Identifying Animals in the Wild:

Final Presentation

Chia Kang Chao (cc3613)

Yen-Cheng Chou (yc2901)

5/8/2014

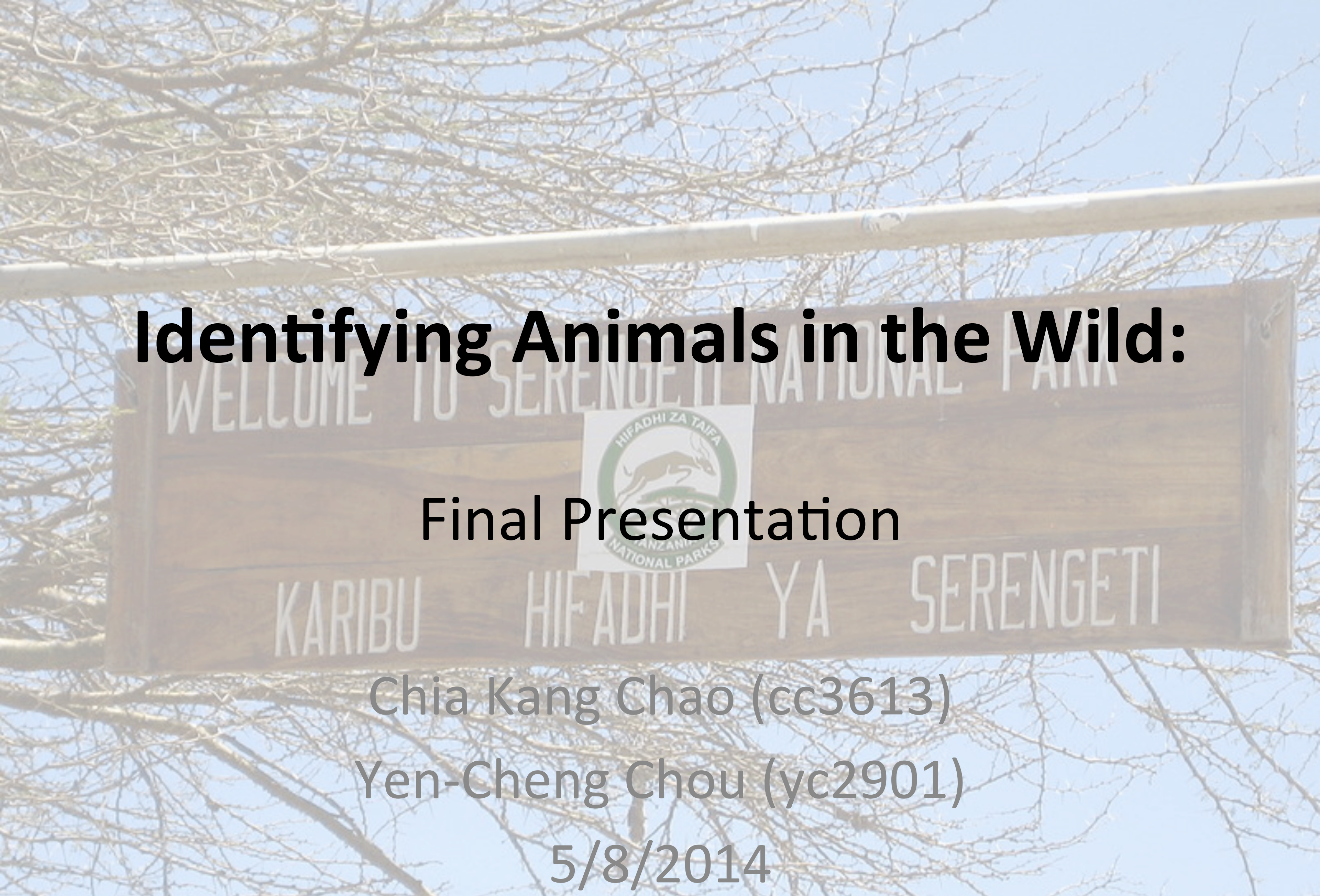


Table of Content

- Background
- Problem Identification
- Dataset
- Approach
- Experimental Result
- Analysis
- Conclusion

Table of Content

Serengeti National Park

- Name derived from *Serengit,* which means *Endless Plain* in Maasai Language
- ~ 70 large animals & ~ 500 bird species
- Host the largest terrestrial mammal migration in the world

Snapshot Serengeti Project

- Classify images collected by camera traps
- 50 animal species
- over 1 million images

Aardvark	Giraffe	Porcupine
Aardwolf	Guinea fowl	Reedbuck
Baboon	Hare	Reptiles
Bat-eared fox	Hartebeest	Rhinoceros
Bird (other)	Hippopotamus	Rodents
Buffalo	Honey-badger	Secretary bird
Bushbuck	Hyena (spotted)	Serval
Caracal	Hyena (striped)	Topi
Cheetah	Impala	Vervet monkey
Civet	Jackal	Warthog
Dik dik	Kori bustard	Waterbuck
Eland	Leopard	Wildcat
Elephant	Lion (female or cub)	Wildebeest
Gazelle (Grant's)	Lion (male)	Zebra
Gazelle (Thomson's)	Mongoose	Zorilla
Genet	Ostrich	Human

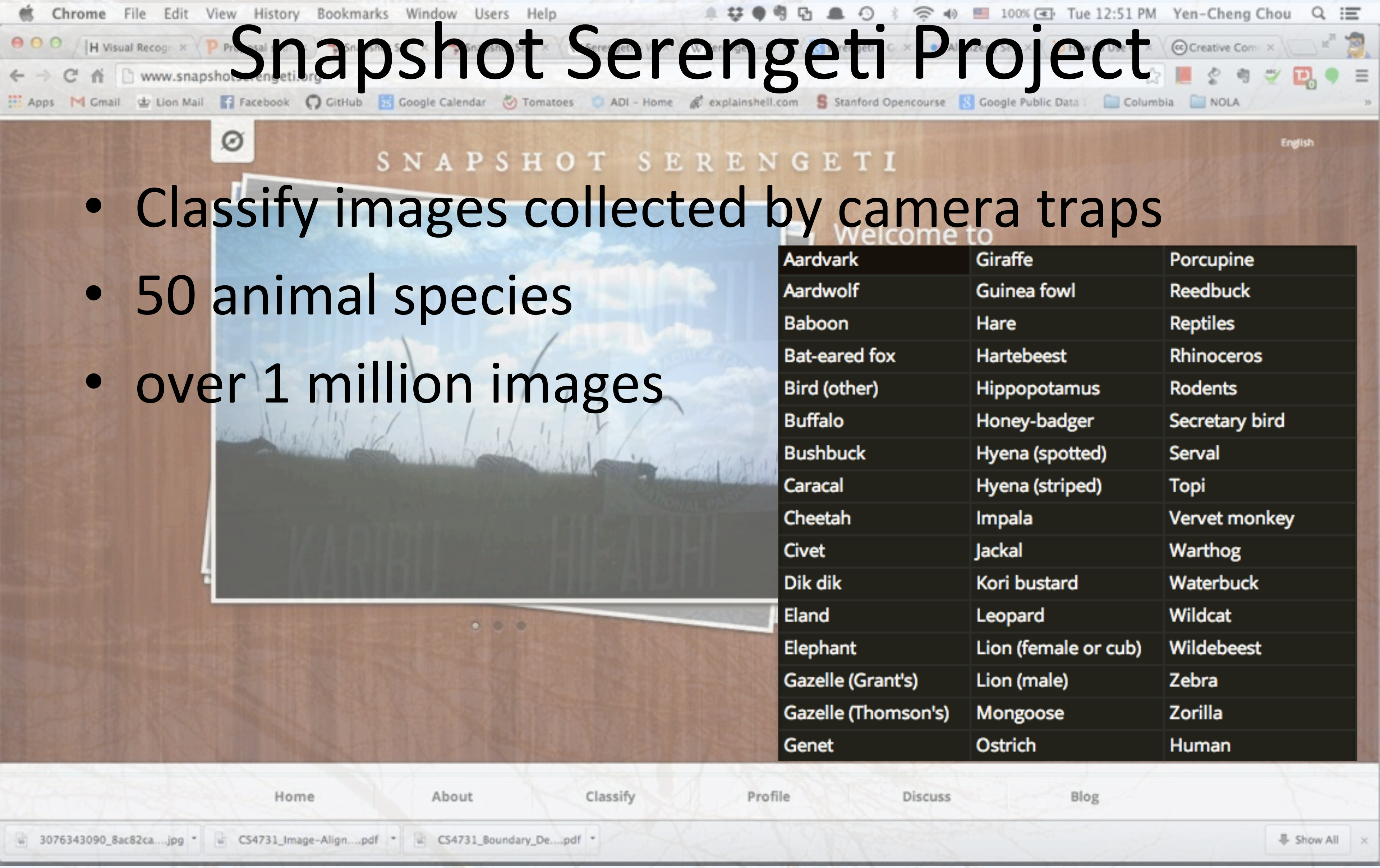


Image cropped on [Snapshot Serengeti]

Table of Content

- Background
- Problem Identification
- Dataset
- Approach
- Experimental Result
- Analysis
- Conclusion

Data is WILD

- Lighting
- Animal size in the photo
- Species look similar

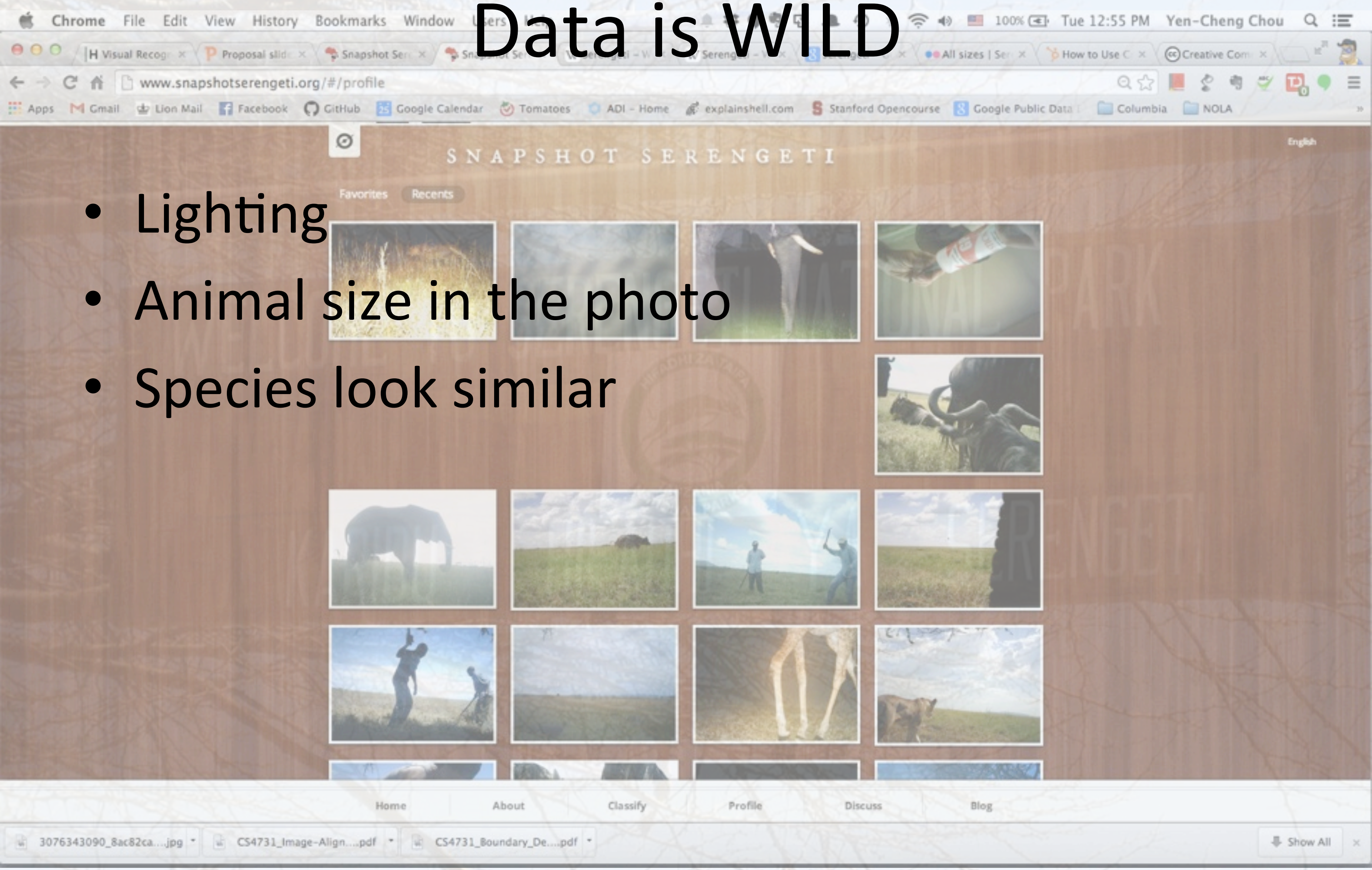


Image cropped on [Snapshot Serengeti]

Lighting

Image cropped on Snapshot Serengeti

Animal size in the photo




Image cropped on [Snapshot Serengeti]

Species look similar

Image cropped on [Snapshot Serengeti]

Table of Content

Antelope/deer classification

SUB-PROBLEM

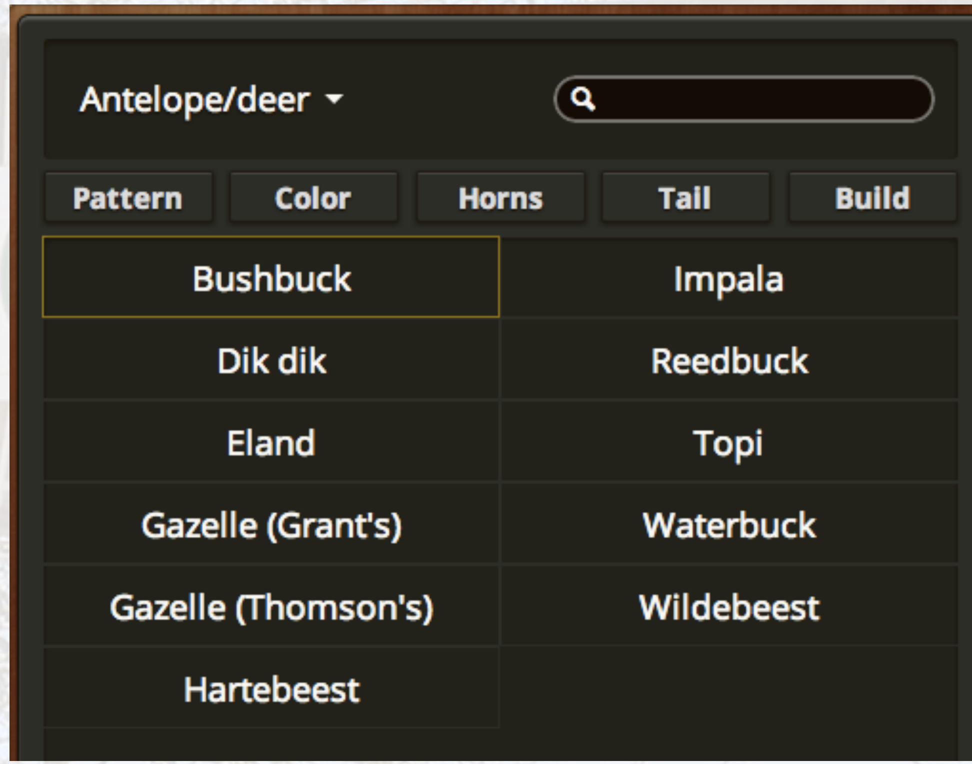


Image cropped on [Snapshot Serengeti]

Dataset Dividing

- Dividing the dataset into
  - Non-existence
  - Easy
  - Hard

Non-existence Group

- Non-existence Group:
  - No animals inside the image
  - Used as negative samples when training the detectors

Nonexistence Group

Easy Group

- Include
  - (1) Full head AND
  - (2)major portion of body
- At least 1 animal has to be identifiable
- Taken in day-time

Easy Group

Hard Group

- Hard Group:
  - Images with animals
  - Occlusion present
  - Only a portion of body part is in image
  - Taken at night

Dataset

Dataset

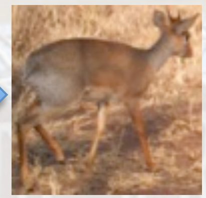
Dataset

Table of Content

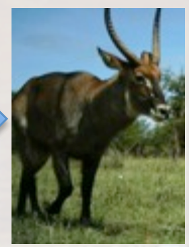
- Background
- Problem Identification
- Dataset
- Approach
  - Detection
  - Classification
- Experimental Result
- Analysis
- Conclusion

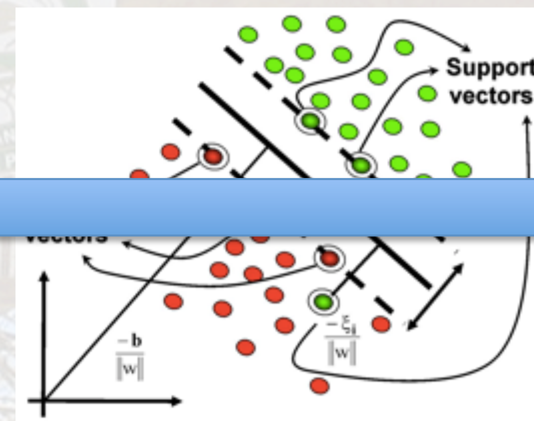
Workflow

Dectection | Classification

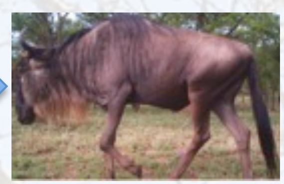


dikdik





waterbuck

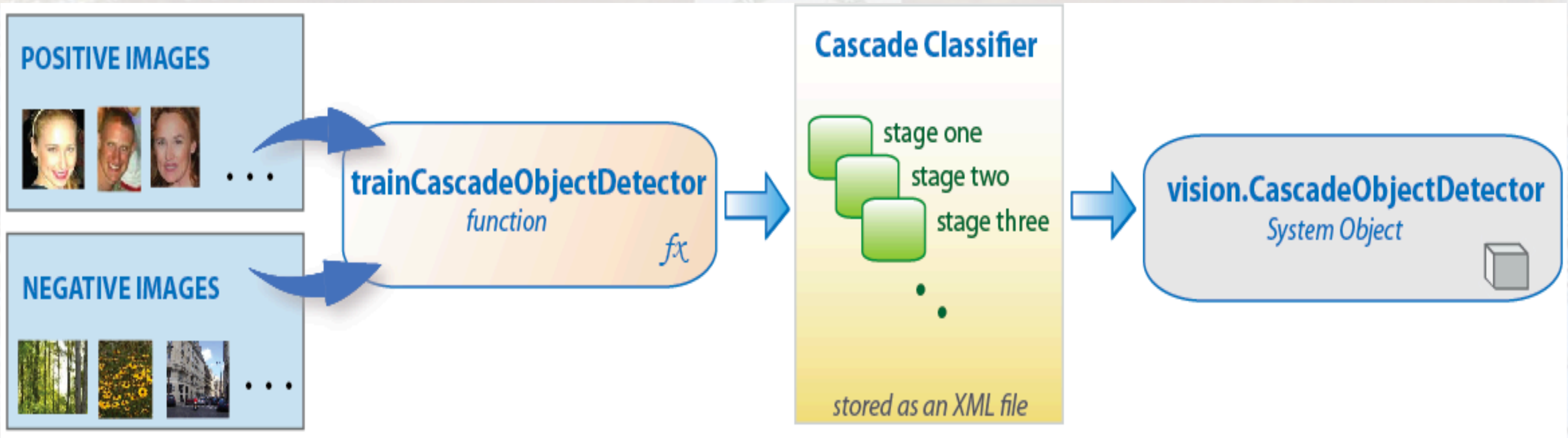


wildbeest

Table of Content

Detection

- Using Cascade Object Detector to identify animals' locations in the images.



Detection

- Positive and negative samples are acquired from ImageNet
- Training takes ~ 1900 positive samples and ~5000 (~3800 are used at each stage) negative samples (previously 600 and 1200)

Detection

- For the training sets:
  - Positive:
    - images of the actual animals being classified
  - Negative:
    - Random images including boxes, trees, cars, etc.
    - Background extracted (thanks to Maja and Guangnan)
    - Images from some Serengeti samples that do not contain the parts wanted.

Detection

- Detectors:
  - Head
    - Animal heads including horns (if present)
  - Body
    - Body portion of the animals which cover part of the legs
  - Legs
    - Legs and hooves

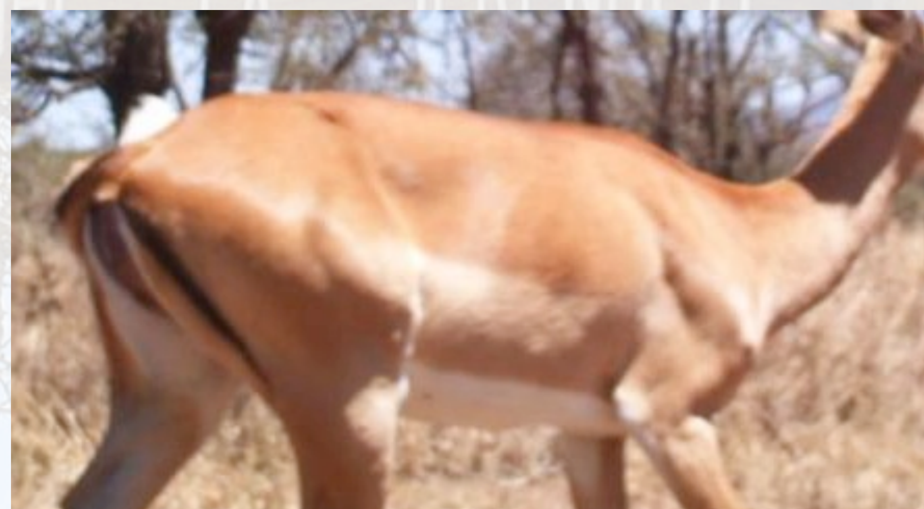
Detection

Raw Image

After detection & processing

← Raw Image

After detection &
processing-->



Detection

- After the detection, crop out the portion of images that contain animals.

- Such images will go through SIFT to obtain the image representation.

Table of Content

Classification

- Image Representation
  - Bag of Visual Words (BoVW) Model
  - Build Visual Word Dictionary
  - Represent an image as a BoVW
- Classification
  - KNNR Classification
  - SVM Classification
- Improvement
  - Background Modeling

Classification

- Image Representation
  - Bag of Visual Words (BoVW) Model
  - Build Visual Word Dictionary
  - Represent an image as a BoVW
- Classification
  - KNNR Classification
  - SVM Classification
- Improvement
  - Background Modeling

SIFT Collection

Build Visual Words Dictionary

- Random collect 100K SIFT descriptors
- User K-mean to get K *Visual Words* (K=1024)
- We get a visual word bank

Represent image as BoVW

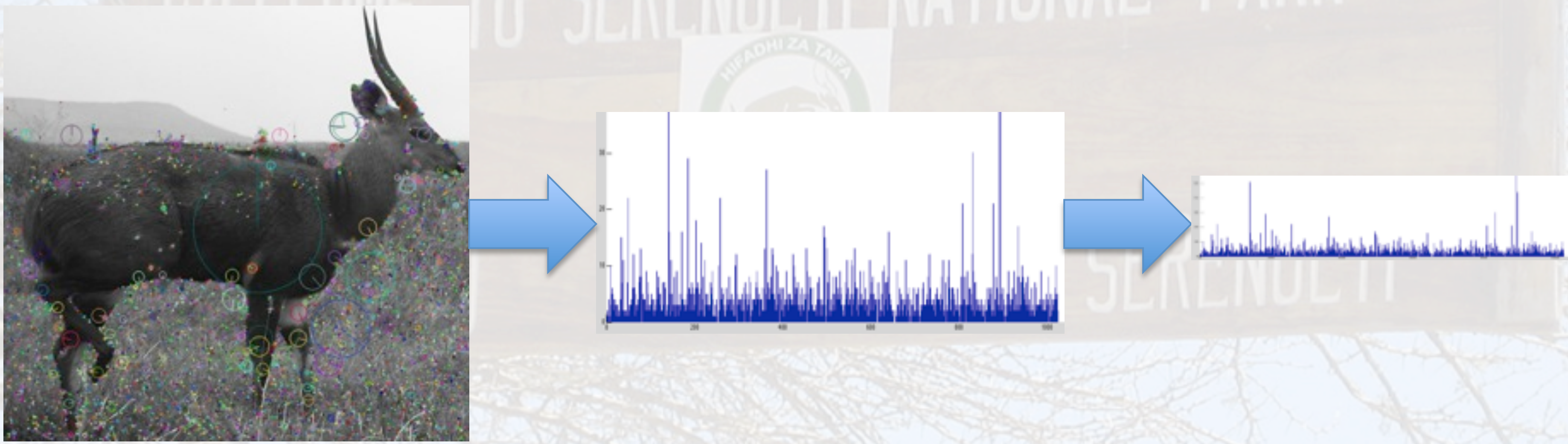


Image
(with features)

Histogram
(on 1024 visual words)

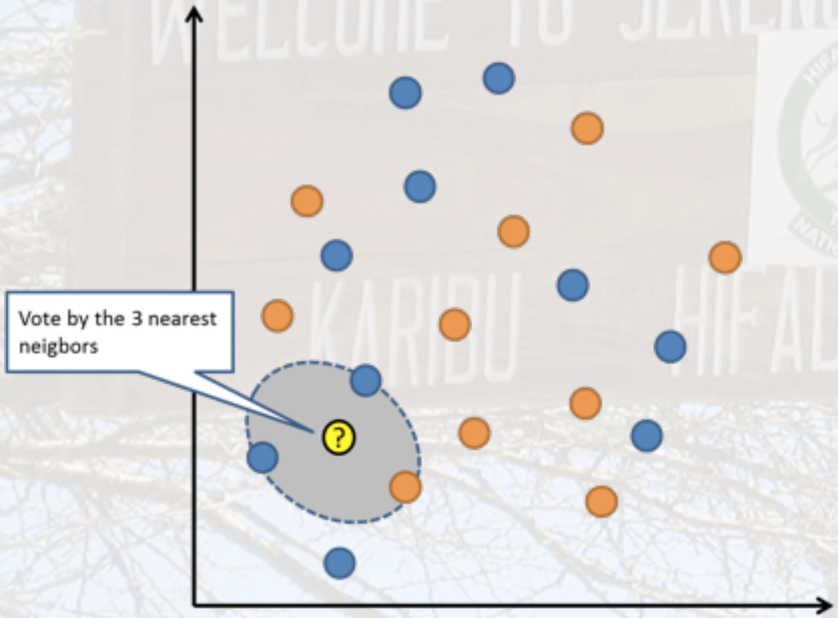
BoVW
(Normalized histogram)

Classification

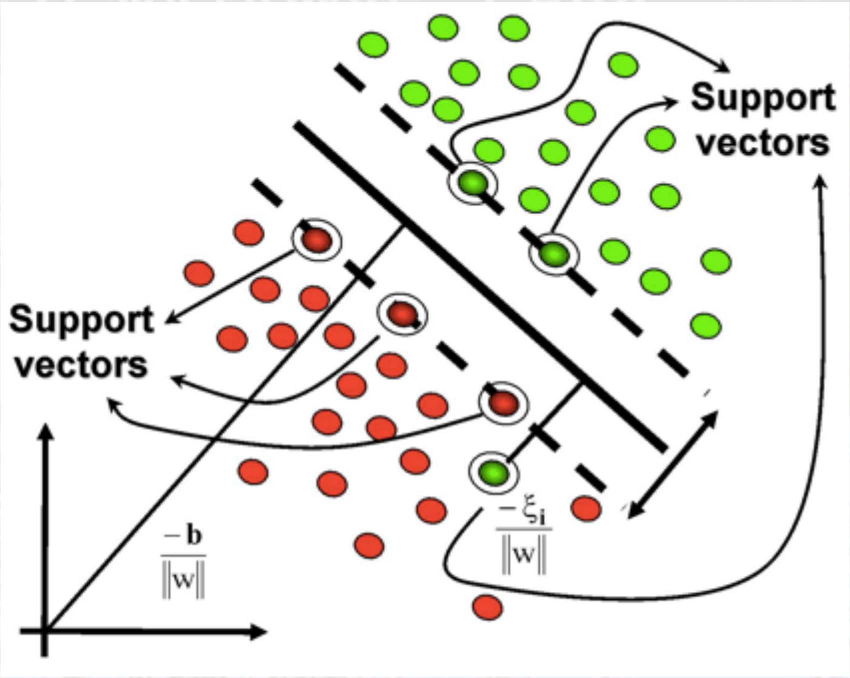
- Image Representation
  - Bag of Visual Words (BoVW) Model
  - Build Visual Word Dictionary
  - Represent an image as a BoVW
- Classification
  - KNNR Classification
  - SVM Classification
- Improvement
  - Background Modeling

Two method to classification

KNN



SVM

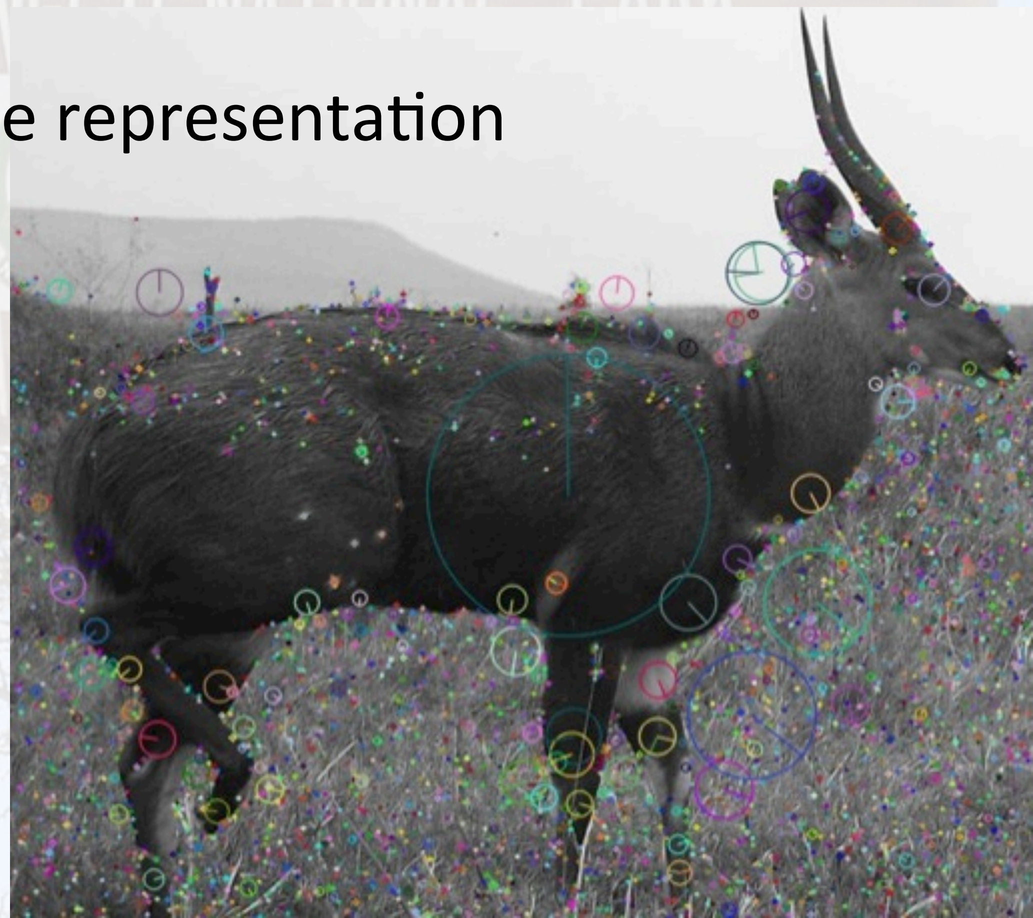


Classification

- Image Representation
  - Bag of Visual Words (BoVW) Model
  - Build Visual Word Dictionary
  - Represent an image as a BoVW
- Classification
  - KNNR Classification
  - SVM Classification
- Improvement
  - Background Modeling

Background Modeling

- Background features exist in the cropped image
- Solution:
  - Extract discriminative representation



Background Modeling

- Assumption
  - Features of each class are drawn from its distribution
  - Distribution of each class has same prior – background distribution
  - Discriminative representation can be extracted based on background distribution
- Use Background image to model the background

Table of Content

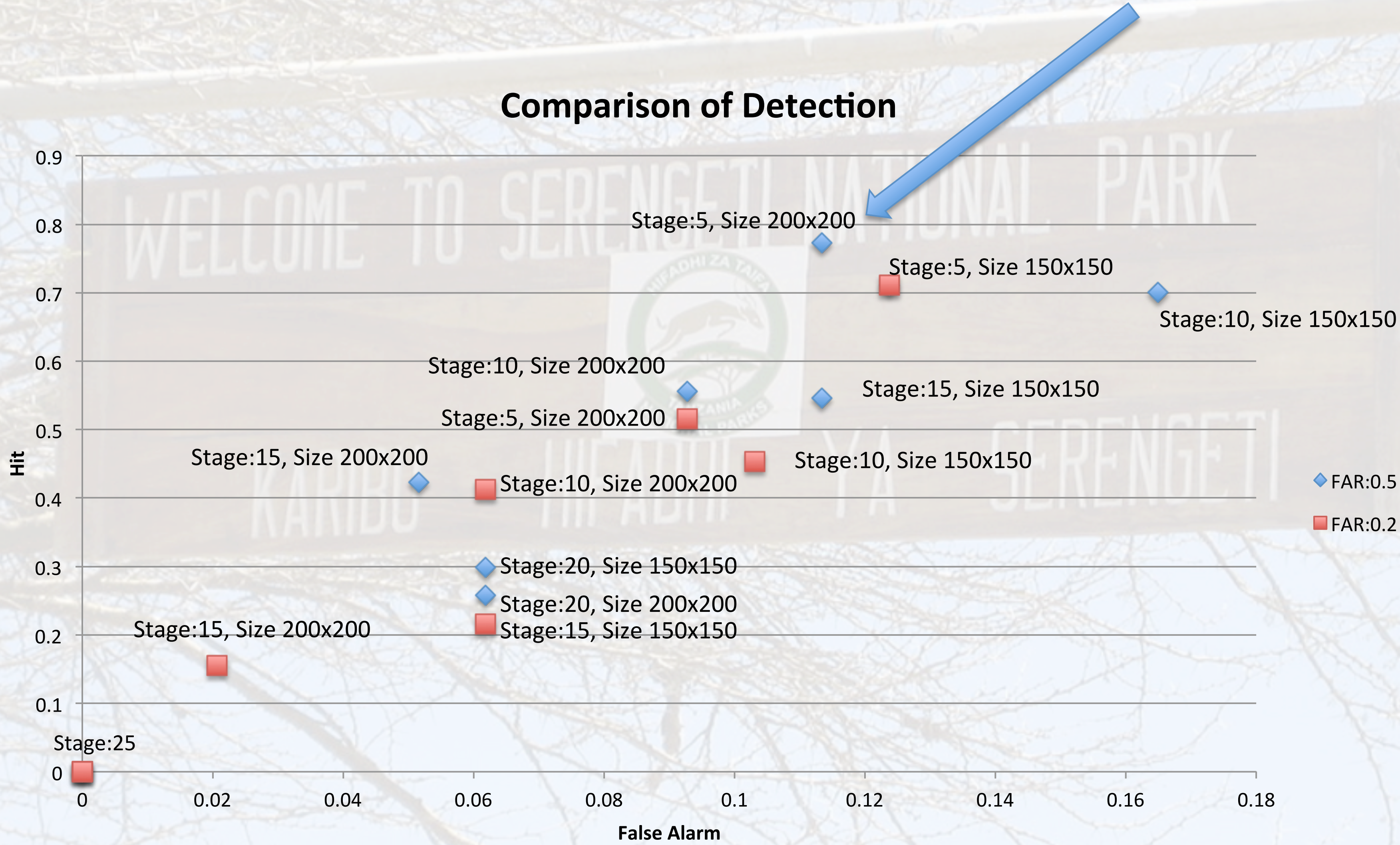
Experimental Results

- Detection Result (isolated)
- Classification Result (isolated)
- Full experiment Result

Training/Testing separation

- Separate 'EASY group' to Training/Testing images for each class Randomly
  - ratio = 2:1
  - total images - 195:97
- Further Test on 'Hard group'

Detection Result (isolated)



Detection Result (isolated)

Test data accuracy record (some examples)				
	Hit	Miss	False Alarm	sum
after feeding in more background data				
FAR:0.5 Stages:5 Size: 200x200	75	11	11	97
FAR:0.5 Stages:15 Size: 200x200	41	51	5	97
FAR:0.5 Stages:15 Size: 150x150	53	33	11	97
FAR:0.5 Stages:10 Size: 150x150	68	13	16	97
FAR:0.5 Stages:10 Size: 200x200	54	34	9	97
FAR:0.2 Stages:5 Size: 200x200	50	38	9	97
FAR:0.2 Stages:5 Size: 150x150	69	18	10	97

Detection Result (isolated)

- 100x100 will have too much “noise” (boxes that are too little covering almost the entire image)

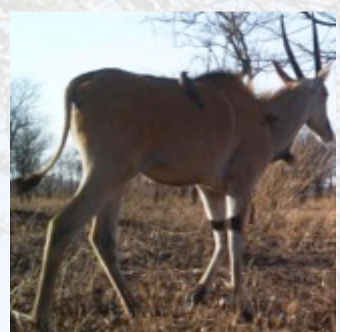
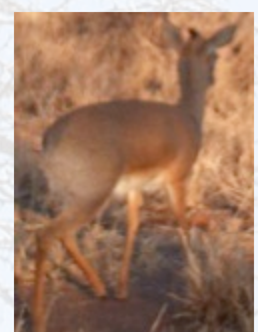
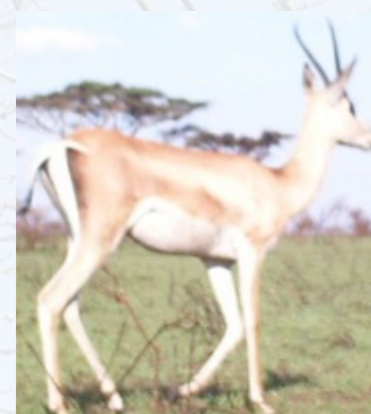
- Stage 25 and above will create almost complete miss for all images.

- Stopped after only 2 FAR setting in the program due to the good result we already found.

Classification Result (isolated)

- Test over Manual cropped Dataset
- Manual cropped Dataset
  - Manual cropped from the original images
  - One deer per cropped image



Classification—Manually Crop

- KNN (K=1)
  - Testing: 69.41%
- SVM
  - Linear-SVM
    - Training: 58.95% (115/195)
    - Testing: 50.59% (43/85)
  - RBF - SVM
    - Training: 57.95% (113/195)
    - Testing: 47.06% (40/85)

Classification--Pipeline

- 1-NNR
  - Testing: 67.05%
  - Hard: 10.78%
- SVM
  - Linear-SVM:
    - Training: 58.97% (115/195)
    - Testing: 52.94% (45/85)
    - Hard: 17.75% (191/1076)
  - RBF – SVM
    - Training: 52.31% (102/195)
    - Testing: 49.41% (42/85)
    - Hard: 16.73% (180/1076)

Discriminative Classification

- Get BoVW from 36 background images (Thanks to Maja and Guangnan group!)
- Get background BoVW by mean over 36 BoVWs.
- Element-wise divide each BoVW to background

Classification—Manually Crop (modified)

Original BoVW

- KNN (K=1)
  - Testing: 69.41%
- SVM
  - Linear-SVM
    - Training: 58.95% (115/195)
    - Testing: 50.59% (43/85)
  - RBF - SVM
    - Training: 57.95% (113/195)
    - Testing: 47.06% (40/85)

Discriminative BoVW

- SVM
  - Linear-SVM
    - Training: 93.85 % (183/195)
    - Testing: 75.29% (64/85)
  - RBF - SVM
    - Training: 89.74% (175/195)
    - Testing: 75.29% (64/85)

Classification—Pipeline (modified)

Original BoVW

- 1-NNR
  - Testing: 67.05%
  - Hard: 10.78%
- SVM
  - Linear-SVM:
    - Training: 58.97% (115/195)
    - Testing: 52.94% (45/85)
    - Hard: 17.75% (191/1076)
  - RBF – SVM
    - Training: 52.31% (102/195)
    - Testing: 49.41% (42/85)
    - Hard: 16.73% (180/1076)

Discriminative BoVW

- SVM
  - Linear-SVM:
    - Training: 94.36% (184/195)
    - Testing: 74.12% (63/85)
    - Hard: 19.80% (213/1076)
  - RBF – SVM
    - Training: 86.15% (168/195)
    - Testing: 57.64% (49/85)
    - Hard: 14.31% (154/1076)

Table of Content

Analysis

- The algorithm works well on ‘Easy’ set, but struggles on ‘Hard’ set.

- SVM does not outperforms 1-NNR classifier. The reason could be due to some pictures are taken one right after another, so the data points are not completely independent.

Analysis

- For linear SVM, the result for manual-cropped images and detect-then-crop images show competitive result, showing that detection can be implemented, at least for images in ‘Easy’ set.
- However, the result could vary since the “easy” set has small sample size.

Table of Content

Conclusion

- In preliminary conclusion, the algorithm works for certain images (day time, one complete animal, etc.) but struggles to classify images with bad lighting, occlusion, and missing body parts.

- A good step for divide-and-conquer.

Literature Research

- Detection:
  - Tracking Animals in Wildlife Using Face Detection—T. Burgahardt
  - Building Models of Animals from Videos—D. Ramanan
  - Hierarchical Recognition: Taxonomy of Animals—S. Yun
- Image Representation:
  - Distinctive Image Features From Scale-Invariant Keypoints", Lowe, D.G.
- Classification:
  - Distance Metric Learning for Large Margin Nearest Neighbor Classification—K. Weinberger