Visipedia Tool Ecosystem for Dataset Curation and Annotation

Serge Belongie

CORNELL NY C TECH
Outline

- Visipedia Project Overview
- Related Work
- Bird Datasets
- ViBE: Visipedia Back End
- Future Work
Outline

- Visipedia Project Overview
- Related Work
- Bird Datasets
- ViBE: Visipedia Back End
- Future Work
What Is Visipedia?

- A user-generated encyclopedia of visual knowledge
- An effort to associate articles with large quantities of well-organized, intuitive visual concepts

http://en.wikipedia.org/wiki/Bird
Motivation

- People will willingly label or organize certain images if:
  - They are interested in a particular subject matter
  - They have the appropriate expertise
While we’re not particularly certain about some of the claims in this bicycle family tree (e.g. freeride bikes spawned downhill bikes, which gave birth to 29ers?), we’re certain you’ll appreciate the artwork. You can head to their website to buy your own copy and laugh at the implication that big wheels evolved into recumbents for only $22.
**COMMENTS**

**Ben - 09/04/13 - 5:44pm**
This is so completely out of order. Why the hell would you pay 22 dollars for a poster that doesn’t make any sense?

**Gillis - 09/04/13 - 6:00pm**
I like how the track bike sits in between the randonneur and touring bikes. And a modern looking TT bike some how comes before Boardman’s Lotus, which both come after fixie’s. This is junk.

**Walter - 09/04/13 - 6:09pm**
So triathlon bikes gave birth to fixies and early eighties long wheel base recumbents came from modern high racers. These folks are creationists.

**NotAMachinist - 09/04/13 - 6:20pm**
It’s sort of cool looking until you really look at it. For instance how does a randonneur differ from a touring or trekking bike? How did cyclocross spawn BMX? The cycling family tree is far more incestuous.

**Joe - 09/04/13 - 6:22pm**
Cyclocross to 20” Dirt, Street, Park to Racing to Freestyle and Flatland????? Not a one of these is right....the whole chart is a nice piece of wallpaper art but that’s it.

**Keith D - 09/04/13 - 6:46pm**
It’s pretty much rubbish.
Motivation

- Construct comprehensive, intuitive knowledge base of visual objects
- Provide better text-to-image search and image-to-article search
Outline

- Visipedia Project Overview
- Related Work
- Bird Datasets
- ViBE: Visipedia Back End
- Future Work
Related Work: Systems

- {Leaf, Dog, Bird}snap [Belhumeur et al.]
- Oxford Flowers [Nilsback & Zisserman]
- STONEFLY9 [Martínez-Muñoz et al.]
- omoby [IQEngines.com]
- 20 Questions game [20q.net]
- ReCAPTCHA [von Ahn et al.]
- Wikimedia Commons
Related Work: Methods

- Relevance Feedback
- Active Learning
- Expert Systems
- Decision Trees
- Feature Sharing & Taxonomies
- Parts & Attributes
- Crowdsourcing & Human Computation
Motivation:
Computer Vision Perspective

● Need for more training data
  ○ Beyond the capacity of any one research group
  ○ Better quality control

● Need for more realistic data
  ○ Let people define what tasks are important
  ○ Study tightly-related categories
Dealing With a Large Number of Related Classes

- Standard classification methods fail because:
  - Few training examples per class available
  - Variation between classes is small
  - Variation within a class is often still high

Brewer’s Sparrow

Vesper Sparrow
(A) Easy for Humans  
Chair? Airplane? ...

(B) Hard for Humans  
Finch? Bunting? ...

(C) Easy for Humans  
Yellow Belly? Blue Belly? ...
Plants vs Birds

2d
- Doesn’t move
- Okay to pluck from tree
- Mostly single color
- Very few parts
- Adequately described by boundary
- Relatively easy to segment

3d
- Moves
- Not okay to pluck from tree
- Many colors
- Many parts
- Not well described by boundary
- Hard to segment

slide credit: Neeraj Kumar
Visual 20 Questions

- “Computer Vision” module = Vedaldi’s VLFeat
- VQ Geometric Blur, color/gray SIFT spatial pyramid
- Multiple Kernel Learning
- Per-Class 1-vs-All SVM
- 15 training examples per bird species
- Choose question to maximize expected Information Gain
Pose Normalized Deep ConvNets

[Van Horn, Branson, Perona, Belongie BMVC 2014]
Algorithm 1 Visual 20 Questions Game

1: \( U^0 \leftarrow \emptyset \)
2: for \( t = 1 \) to 20 do
3: \( j(t) = \max_k I(c; u_k | x, U^{t-1}) \)
4: Ask user question \( u_{j(t)} \), and \( U^t \leftarrow U^{t-1} \cup u_{j(t)} \).
5: end for
6: Return class \( c^* = \max_c p(c|x, U^t) \)
Outline

- Visipedia Project Overview
- Related Work
- Bird Datasets
- ViBE: Visipedia Back End
- Future Work
Birds-200 Dataset

6033 images over 200 bird species
Image Harvesting

- Flickr: text search on species name
- MTurk: presence/absence and bounding boxes
Illustrations of Eastern Bluebird

Click to select images below which you think contain Eastern Bluebird.
Remember that the color of the bird in a photo is sometimes slightly different from the illustrations.

Click here to view detailed instructions. Please read at least once.

Select images by clicking on them so that a green border appears.

Click here to provide feedback on this HIT. (Will expand this section to show a form.)

Tick this box if you are particularly interested in birds and consider yourself an expert on Eastern Bluebird (this will not affect your payment).
The human annotation process

- Modeling various aspects of annotation:
  - *Worker competency* – accuracy in labeling
  - *Worker expertise* – better at labeling some things than others, based on their strengths
  - *Worker bias* – how one weighs errors
  - *Task difficulty* – ambiguous images are universally hard to label
  - *True label* – the ground truth label value

- We leverage the "Multidimensional Wisdom of Crowds" [Welinder et al. 2010]
Types of annotator errors

Task: Find the Indigo Bunting

Select images by clicking on them so that a green border appears.

Indigo Bunting

Blue Grosbeak
Task: Find the Indigo Bunting

- 6% error
- 15% error
- 31% error
- 50% error

hit rate (correct detection) vs. rate of correct rejection
Task: Find the Indigo Bunting

hit rate (correct detection)

rate of correct rejection

competent

6% error

15% error

31% error

50% error
Task: Find the Indigo Bunting

hit rate (correct detection) vs. rate of correct rejection

- 6% error
- 15% error
- 31% error
- 50% error

dots represent bots
Task: Find the Indigo Bunting

hit rate (correct detection)

rate of correct rejection

6% error
15% error
31% error
50% error

optimists
Task: Find the Indigo Bunting

- 6% error
- 15% error
- 31% error
- 50% error

hit rate (correct detection)
rate of correct rejection

pessimists
Image formation process

Object presence or absence

$Z_i$ 

$I_i$ 

$X_i$ 

Factors influencing appearance

(specimen, pose, location, weather, viewpoint, camera, ...)
Entire annotation process

\[
p(L, x, w, \tau) = \prod_{j=1}^{M} p(\tau_j | \gamma)p(w_j | \alpha) \prod_{i=1}^{N} \left( p(x_i | \theta_z, \beta) \prod_{j \in I_i} p(l_{ij} | x_i, w_j, \tau_j) \right)
\]
Image difficulty

Ground truth decision plane at $x_i = 0$

$p(x_i \mid z_i = 0)$

$p(x_i \mid z_i = 1)$

A

B

C

$-3$ $-2$ $-1$ $0$ $\tau_j$ $1$ $2$ $3$

$Y_{ij}$

[Welinder et al., 2010]
Annotator competence

High competence

Low competence

[Welinder et al., 2010]
Multidimensional ability of annotators

\[ x_i = (x_i^1, x_i^2) \]

\[ p(x_i \mid z_i = 1) \]

\[ p(x_i \mid z_i = 0) \]
Multidimensional ability of annotators

\[ x_i = (x_i^1, x_i^2) \]

\[ p(x_i \mid z_i = 1) \]

\[ p(x_i \mid z_i = 0) \]
Multidimensional ability of annotators
Worker “schools of thought”

Ducks

Ducks and grebes

Ducks, grebes, and geese

*
Discussion: quality management

- Models can capture multidimensionality of annotation process
- How well does this generalize to continuous annotations?

Different tasks require different reviewing strategies.
Predicting quality accurately can reduce the number of labels needed.
Please draw a rectangle around the bird in the image. The rectangle should be fit around the bird tightly.

**Good Rectangles:**

- [Image of a correctly drawn rectangle around a bird]
- [Image of another correctly drawn rectangle around a bird]
- [Image of a correctly drawn rectangle around a bird]
- [Image of a correctly drawn rectangle around a bird]

**Bad Rectangles:**

- [Image of an incorrectly drawn rectangle around a bird]
- [Image of an incorrectly drawn rectangle around a bird]
- [Image of an incorrectly drawn rectangle around a bird]
- [Image of an incorrectly drawn rectangle around a bird]

Tick any of the boxes below if they are true about the image.

- [ ] Bird is **truncated**
- [ ] Bird is **occluded**
- [ ] There is more than one bird
- [ ] There is no bird in the image

1/5  [Previous]  [Next]  [✓ Submit]

Please provide feedback in the box below if you have comments:

[Blank input field]
Attribute Labeling

- Attributes from whatbird.com
- 25 visual attributes → 288 binary attributes
  - similar to “dichotomous key” in biology
- MTurk interface
  - \{guessing, probably, definitely\}
- 3-5x redundancy factor
You will be asked to answer a series of questions based on identifying visual features from the bird image on the left. Closely follow the specific instructions for each question. Holding the mouse over each selectable option for 1 second will provide additional instructions or examples.

What is the pattern of the breast of the bird? 1/12

Select one. If the breast isn't visible, make your best guess, then select "Guessing".

- Solid
- Multi-Colored
- Striped
- Spotted

Go Back  Guessing  Probably  Definitely
You will be asked to answer a series of questions based on identifying visual features from the bird image on the left. Closely follow the specific instructions for each question. Holding the mouse over each selectable option for 1 second will provide additional instructions or examples.

What is the **color of the crown?**

Select at least one. If the crown isn't visible, make your best guess, then select "Guessing". If the color is a mixture of two colors, select both (e.g., for blue-green select blue and green). If the crown has two distinct colors, select both (e.g., for yellow with black stripes, select yellow and black).
You will be asked to answer a series of questions based on identifying visual features from the bird image on the left. Closely follow the specific instructions for each question. Holding the mouse over each selectable option for 1 second will provide additional instructions or examples.

What is the **shape of the bill/beak?**

*Select one. If the beak isn't visible, make your best guess, then select "Guessing".*

- All-purpose
- Cone
- Curved (up or down)
- Dagger
- Hooked
- Hooked Seabird
- Needle
- Spatulate
- Specialized

[Go Back] [Guessing] [Probably] [Definitely]
You will be asked to answer a series of questions based on identifying visual features from the bird image on the left. Closely follow the specific instructions for each question. Holding the mouse over each selectable option for 1 second will provide additional instructions or examples.

What is the **color of the wings** of the bird?

Select at least one. If the wings aren't visible, make your best guess, then select "Guessing". If the color is a mixture of two colors, select both (e.g., for blue-green select blue and green). If the wings have two distinct colors, select both (e.g., for yellow with black stripes, select yellow and black).

- White
- Black
- Grey
- Buff
- Brown
- Rufous
- Red
- Pink
- Orange
- Yellow
- Green
- Olive
- Blue
- Purple
- Shiny / Iridescent

Go Back  Guessing  Probably  Definitely
MTurker Label Certainty
MTurker Feedback

- “These hits were fun. Will you be posting more of them anytime soon? Thanks!”
- “These are Beautiful birds and I am enjoying this hit collection”
- “I really enjoy doing your hits, they are fun and interesting. Thanks.”
- “Love doing these because I'm a bird watcher.”
- “the birds are so cute..hope u can send more kind of birds”
- “I haven't really studied birds, but doing these HITs has made me realize just how beautiful they are. It has also made me aware of the many different types of birds. Thank you”
- “I REALLY LOVE THE COLOR OF THE BIRDS.”
- “Thank you for providing this job. The fact that the images are beautiful to look at make it a lot more enjoyable to do!”
- “Enjoyable to do.”
- Hourly Wage ≈ $1.25
Try out on a new dataset for fine-grained recognition, featuring 550 of North America’s most common birds. The full dataset will be available in the fall. Join the competition today and download the "taster" dataset!

http://birds.cornell.edu/nabirds

CCUB NABirds includes:

- More than 700 visual categories, organized taxonomically
- Photos curated in collaboration with domain experts
- Data organized in a researcher-friendly, widely-used PASCAL VOC format

For more information contact: Ryan Farrell (farrell@eecs.berkeley.edu)
CCUB Taster25 Results

Baseline Performance:
The winning ILSVRC '11 approach of Florent Perronnin and Jorge Sanchez.

- Dense SIFT and Color Descriptors
- Linear SVMs with SGD
- Same parameters used in ILSVRC

Using the winning ILSVRC '11 approach by [F. Perronnin, et al.], training on 25 images/category
**CCUB Taster25 Results**

Using the winning ILSVRC ’11 approach by [F. Perronnin, et al.],
training on 50 images/category
Try out on a new dataset for fine-grained recognition, featuring 550 of North America’s most common birds. The full dataset will be available in the fall. Join the competition today and download the “taster” dataset!

http://birds.cornell.edu/nabirds

For more information contact: Ryan Farrell (farrell@eecs.berkeley.edu)
Outline

- Visipedia Project Overview
- Related Work
- Bird Datasets
- ViBE: Visipedia Back End
- Future Work
Vibe Demo

http://visipedia.org
http://vibe.visipedia.org
Visipedia Backend

Storage and collaboration infrastructure to support visual search applications.

Storage
Cloud storage and access for your image datasets and annotations.

Collaborate
Divide and conquer your data collection and curation tasks by sharing your data with collaborators.

Organize
Build a hierarchical representation of your domain and use it to organize your images.

Annotate
Use our annotation templates to create your custom annotation tasks.

Deploy
Integrate Vibe storage functionality into your app or website for easy image upload and annotation by your users.

Analyze
Hook Vibe into your classification pipeline to analyze how images are being classified.
<table>
<thead>
<tr>
<th>Species</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Black-bellied Whistling Duck</td>
<td>169</td>
</tr>
<tr>
<td>Fulvous Whistling Duck</td>
<td>44</td>
</tr>
<tr>
<td>Greater White-fronted Goose</td>
<td>152</td>
</tr>
<tr>
<td>Snow Goose (White morph)</td>
<td>177</td>
</tr>
<tr>
<td>Snow Goose (Blue morph)</td>
<td>118</td>
</tr>
<tr>
<td>Ross's Goose</td>
<td>155</td>
</tr>
<tr>
<td>Brant</td>
<td>149</td>
</tr>
<tr>
<td>Cackling Goose</td>
<td>158</td>
</tr>
<tr>
<td>Canada Goose</td>
<td>163</td>
</tr>
<tr>
<td>Mute Swan</td>
<td>225</td>
</tr>
<tr>
<td>Trumpeter Swan</td>
<td>167</td>
</tr>
<tr>
<td>Tundra Swan</td>
<td>196</td>
</tr>
<tr>
<td>Muscovy Duck</td>
<td>91</td>
</tr>
<tr>
<td>Wood Duck (Breeding male)</td>
<td>281</td>
</tr>
<tr>
<td>Wood Duck (Female/Eclipse male)</td>
<td>186</td>
</tr>
<tr>
<td>Gadwall (Breeding male)</td>
<td>226</td>
</tr>
<tr>
<td>Gadwall (Female/Eclipse male)</td>
<td>133</td>
</tr>
<tr>
<td>American Wigeon (Breeding male)</td>
<td>190</td>
</tr>
<tr>
<td>American Wigeon (Female/Eclipse male)</td>
<td>139</td>
</tr>
<tr>
<td>Eurasian Wigeon (Breeding male)</td>
<td>154</td>
</tr>
<tr>
<td>Eurasian Wigeon (Female/Eclipse male)</td>
<td>24</td>
</tr>
<tr>
<td>American Black Duck</td>
<td>121</td>
</tr>
<tr>
<td>Mottled Duck</td>
<td>125</td>
</tr>
<tr>
<td>Blue-winged Teal (Male)</td>
<td>121</td>
</tr>
<tr>
<td>Blue-winged Teal (Female/Juvenile)</td>
<td>187</td>
</tr>
<tr>
<td>Cinnamon Teal (Male)</td>
<td>211</td>
</tr>
<tr>
<td>Cinnamon Teal (Female/Juvenile)</td>
<td>103</td>
</tr>
<tr>
<td>Northern Shoveler (Breeding male)</td>
<td>250</td>
</tr>
<tr>
<td>Northern Shoveler (Female/Eclipse male)</td>
<td>150</td>
</tr>
</tbody>
</table>
Merlin is equipped with computer vision to identify birds in images. Thanks to help from the community who shared and annotated images, Merlin has been trained to identify 125 species. At this stage, Merlin is well-suited to handle high-quality images taken with DSLR cameras. Results may not be as good for images taken on iPhones and other mobile devices, however, we encourage you to see how Merlin does with lower-quality images! All images uploaded to this test interface will be used to help improve the system in the future. We appreciate your feedback as Merlin development continues.
Select your photo.

Or drag and drop your file into this box.

1. Upload one image at a time.
2. Upload a jpeg or png image.
3. Image must be less than 10 MB.
Did you take this photo?

Yes
No
Crop the bird by clicking and dragging a box.
Click on the bill tip.
Click on the eye. If both eyes are visible, click on the side of the head that is more visible.
Where did you see the bird?

1. Search or click on the map to place a marker where the photo was taken.
2. Then, click Next to confirm the location.

Next

Search by City and State or Zip Code
When did you see the bird?

August 6

Next
Creating list of possible birds...
Can't find your bird?

Merlin considers the 400 most familiar species in North America. The bird you're seeing might not be in our database yet. Browse species.
Outline

- Visipedia Project Overview
- Related Work
- Bird Datasets
- ViBE: Visipedia Back End
- Future Work
Future Work

- Beyond Birds
- Attribute Induction
- Relevance Feedback
Click on the birds to the right that are clearly **dissimilar** species from the bird above.

\[ T^k = \{(i, j, l) | x_i \text{ is more similar to } x_j \text{ than } x_l \} \]

\[ \|z_i - z_j\|_2 < \|z_i - z_l\|_2 \iff s(i, j) > s(i, l) \]
Thank You

- Caltech: Steve Branson, Grant Van Horn, Pietro Perona
- UCSD: Catherine Wah
- Cornell: Jessie Barry, Miyoko Chu
- BYU: Ryan Farrell
- Google Focused Research Award
Extra Slides
Figure 1.4. In our crowdsourcing framework, workers are asked which pair images in a triplet look most similar. Given a large collection of such relative measurements, we can learn a similarity function on image patches that captures human perception. Left: triplet depicting three regions of interest from a slide image. Middle: zoomed-in regions. Right: the chosen pair (clicked by the worker) is indicated in green.
Populating Visipedia

- Populate Wikipedia articles with more visual data using large quantities of unlabeled data on the web
Attribute-Based Classification

- Train classifiers on attributes instead of objects
- Attributes are shared by different object classes
- Attributes provide the ingredients necessary to recognize each object class

Lampert et al. 2009
Farhadi et al. 2009
Attribute-Based Classification

- Number of attributes is less than number of classes
- Attribute classification tasks might be easier
- Makes it easier to incorporate human knowledge

Brewer's Sparrow

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Order</td>
<td>PASSERIFORMES</td>
</tr>
<tr>
<td>Family</td>
<td>Sparrows (Emberizidae)</td>
</tr>
<tr>
<td>Code 4</td>
<td>BRSP</td>
</tr>
<tr>
<td>Code 6</td>
<td>SPIBRE</td>
</tr>
<tr>
<td>ITIS</td>
<td>179440</td>
</tr>
</tbody>
</table>

**BODY**
- Length Range: 14 cm (5.5 in)
- Weight: 14 g (0.5 oz)
- Size: Small (5 - 9 in)
- Color Primary: Brown, Gray
- Underparts: Pale Gray
- Upperparts: Gray-brown with black streaking
- Back Pattern: Striped or streaked
- Belly Pattern: Solid

**HEAD**
- Bill Shape: Cono
- Eye Color: Dark hazel or blackish brown at all ages.
- Head Pattern: Eye line, Striped, Streaked, Eyeing, Malar or malar stripe
- Crown Color: Gray-brown with fine black streaking
- Forehead Color: Gray-brown with fine black streaking
- Nape Color: Gray-brown with fine black streaking
- Throat Color: White with gray wash
- Cere color: No Data

www.whatbird.com
Annotator bias

A

B

C

\[ p(x_i \mid z_i = 0) \quad p(x_i \mid z_i = 1) \]

\[ p(y_{ij} \mid x_i) \]

[Welinder et al., 2010]