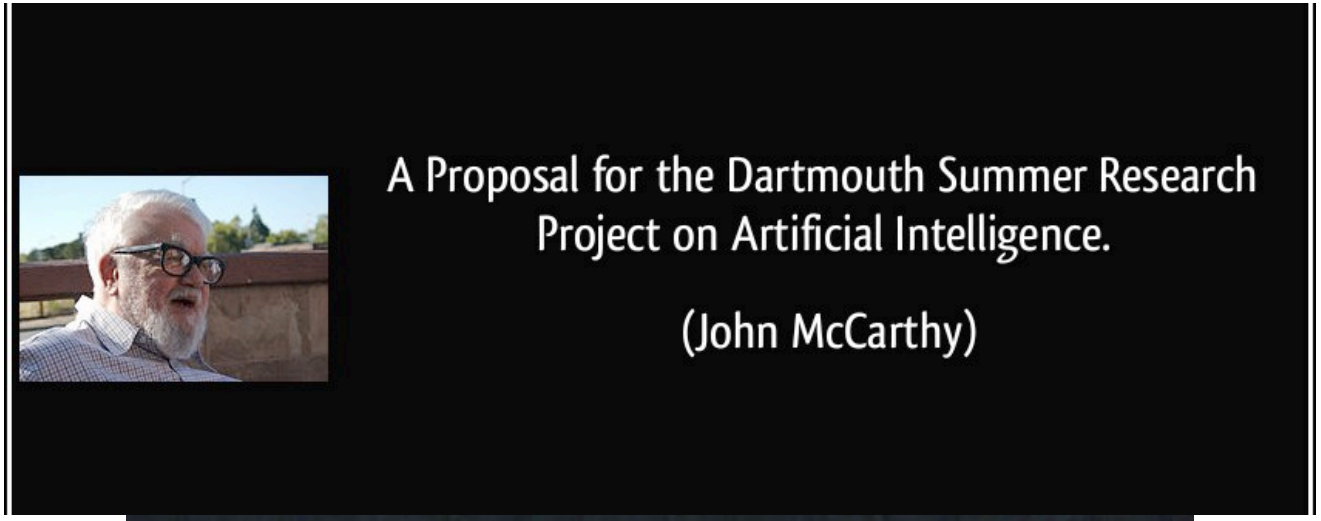


OR

Vision is HARD!



A Brief History of AI



A Brief History of AI

- "We propose that a 2 month, 10 man study of artificial intelligence be carried out during the summer of 1956 at Dartmouth College in Hanover, New Hampshire."

- The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it.

- An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves.

- We think that a significant advance can be made in one or more of these problems if a carefully selected group of scientists work on it together for a summer."

AI Predictions: Experts

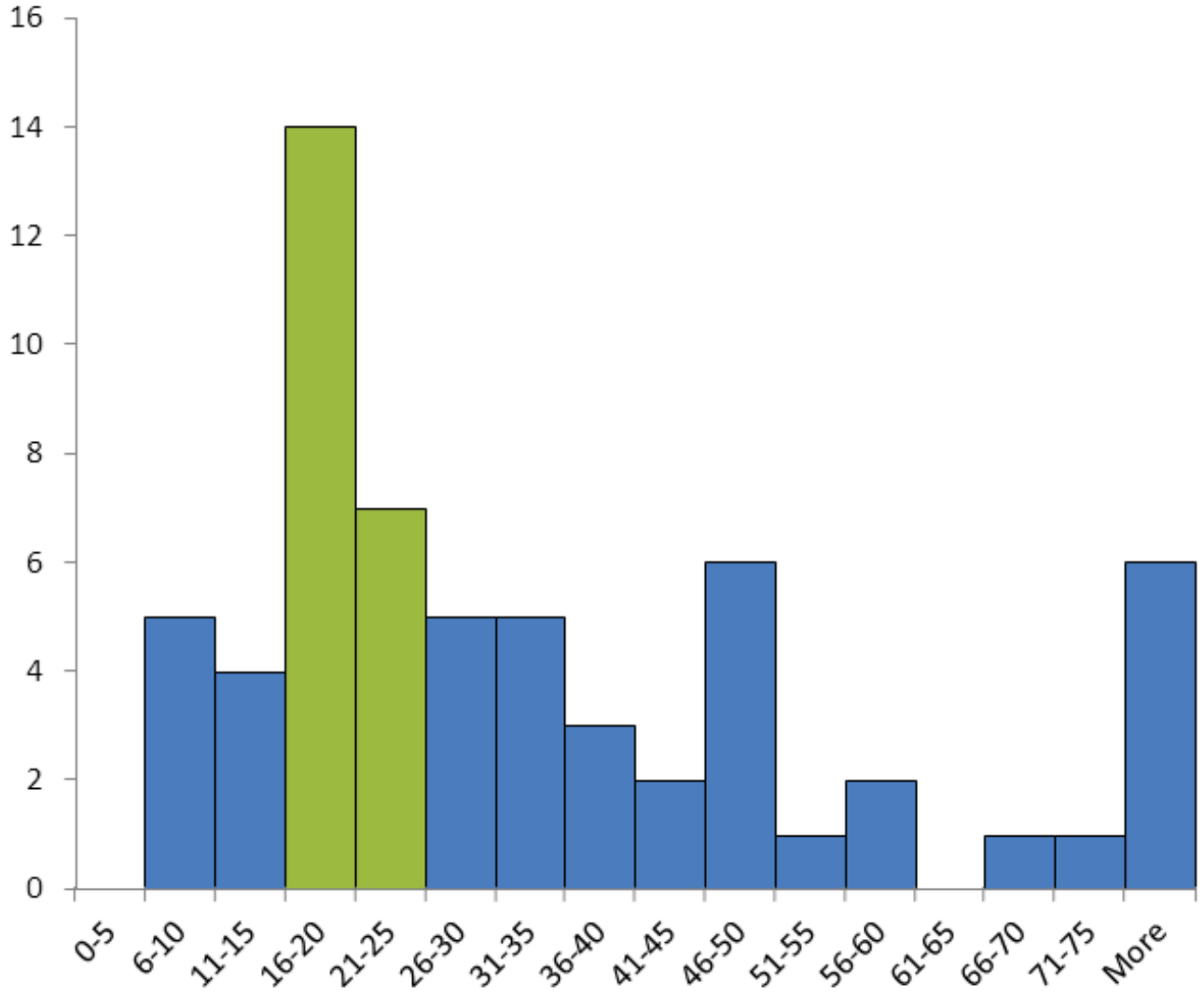


Image Credit: http://intelligence.org/files/PredictingAI.pdf

AI Predictions: Non-Experts

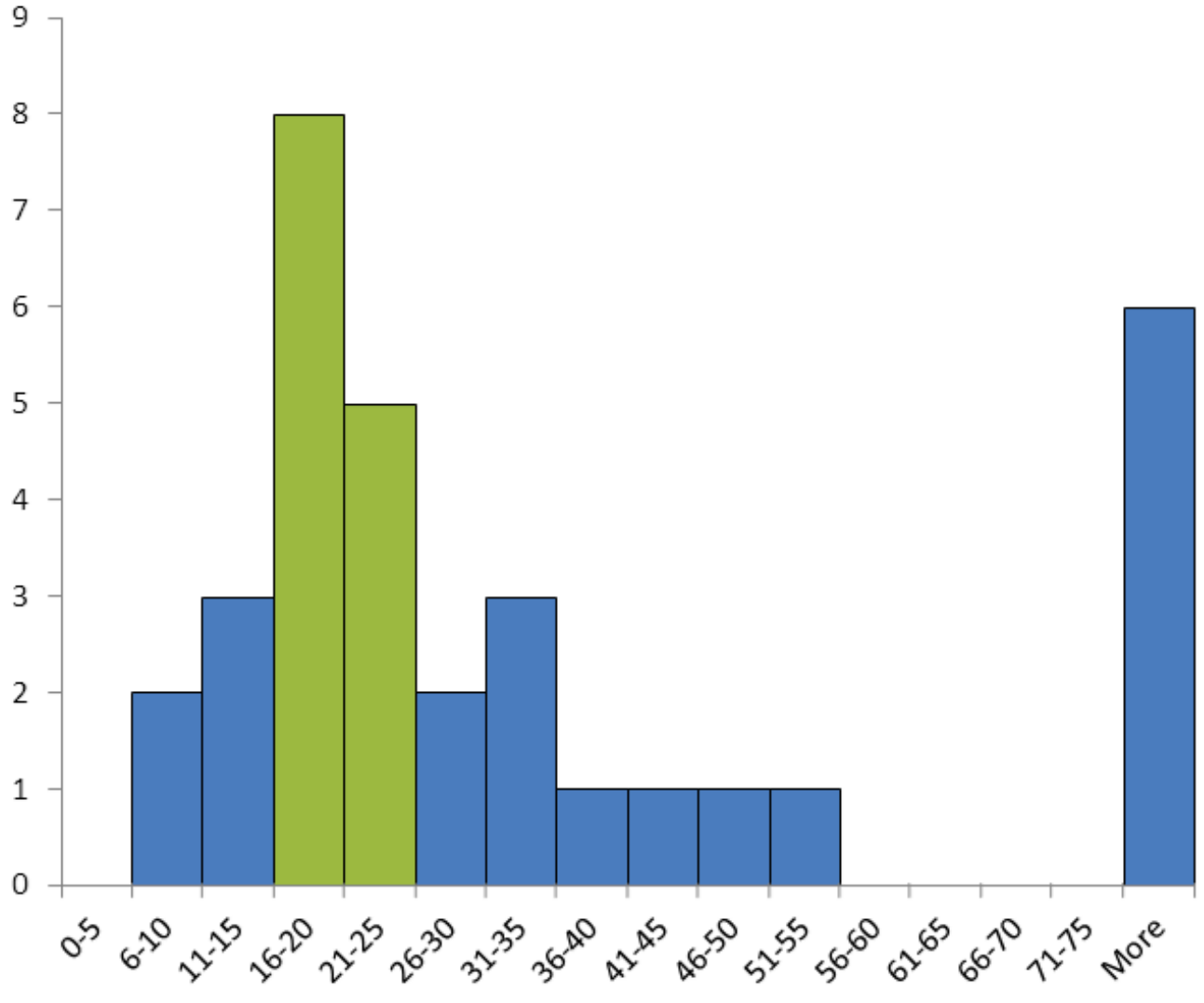


Image Credit: http://intelligence.org/files/PredictingAI.pdf

AI Predictions: Failed

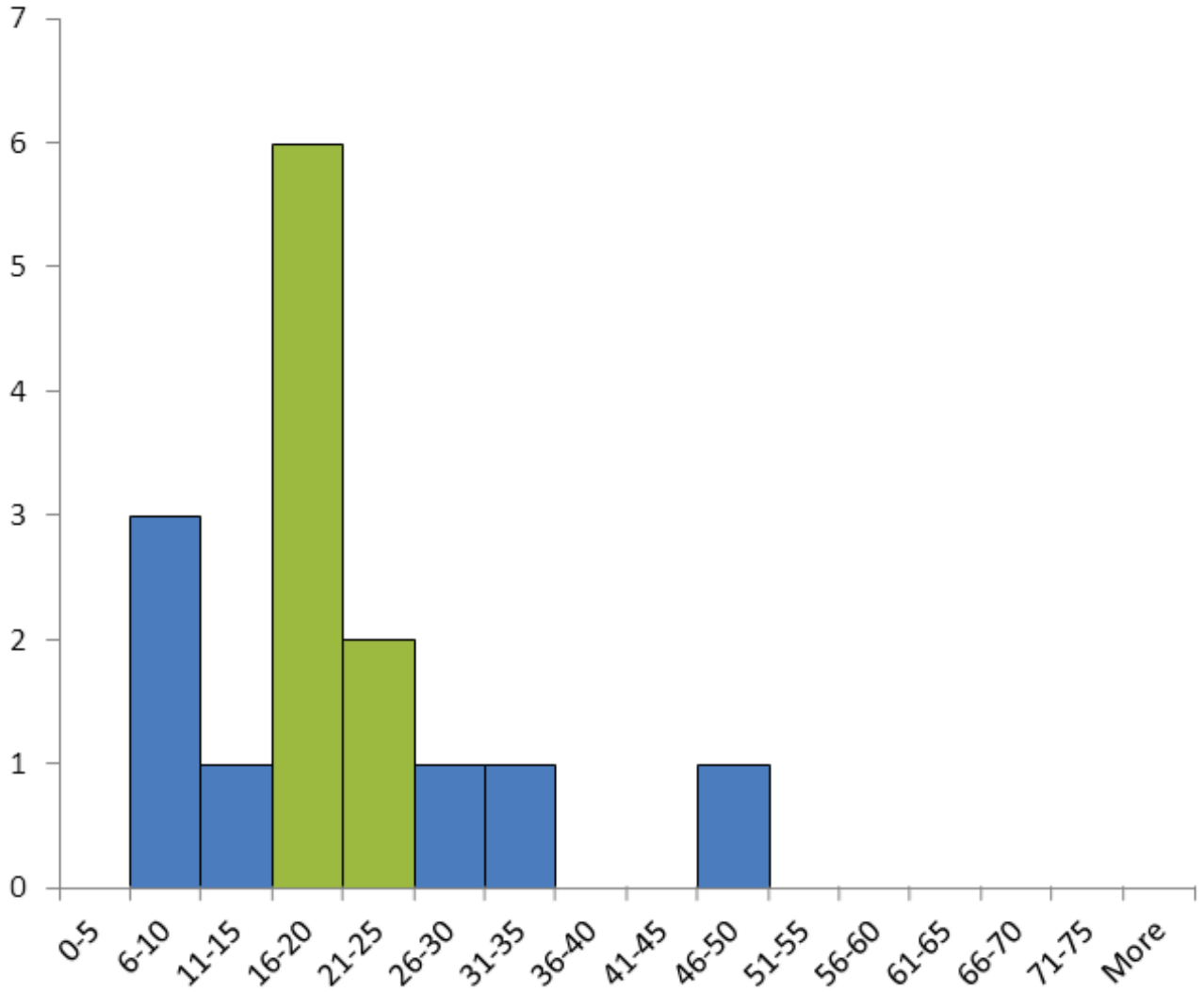
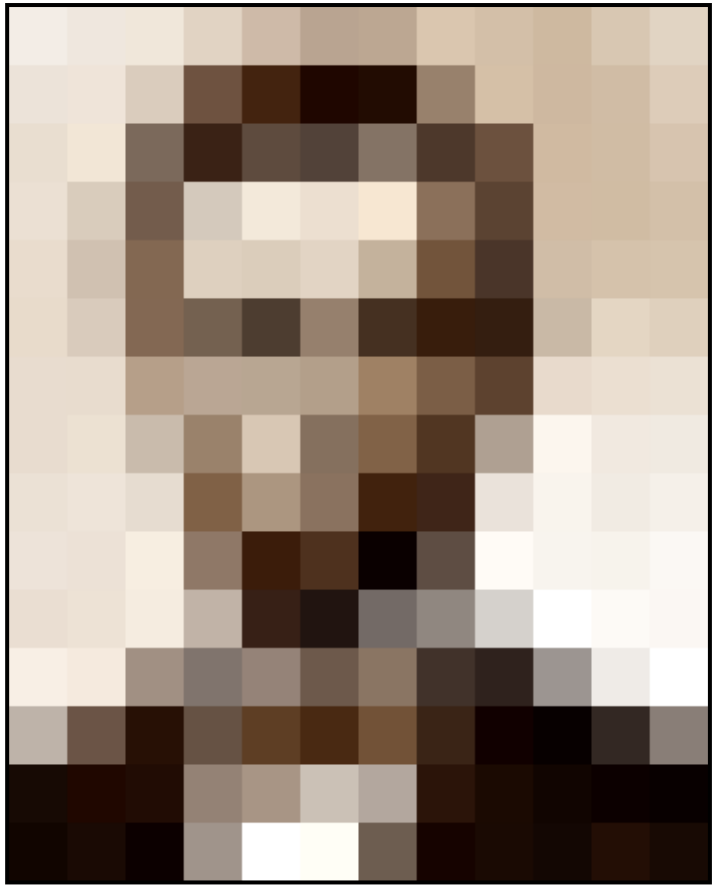


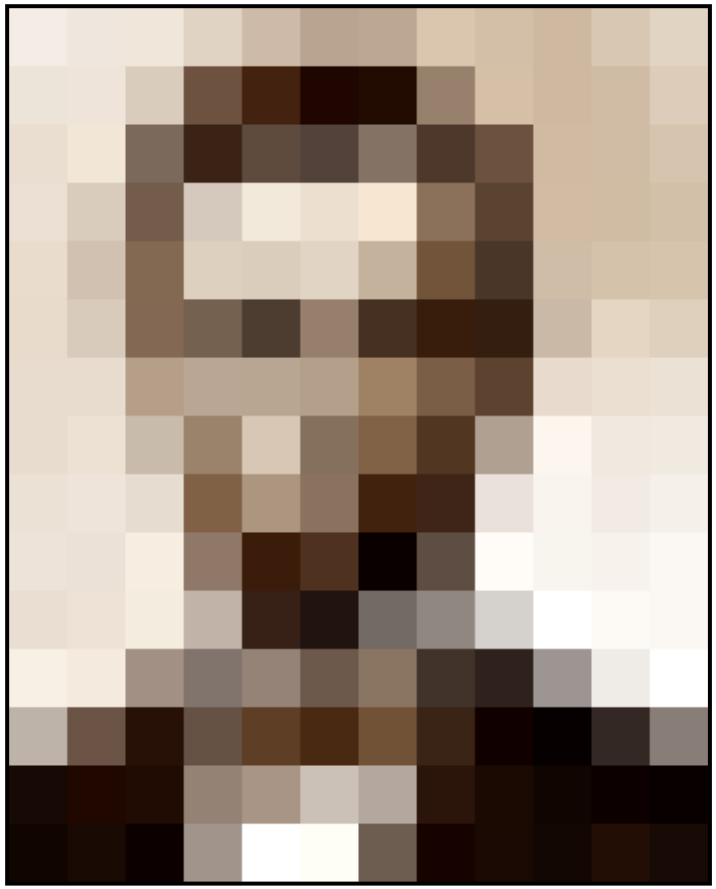
Image Credit: http://intelligence.org/files/PredictingAI.pdf

What humans see



Slide Credit: Larry Zitnick

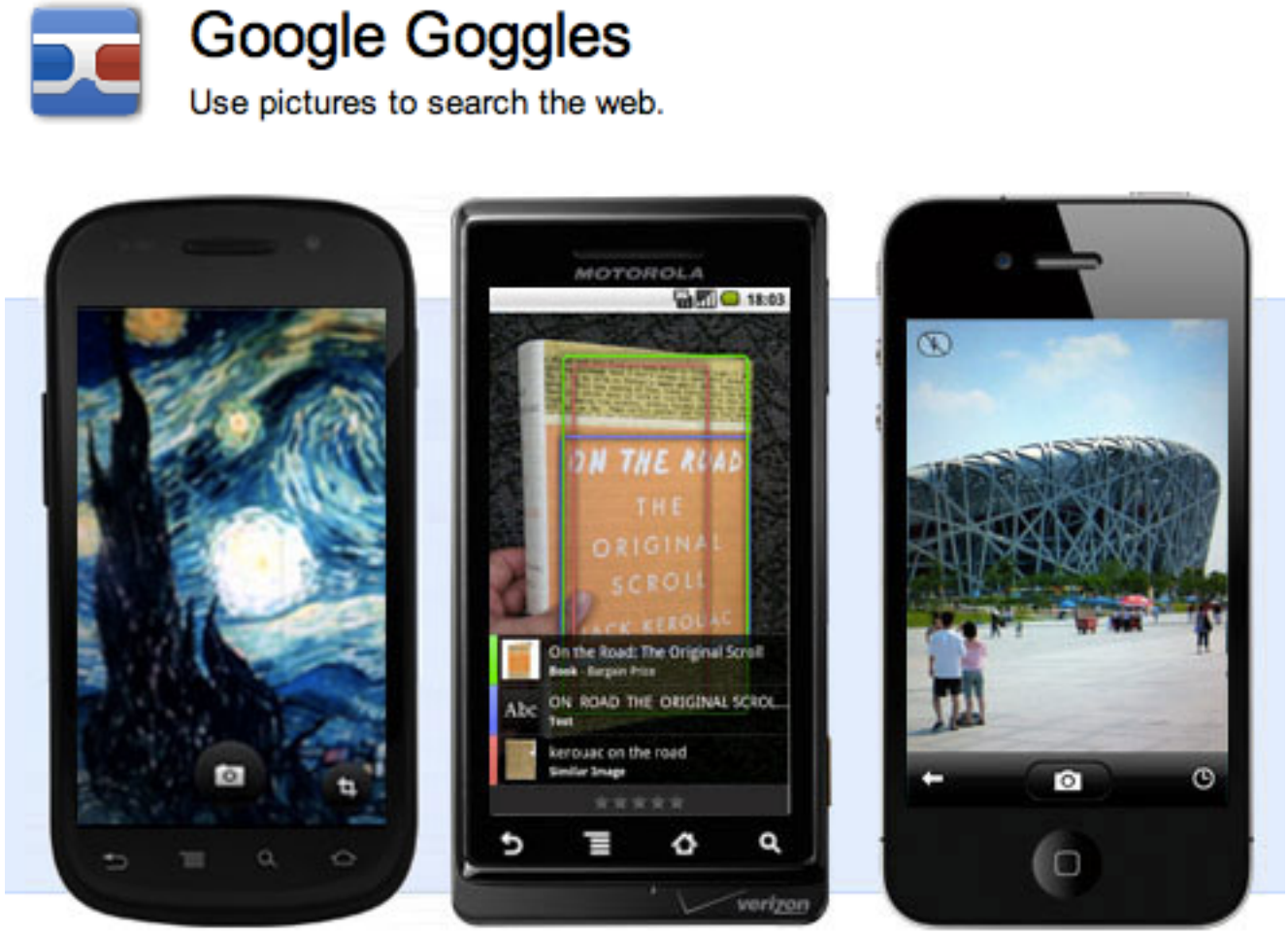
What computers see



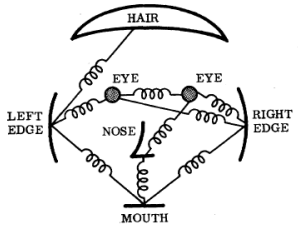
Slide Credit: Larry Zitnick

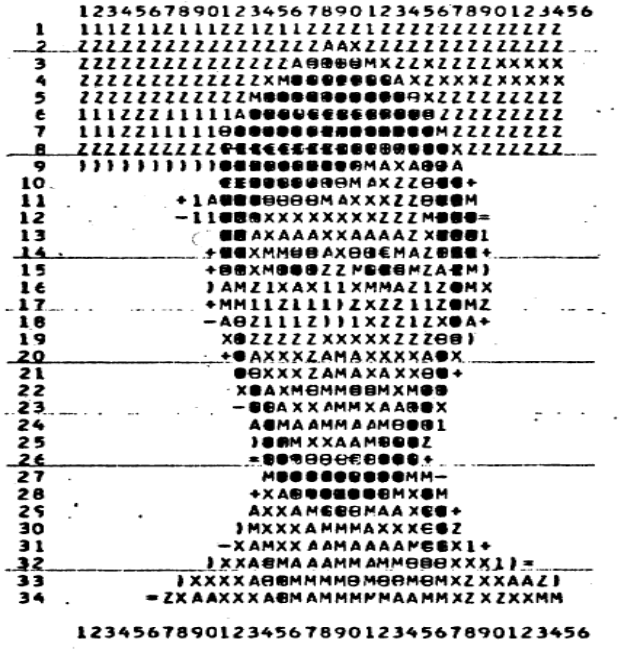
We've come a long way…

We've come a long way…



We've come a long way…





[Fischler and Elschlager, 1973]

We've come a long way…

Datasets and computer vision

UIUC Cars (2004)

S. Agarwal, A. Awan, D. Roth

CMU/VASC Faces (1998)

H. Rowley, S. Baluja, T. Kanade

FERET Faces (1998)

P. Phillips, H. Wechsler, J. Huang, P. Raus

COIL Objects (1996)

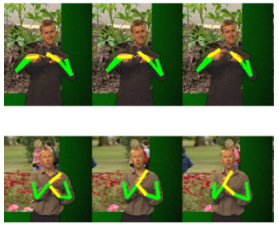
S. Nene, S. Nayar, H. Murase

MNIST digits (1998-10)

Y LeCun & C. Cortes

KTH human action (2004)

I. Leptev & B. Caputo

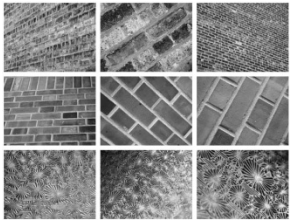


Sign Language (2008)

P. Buehler, M. Everingham, A. Zisserman

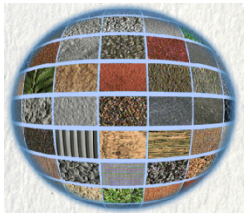
Segmentation (2001)

D. Martin, C. Fowlkes, D. Tal, J. Malik.



3D Textures (2005)

S. Lazebnik, C. Schmid, J. Ponce

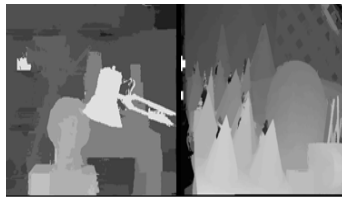


CuRRET Textures (1999)

K. Dana B. Van Ginneken S. Nayar J. Koenderink

CAVIAR Tracking (2005)

R. Fisher, J. Santos-Victor J. Crowley



Middlebury Stereo (2002)

D. Scharstein R. Szeliski

Backpack

Flute

Strawberry

Traffic light

Backpack

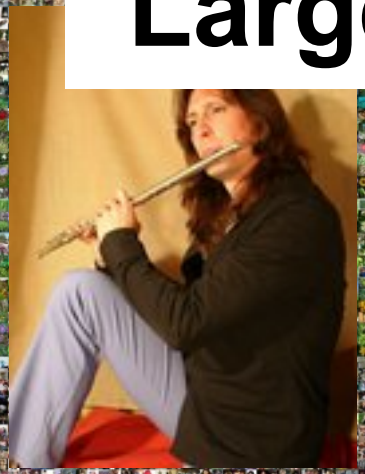
Matchstick

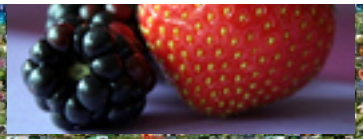
Bathing cap

Sea lion

Racket

Large-scale recognition





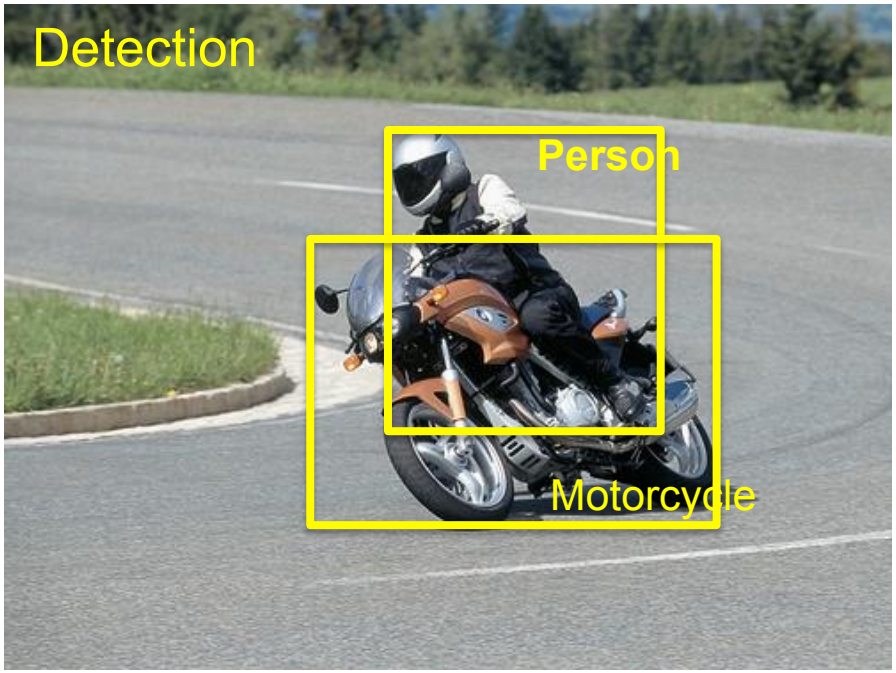
PASCAL VOC 2005-2012

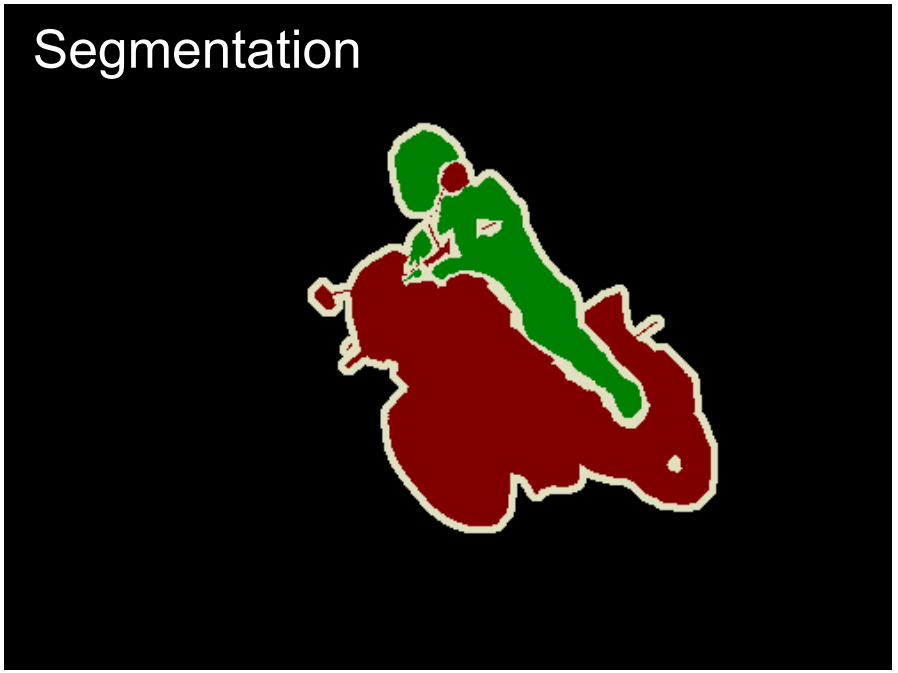
Everingham, Van Gool, Williams, Winn and Zisserman.

The PASCAL Visual Object Classes (VOC) Challenge. IJCV 2010.

20 object classes **22,591 images**

Classification: person, motorcycle





Action: riding bicycle

ImageNet Large Scale Visual Recognition Challenge (ILSVRC)

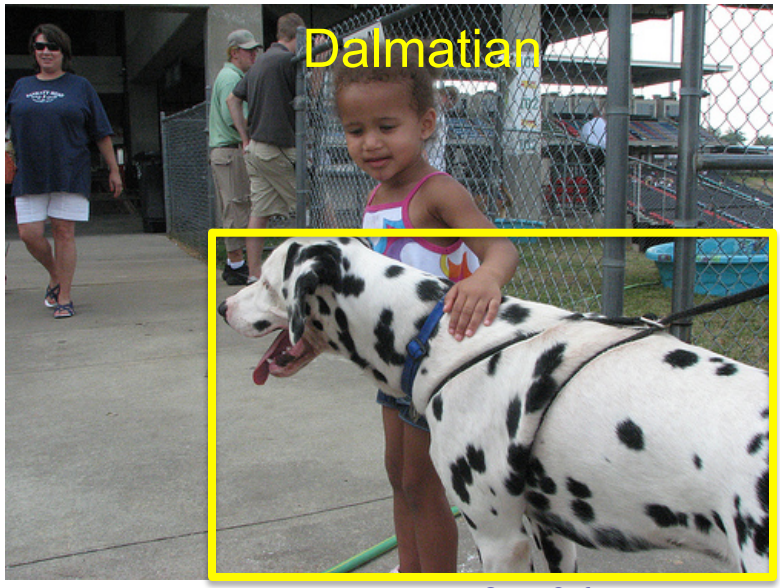
~~20 object classes 22,591 images~~

Classification:

1000 object classes 1.4M/50k/100k images

Detection:

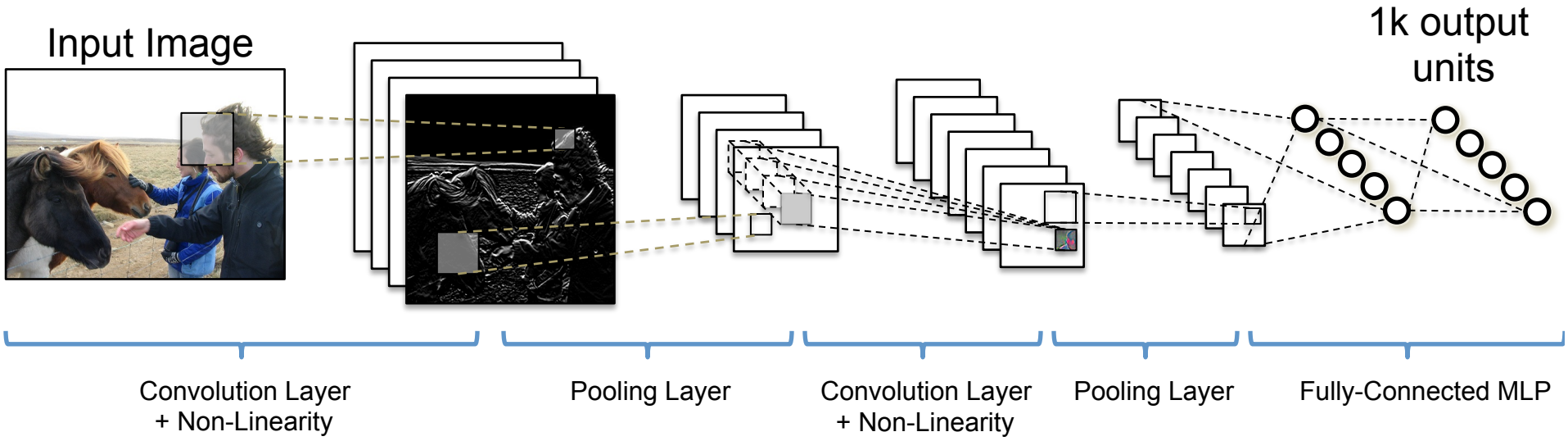
200 object classes 400k/20k/40k images



http://image-net.org/challenges/LSVRC/{2010,…,2014}

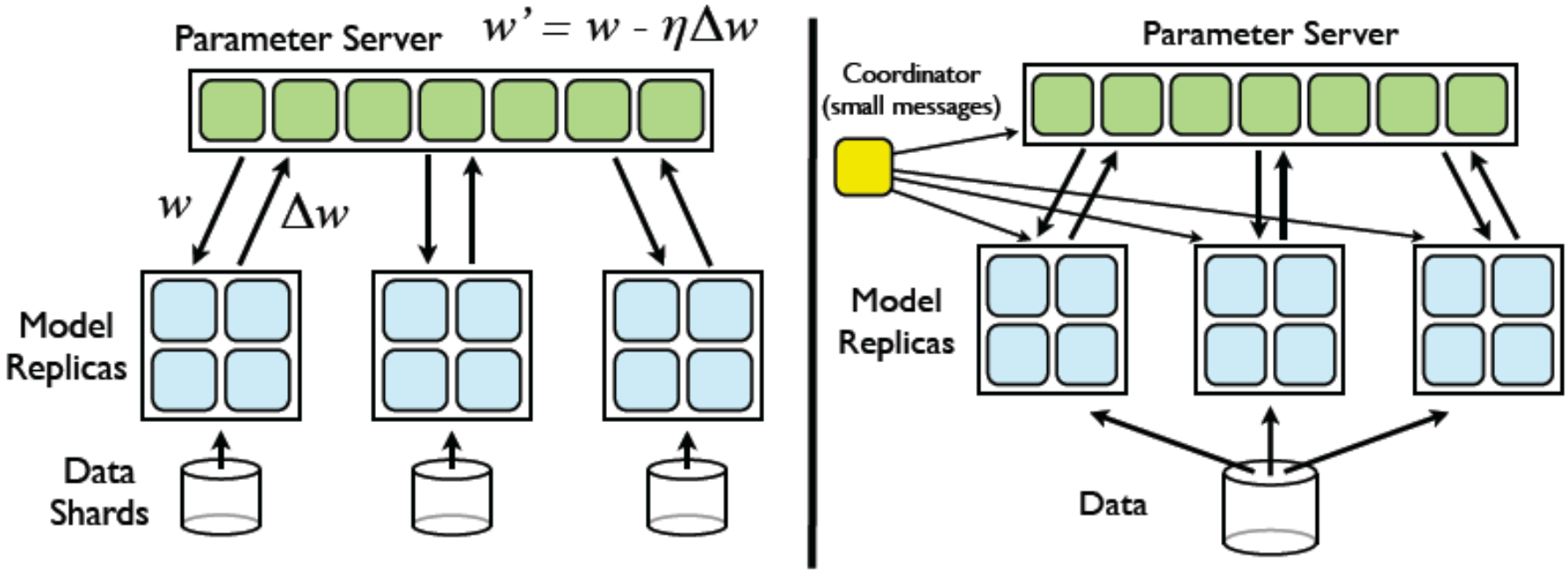
Data Enabling Richer Models

- [Krizhevsky et al. NIPS12, Donahue ICML14]
  - 54 million parameters
  - Trained on 1.4M images in ImageNet



Data Enabling Richer Models

- DistBelief [Dean et al. NIPS12]



Data Enabling Richer Models

- [Le et al. ICML12]
  - 2,000 machines / 32,000 cores for 1 week

- DistBelief [Dean et al. NIPS12]
  - 16 million images and 21k categories
  - 1.7 Billion parameters
  - 12,000 cores

Historical Perspective

- Challenges in computer vision research: future directions of research.
  Shahriar Negahdaripour and Anil K. Jain.
  NSF Workshop 1991

- Panel stressed the need for:
  - more experimental validation of models on large datasets
  - sharing of images, algorithms, and models between research groups
  - greater interaction between academia and industry
  - the need for complete computer vision systems that perform real world tasks

Back to Present

- Frontiers in Computer Vision.
  Alan Yuille and Aude Oliva.
  NSF Workshop Nov 2010

- Noticeable changes since 1991:
  - Computers are much faster, have far greater memory, and are much cheaper.
  - Computer vision researchers have continued to learn, adapt, develop, and apply tools from mathematics, statistics, computer science, and engineering.
  - New tools specific to vision (e.g., SIFT and HOG)
  - The use of benchmarked image databases and learning algorithms has become common

Back to Present

- Frontiers in Computer Vision.
  Alan Yuille and Aude Oliva.
  NSF Workshop Nov 2010

- Remaining concerns:
  - increased the fragmentation of the field
  - there remains lack of scholarship and little progress made on building on research done by others.
  - computer vision datasets do not compare yet to the complexity of the natural world
  - academic research is seen as being neither realistic enough to help develop practical real world systems nor insightful enough to yield new theories

Challenges

- Big data is an enabler and an isolator!
- All researchers repeatedly solving the same problems

  - Build and maintain a cluster
    - Job scheduler (PBS, Torque)
    - Distributed storage (Hadoop FS)

    → Logistical

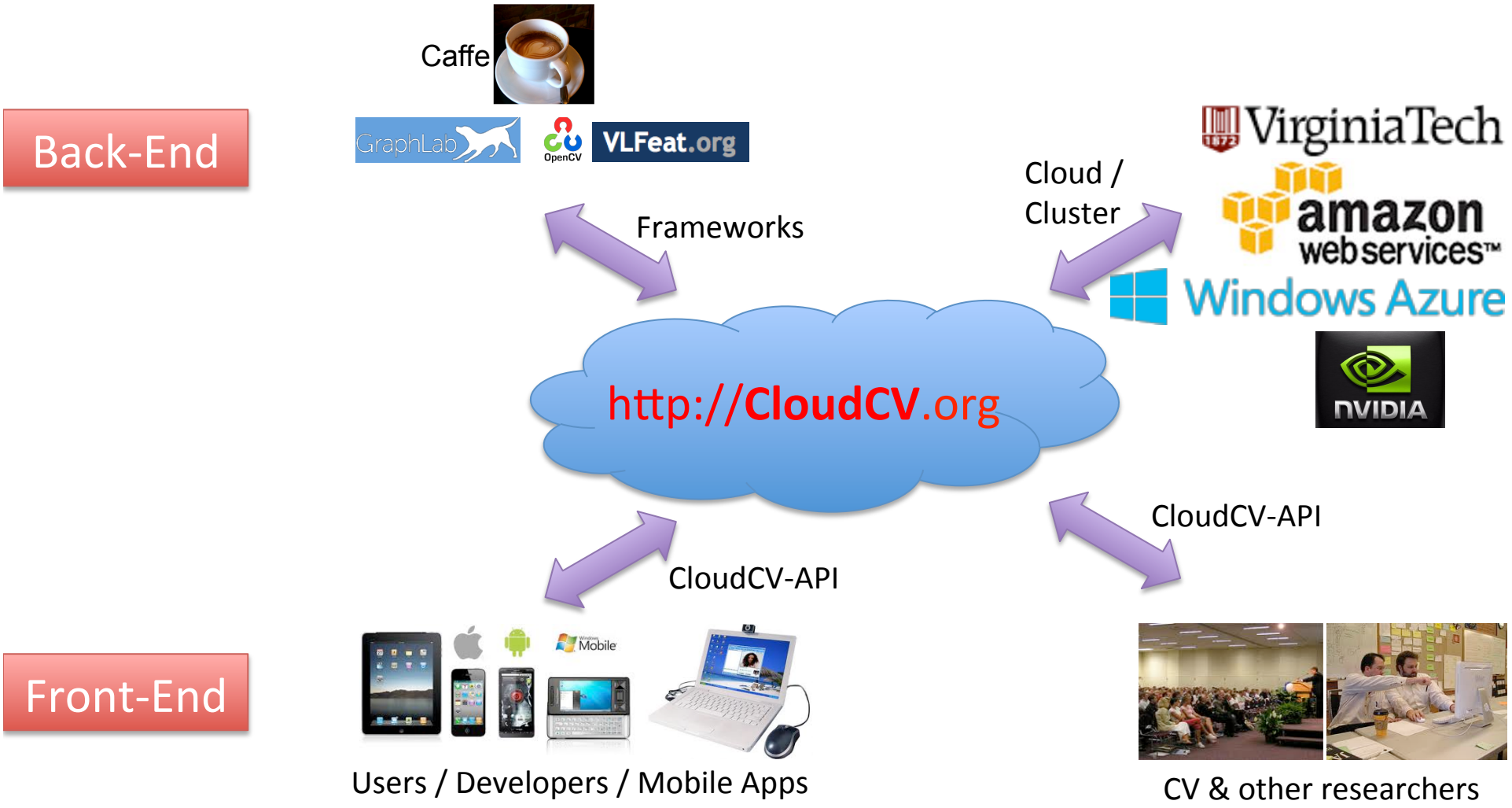
  - Scale vision algorithms
    - Identify model/data parallelism
    - Design & implement multi-threaded vision primitives

    → Computer Vision

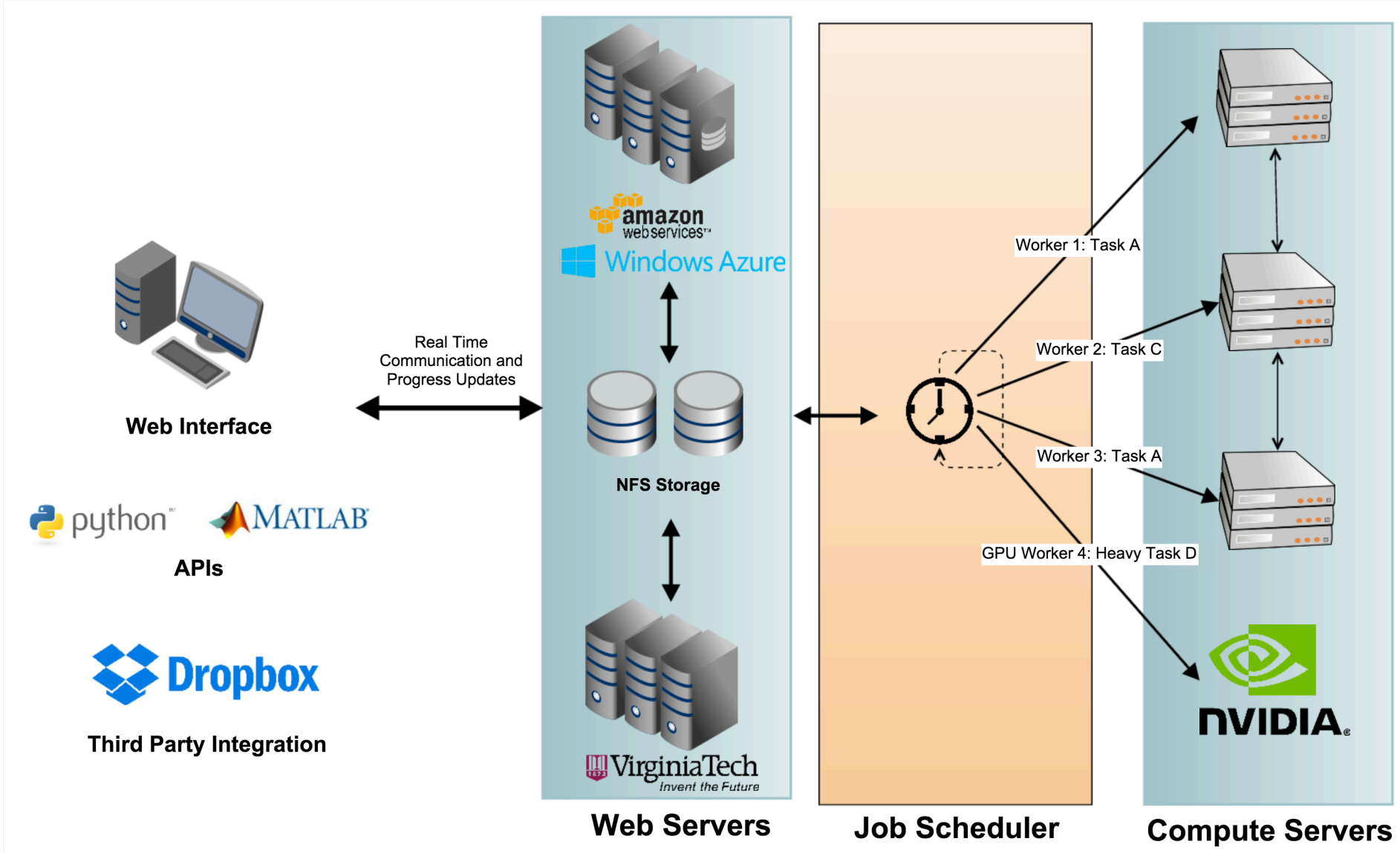
  - Distributed computing
    - Implement mechanisms to avoid race conditions & dead-locks
    - Ensure data consistency, locking, good scheduling

    → Distributed Computing

CloudCV


Back-End
Caffe
GraphLab
OpenCV
VLFeat.org
Frameworks
Cloud /
Cluster
VirginiaTech
amazon
web services™
Windows Azure
NVIDIA
http://CloudCV.org
CloudCV-API
CloudCV-API
Front-End
Users / Developers / Mobile Apps
CV & other researchers

CloudCV: Architecture



CloudCV: Big Picture

- Goal: For developers
  - Reduced barrier to entry
  - Democratize Computer Vision
- Goal: For researchers
  - Easy comparison to baselines
  - Access to state-of-art techniques “off-the-shelf”
- Mini-steps
  - What we have today
  - A few algorithms
  - A few ways to reach CloudCV
  - Where we are headed

CloudCV

- Demo 1
  - Support for ImageNet Challenge
- Demo 2
  - Image Classification
- Demo 3
  - Training a new classifier for your categories
- Demo 4
  - Finding Important People in Images
- Demo 5
  - GigaPixel Image Stitching

“Demo” 1

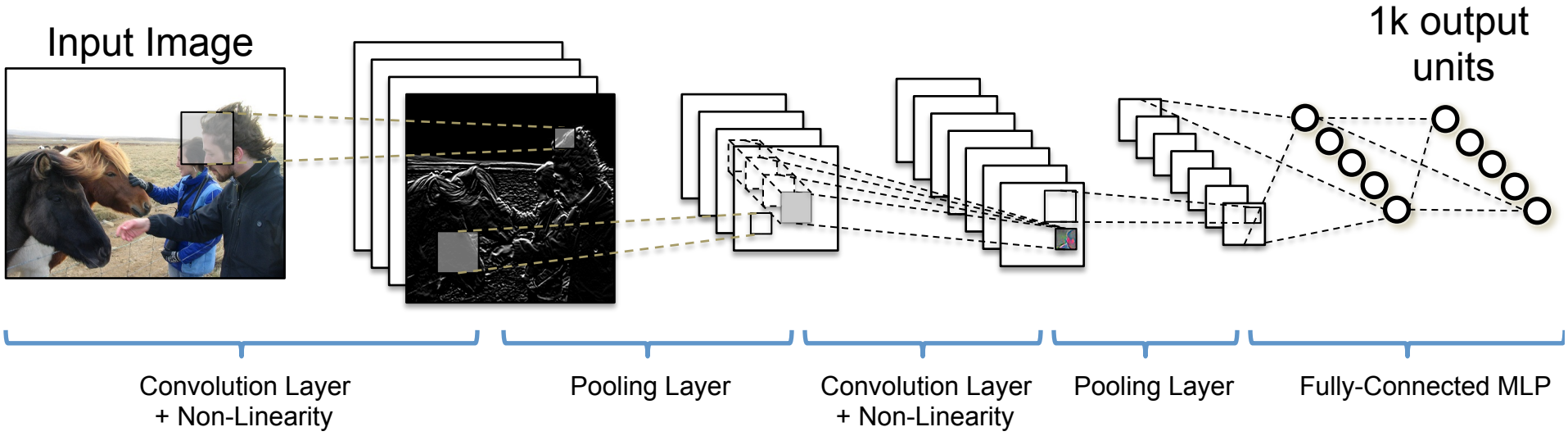
- ImageNet Challenge (ILSVRC13)
  - Training: 1.4 million
  - Val: 50k
  - Test: 100k
- Features
  - 16 “industry standard”
    - DeCAF, GIST, HOG2x2, Dense/Sparse SIFT, LBP, Self-Similarity …
- Webpage
  - http://cloudcv.org/objdetect/#features
- Total: 400 GB, *19 months or 1.5 years of CPU-time*

CloudCV

- Demo 1
  - Support for ImageNet Challenge
- Demo 2
  - Image Classification
- Demo 3
  - Training a new classifier for your categories
- Demo 4
  - Finding Important People in Images
- Demo 5
  - GigaPixel Image Stitching

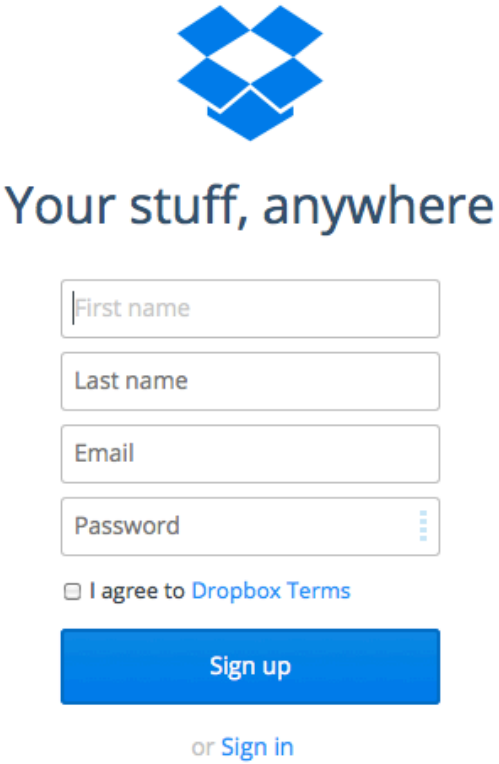
Demo 2

- [Krizhevsky et al. NIPS12, Donahue ICML14]
  - Trained on 1.4M images in ImageNet
  - 1000 categories
  - Available in Caffe framework from BVLC
  - http://cloudcv.org/classify/



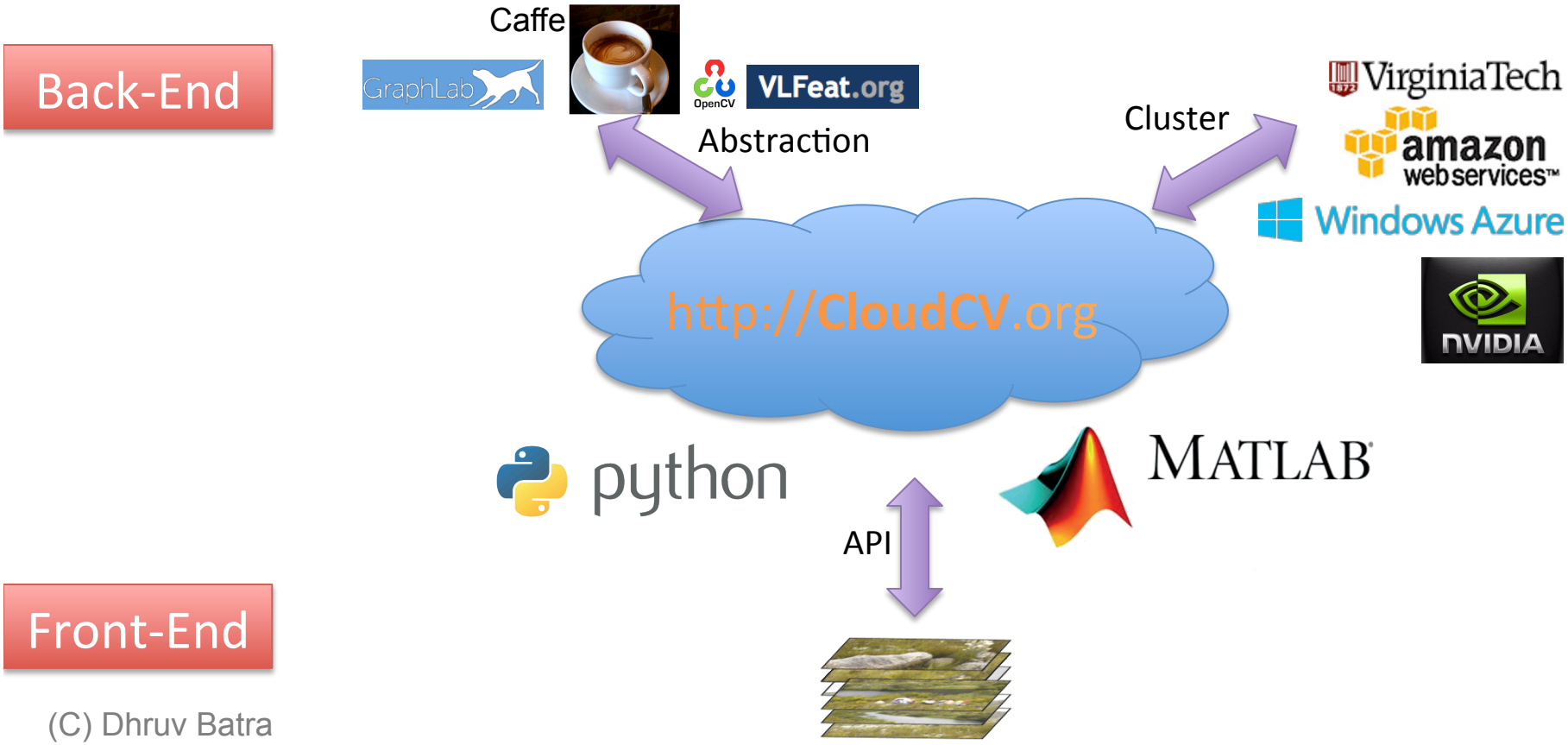
Demo 2

- Drop-box integration
  - Files can live on dropbox
  - http://cloudcv.org/decaf-server/



Demo 2

- How about if you want to write code?
  - Python-API: https://github.com/batra-mlp-lab/pcloudcv
    - "python run.py myconfig.json –nologin"
  - Matlab-API: https://github.com/batra-mlp-lab/mcloudcv



CloudCV

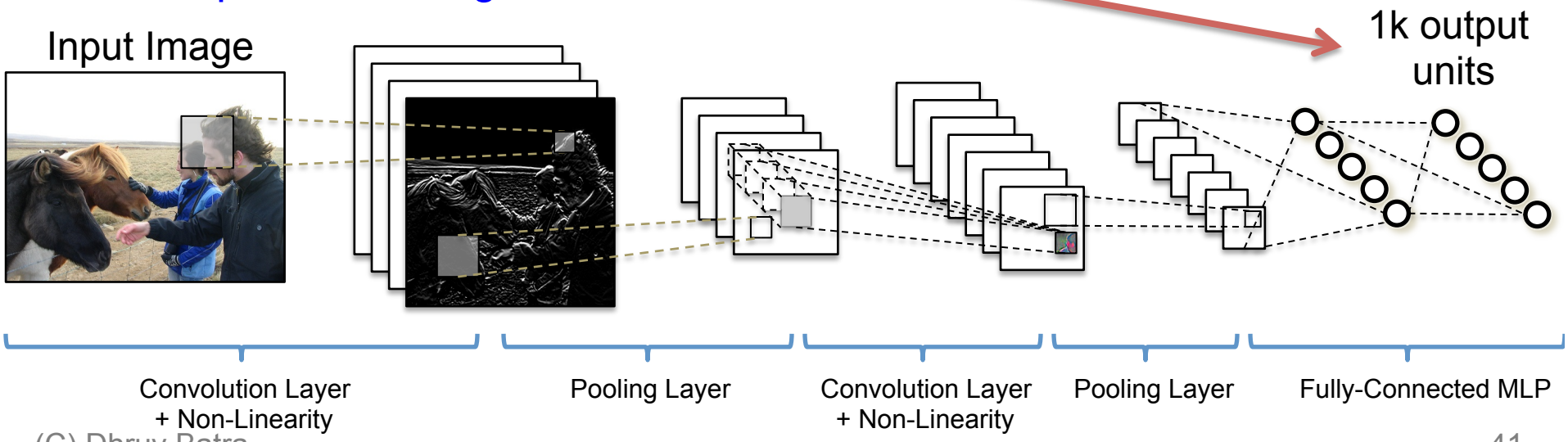
- Demo 1
  - Support for ImageNet Challenge
- Demo 2
  - Image Classification
- Demo 3
  - Training a new classifier for your categories
- Demo 4
  - Finding Important People in Images
- Demo 5
  - GigaPixel Image Stitching

Demo 3

- [Krizhevsky et al. NIPS12, Donahue ICML14]
  - Trained on 1.4M images in ImageNet
  - 1000 categories
  - Available in Caffe framework from BVLC

How about adding a 1001th category?
Your company logo classifier?
In a few seconds, not weeks?

http://cloudcv.org/trainaclass/



CloudCV

- Demo 1
  - Support for ImageNet Challenge
- Demo 2
  - Image Classification
- Demo 3
  - Training a new classifier for your categories
- Demo 4
  - Finding Important People in Images
- Demo 5
  - GigaPixel Image Stitching

Who is the most important person in the photo?

Why is this useful?

- Better image descriptions
- Automatic photo cropping

Two people walking past a crowd

Why is this useful?

- Better image descriptions
- Automatic photo cropping
- Sort consumer photos

How do we do this?

- Collect a large dataset
  - VT Person Importance Dataset
  - Images scraped from Flickr
  - Annotations using Mechanical Turk
- For each face measure:
  - Distance from center
  - Scale
  - Sharpness
  - Face Pose
  - Face Occlusion
- Train a relative importance predictor

Results

- http://cloudcv.org/vip/

Method	Accuracy
Our Approach	**78.91%**
Center Baseline	68.46%
Scale Baseline	67.86%
Sharpness Baseline	71.03%

- Technical Details:
  - VIP: Finding Important People in Images
  - Clint S. Mathialagan, Andrew C. Gallagher, Dhruv Batra
  - http://arxiv.org/abs/1502.05678

CloudCV

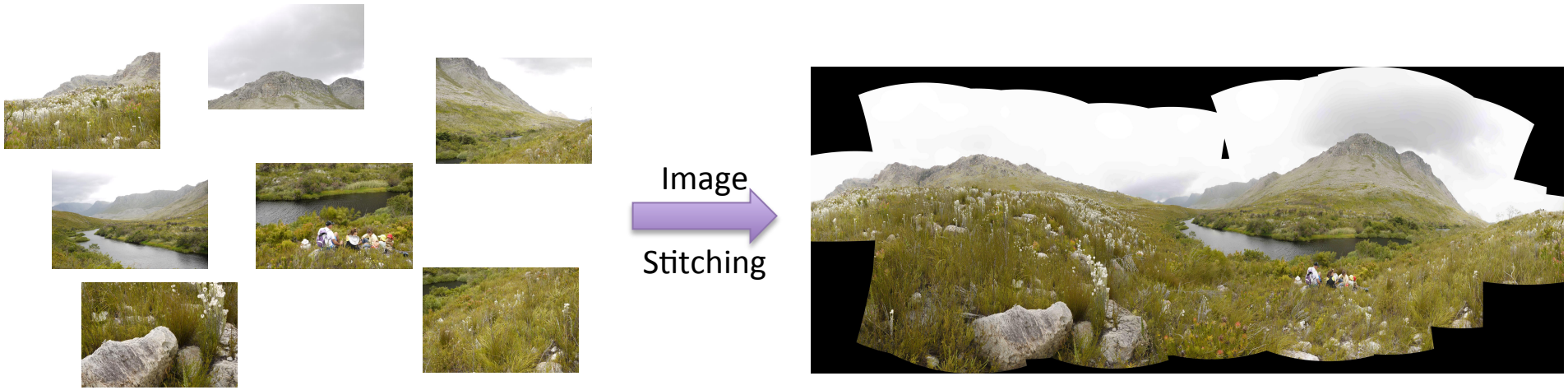
- Demo 1
  - Support for ImageNet Challenge
- Demo 2
  - Image Classification
- Demo 3
  - Training a new classifier for your categories
- Demo 4
  - Finding Important People in Images
- Demo 5
  - GigaPixel Image Stitching

Parallelization

- Some steps in vision embarrassingly parallel
  - Ideal for MapReduce

- However
  - Most pipelines in Computer Vision are not!
  - Example
    - Image Stitching

GigaPixel Image Stitching

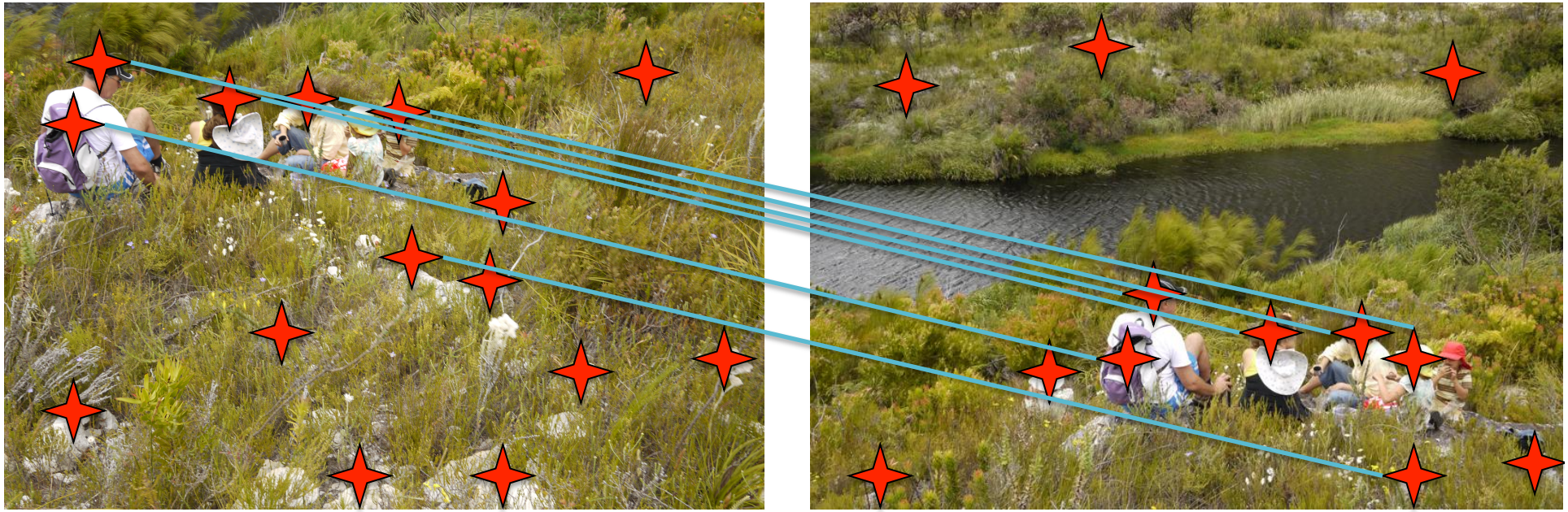


GigaPixel Image Stitching

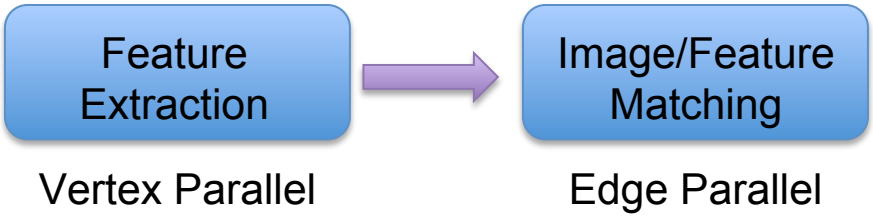
GigaPixel Image Stitching

Feature Extraction

Vertex Parallel

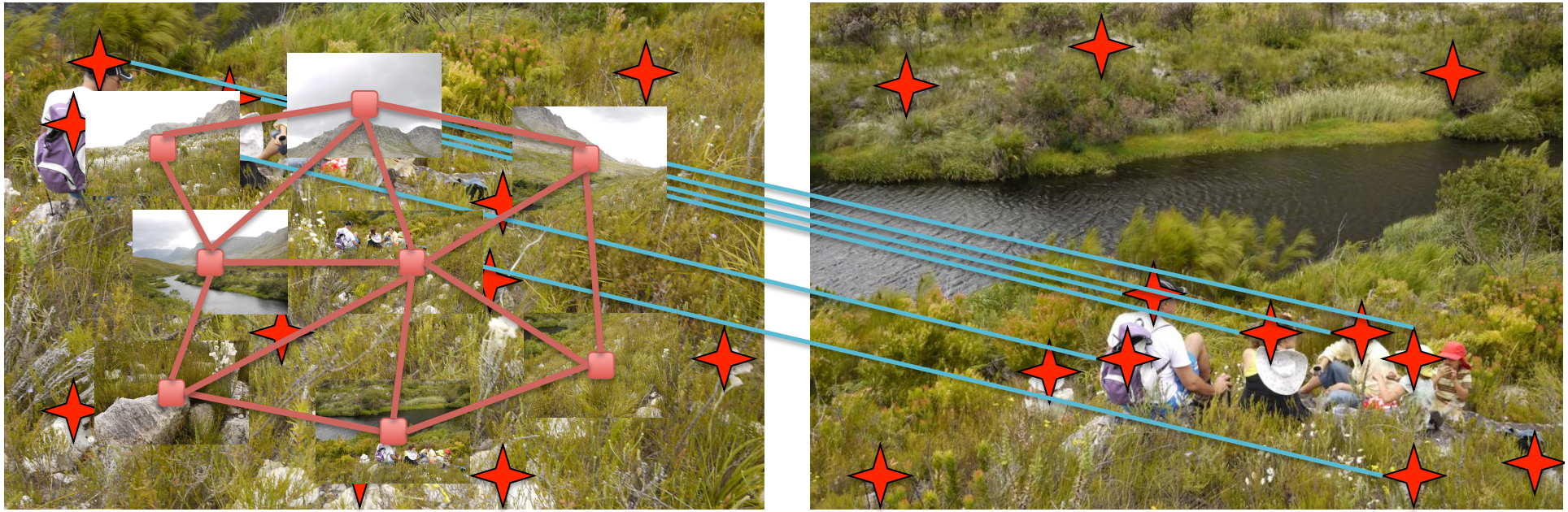


GigaPixel Image Stitching

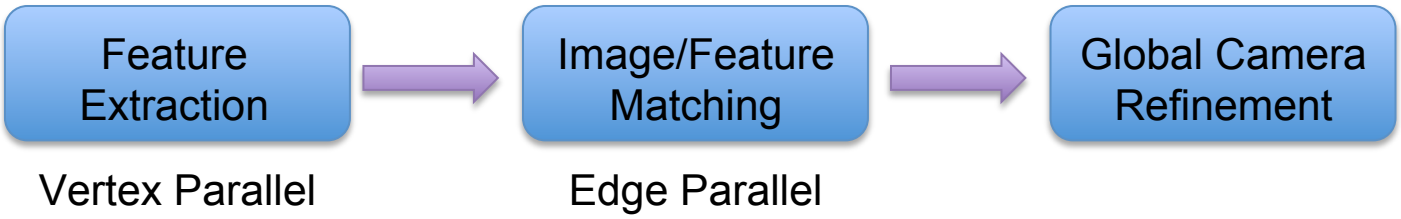


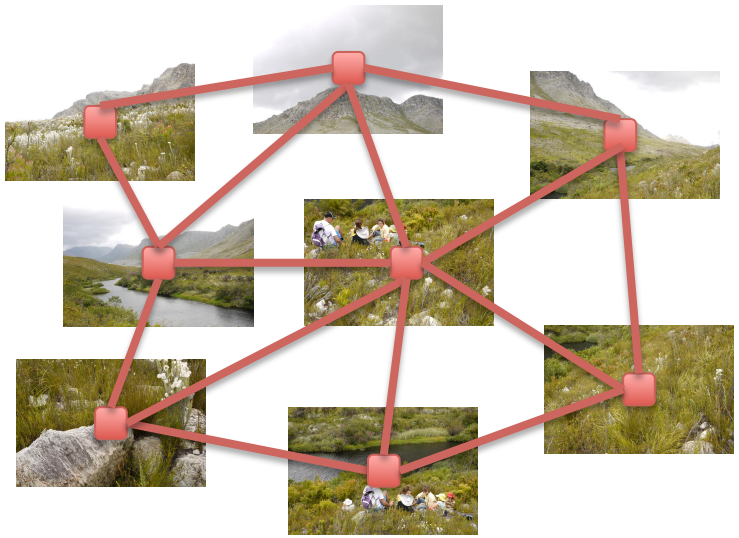
Vertex Parallel

Edge Parallel



GigaPixel Image Stitching



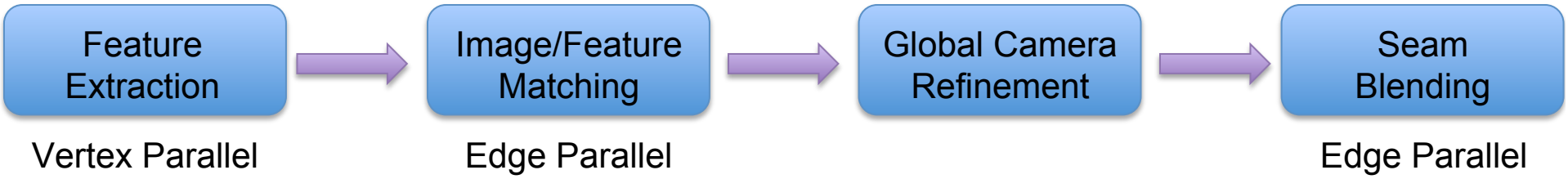


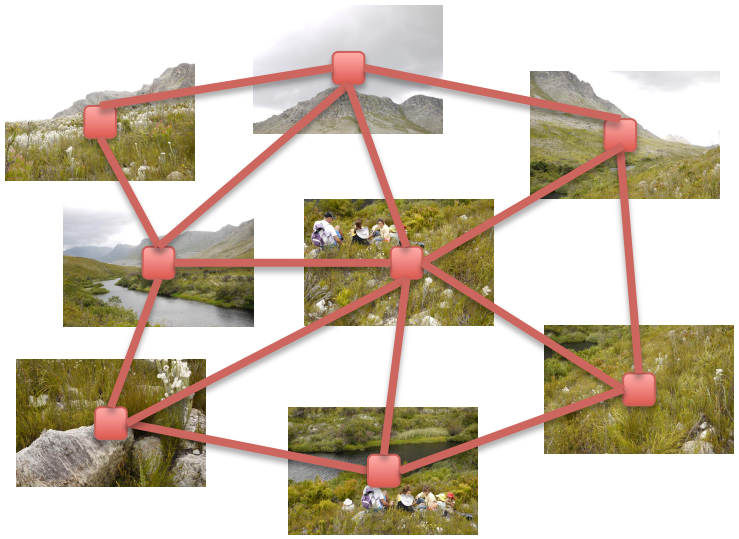
Bundle Adjustment

$$\min_{\hat{P}_i, \hat{X}_p} \sum_{\text{image } i} \sum_{\text{point } p} d(x_{ip}, \hat{P}_i \hat{X}_p)$$

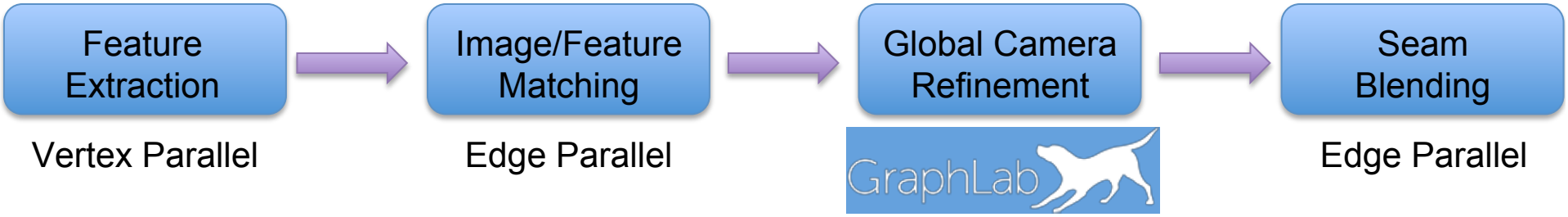
Non-linear optimization
over camera parameters P_i
and 3D locations of points X_p

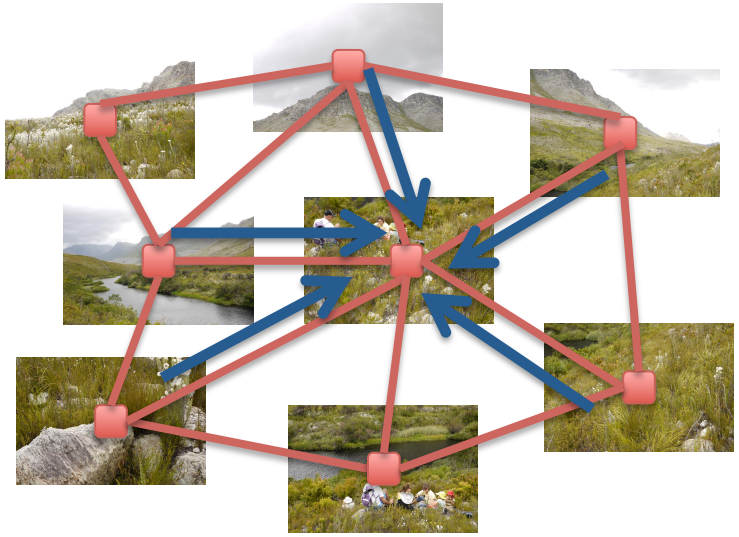
GigaPixel Image Stitching





GigaPixel Image Stitching





Bundle Adjustment

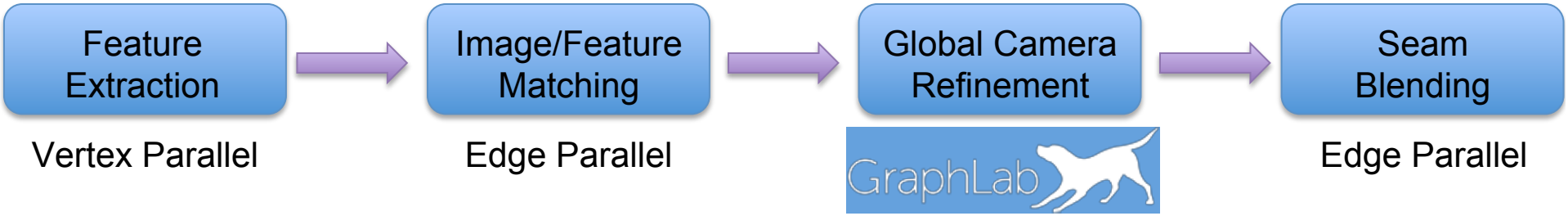
$$\min_{\hat{P}_i, \hat{X}_p} \sum_{\text{image } i} \sum_{\text{point } p} d(x_{ip}, \hat{P}_i \hat{X}_p)$$

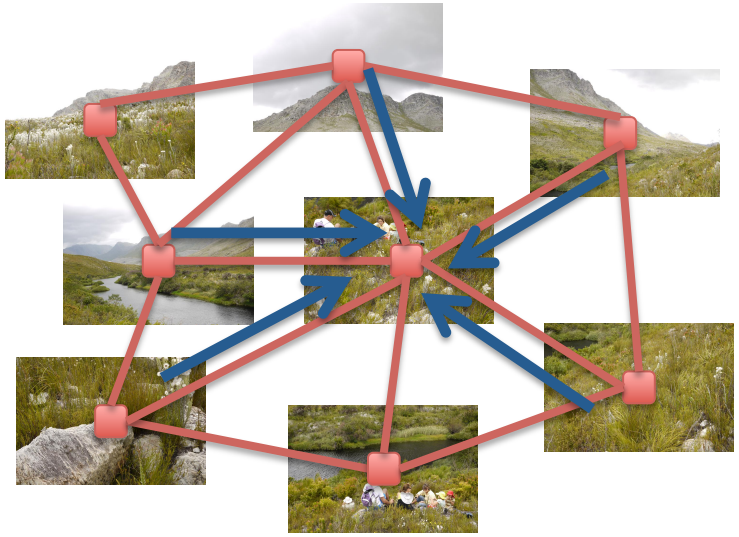
Levenberg–Marquard Updates

Graph-Parallel

GigaPixel Image Stitching

- http://cloudcv.org/image-stitch/





Bundle Adjustment

$$\min_{\hat{P}_i, \hat{X}_p} \sum_{\text{image } i} \sum_{\text{point } p} d(x_{ip}, \hat{P}_i \hat{X}_p)$$

Levenberg–Marquard Updates

Graph-Parallel

CloudCV

- Demo 1
  - Support for ImageNet Challenge
- Demo 2
  - Image Classification
- Demo 3
  - Training a new classifier for your categories
- Demo 4
  - Finding Important People in Images
- Demo 5
  - GigaPixel Image Stitching

Where is CloudCV headed?

- Back-end
  - Open model for contributing code
- Dynamic Database
  - If “familiar” image, we can get you results without computing
  - If new image, we’ll cache the results for the next person
- Lots of challenges unsolved
  - Bandwidth, optimal compression
  - Computation on front end vs back end
  - Compressions on front end that bound performance?
    - Coresets, summarization, etc

Where is CloudCV headed?

- Long way to go
- But we think this is exciting!
- Think about the first APIs for
  - Designing webpages
  - User authentication, Credit-card processing
  - Search, Maps, Twitter feeds, …
- We want to do that for the scientific research and development community.

Acknowledgements

- Collaborator and Mentor
  - Carlos Guestrin (UW / Graphlab / Dato)

- Sponsors

Virginia Tech Institute for Critical Technology and Applied Science

NVIDIA

amazon web services™

Windows Azure

CloudCV Team:

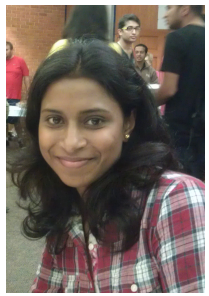
Harsh Agrawal

Clint Solomon

Neelima Chavali

Yash Goyal

Prakriti Banik



Thanks!