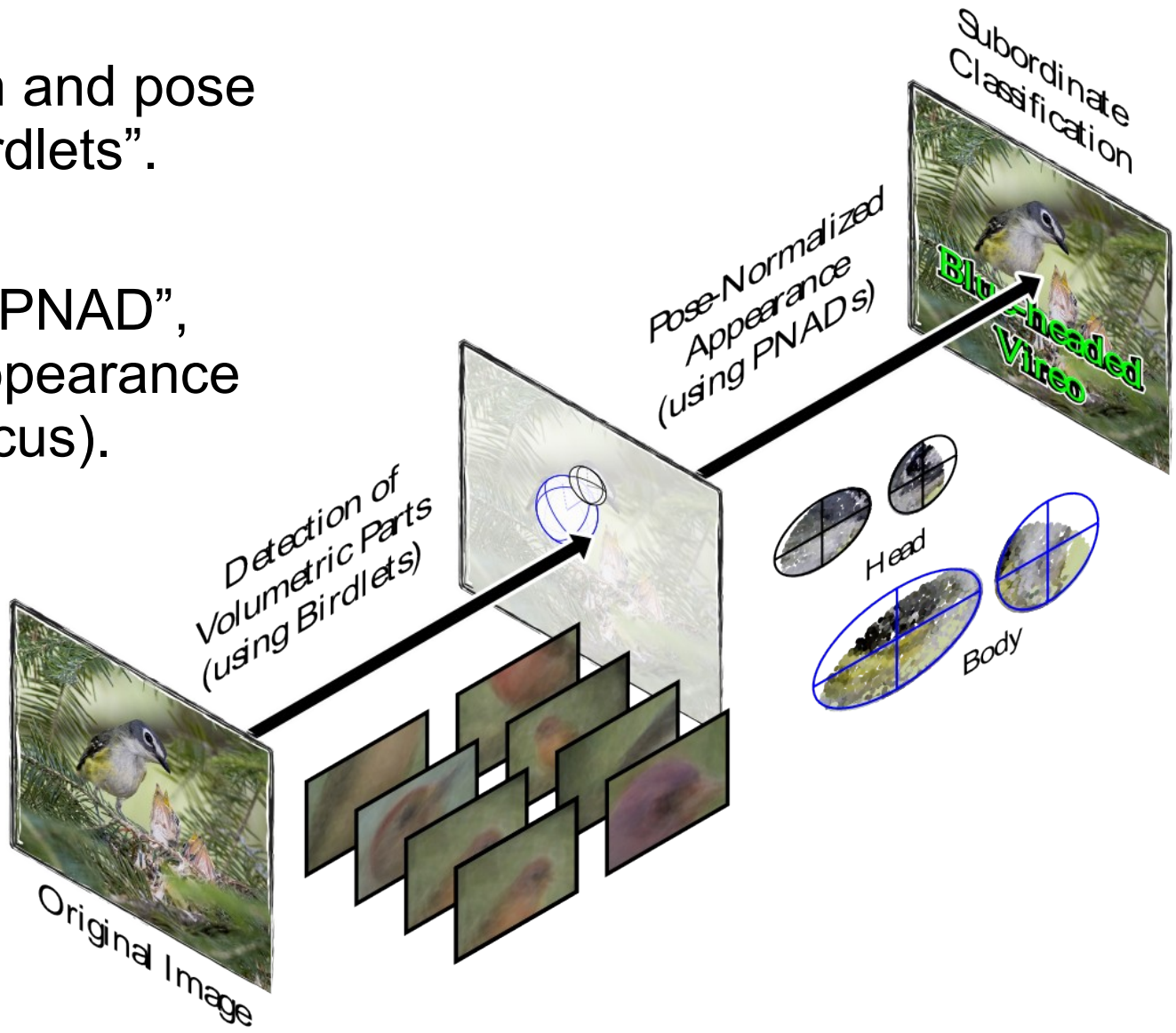
Birdlets: Subordinate Categorization Using Volumetric Primitives and Pose-Normalized Appearance

by Farrell et al. at University of California, Berkeley
Summary by Magnus Burenius, KTH

1. Basic level detection and pose estimation using “Birdlets”.
2. Subordinate level classification using “PNAD”, Pose Normalized Appearance Descriptor. (Their focus).



Basic Geometric Assumption

- Use two ellipsoids to describe the animal.
- Use ellipsoids that are stretched along a single axis.
- Assume scaled orthographic projection.
- The pose of an ellipsoid is defined by its 2D-location, 3D-rotation, stretch & scale (7 parameters).

1. Basic level detection

Detect the parts of the object, i.e. the ellipsoids (using Birdlets).

2. Subordinate level classification

Look at differing properties of these parts: appearance, shape, size, etc (using PNAD).

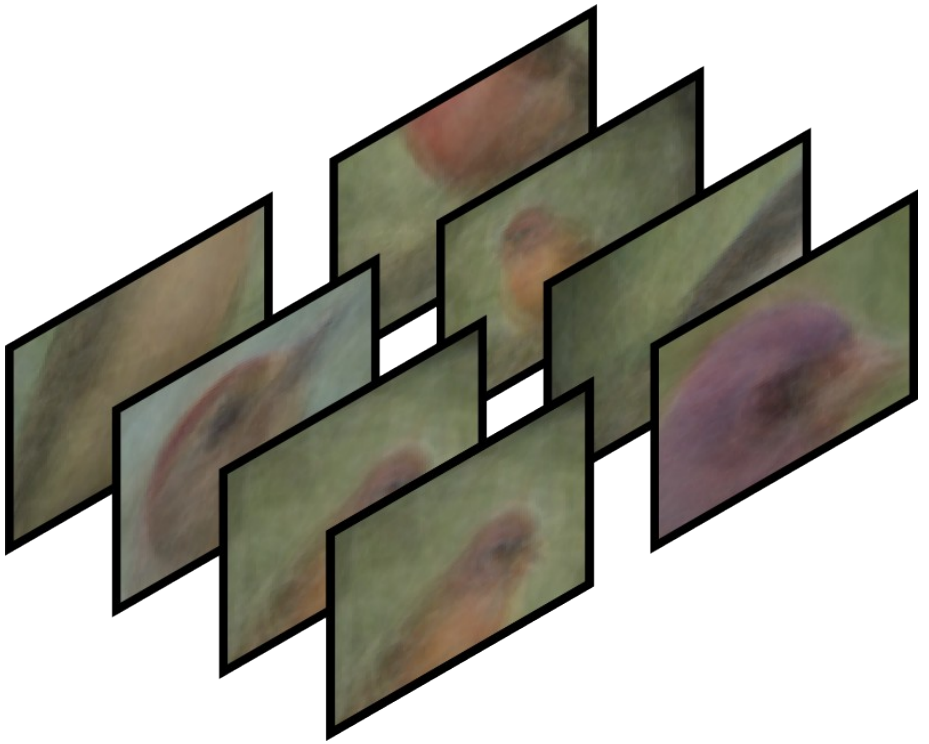
Basic level detection and pose estimation using Birdlets - Annotation

Annotate training data with pose: 2D-location, 3D-rotation, stretch & scale for each ellipsoid.

2*7=14 parameters. Or 14-1=13?

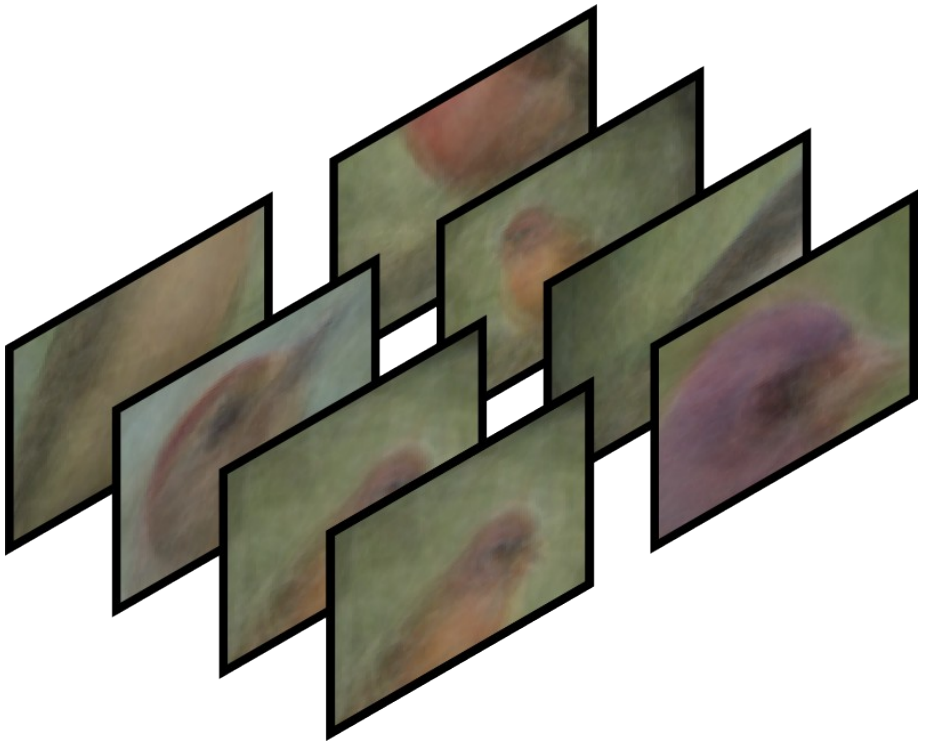
Basic level detection and pose estimation using Birdlets - Training 1

- Define a distance function for poses, taking rotation and scale into account.
- Group training images with similar poses. These groups are called Birdlets (based on the Poselet framework).
- Within each group the images are aligned by computing a 2D similarity transformation and the pixels are warped to a canonical rectangle.



Basic level detection and pose estimation using Birdlets - Training 2

- They use rectangles of 96x64 pixels.
- For each 8x8 region a HOG is computed and concatenated into a feature vector.
- For each Birdlet a classifier is learnt using SVM.
- Use a retraining stage, where false positives of the initial classifier are collected before the final classifier is learnt.

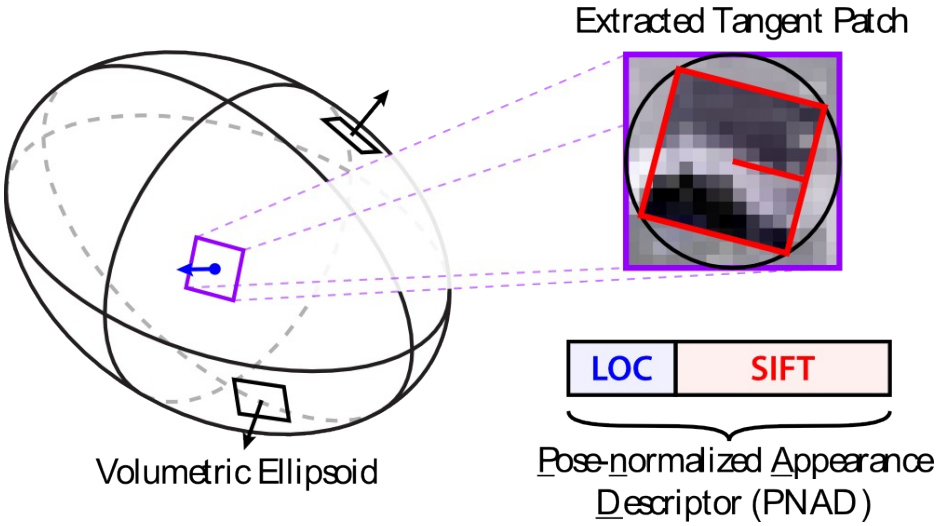


Basic level detection and pose estimation using Birdlets - Runtime

- Use a sliding window approach and apply the classifier of each Birdlet over the image.
- Record high responses and their pose. Each response votes for a pose.
- Finally cluster the recorded responses into one or more final responses (a bit tricky).

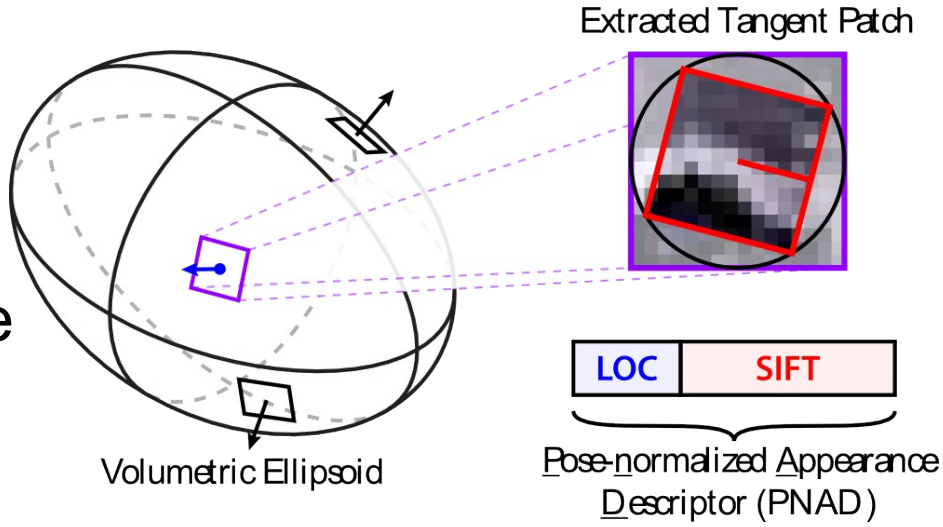
Subordinate level classification using PNAD (Their main focus)

- PNAD - Pose Normalized Appearance Descriptor
- Assume that we have an image and a known ellipsoid pose.
- Sample (visible) points on the ellipsoid randomly.



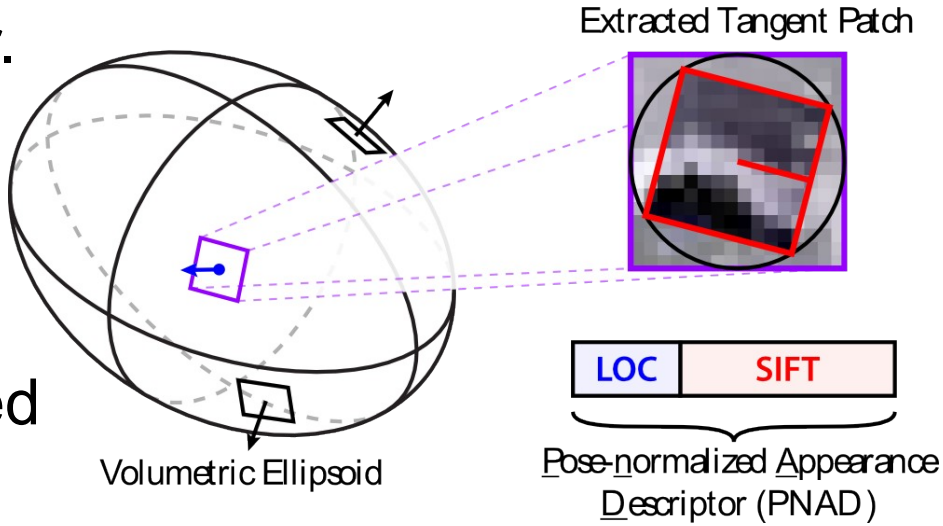
Subordinate level classification using PNAD

- Consider a small square patch in the tangent plane of the sampled point.
- To get the pixel content of the tangent patch it is projected to the image.
- This gives a parallelogram which can be warped to a square.



Subordinate level classification using PNAD

- A color-SIFT descriptor of this patch is computed.
- The surface position and the SIFT is concatenated to a single vector.
- Can also concatenate extra information like, ellipsoid stretch and size.
- This vector is the Pose Normalized Appearance Descriptor PNAD, which couples the pose and appearance of a patch.
- For each image several PNADs are computed as a description of the visible region of the ellipsoid.



Subordinate level classification

- They train a classifier based on Stacked Evidence Trees.
- These rely on a Random Forrest constructed such that all leaves are required to have a specified minimum number of training samples (they use 20).
- When a query (PNAD) is passed through the forest and reaches a leaf, the classifier returns the distribution of class labels that reached that leaf during training.
- The resulting class distribution for each query are aggregated into a single evidence vector (each PNAD votes for several class labels).

Subordinate level classification

- A second-stage (“stacked”) multi-class adaboost classifier is then applied to the class distribution evidence vector, producing the final category prediction. Why?

- The Stacked Evidence Tree model was selected for the way it complements the Pose Normalized Appearance model, providing a way to handle the occlusion.
(Compare Bag of Words)

Experimental Results?

Successful detections

Unsuccessful detections