Fine-grained Categorization of Botanical Shapes

Donald Geman
Johns Hopkins University

IPAM Summer School
Institute for Pure and Applied Mathematics
Los Angeles, August 9, 2013

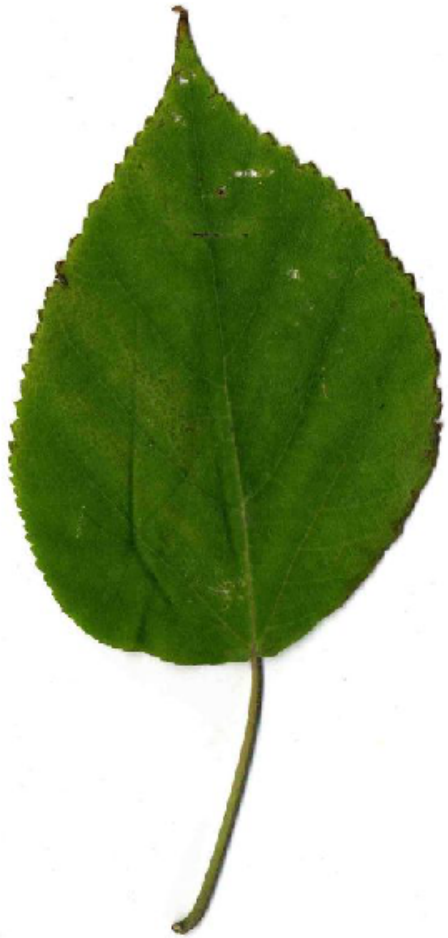
COLLABORATORS

- Asma Rejeb Sfar (PhD student, INRIA)
- Nozha Boujemaa (INRIA)

Outline

- Introduction
- Representation
- Learning
- Classification
- Conclusion

PROBLEM



What plant species is this?

Why identify plants?

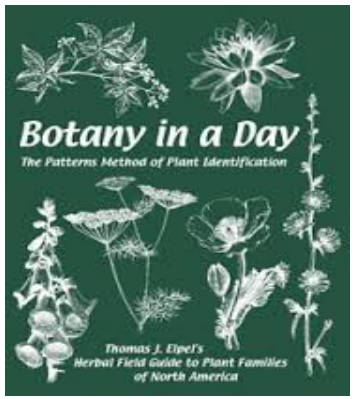
- Field studies
- Conserve biodiversity
- Improve agricultural productivity
- Develop educational tools, etc.

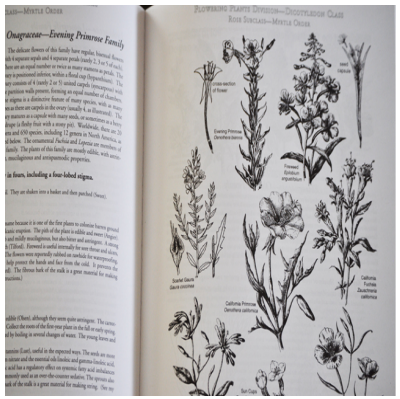
Why leaves?

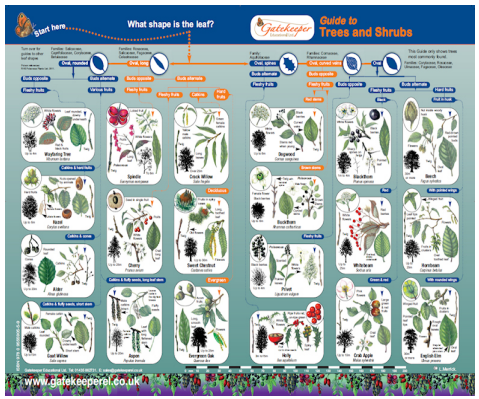
- Discriminating for taxonomic identity.
- Present for much of the year (unlike more transient organs).
- Easily collected.

Manual process

- Field guide:
  - Pictures organized by family, shape, location or other descriptors;
  - Includes identification keys to assist with identification.







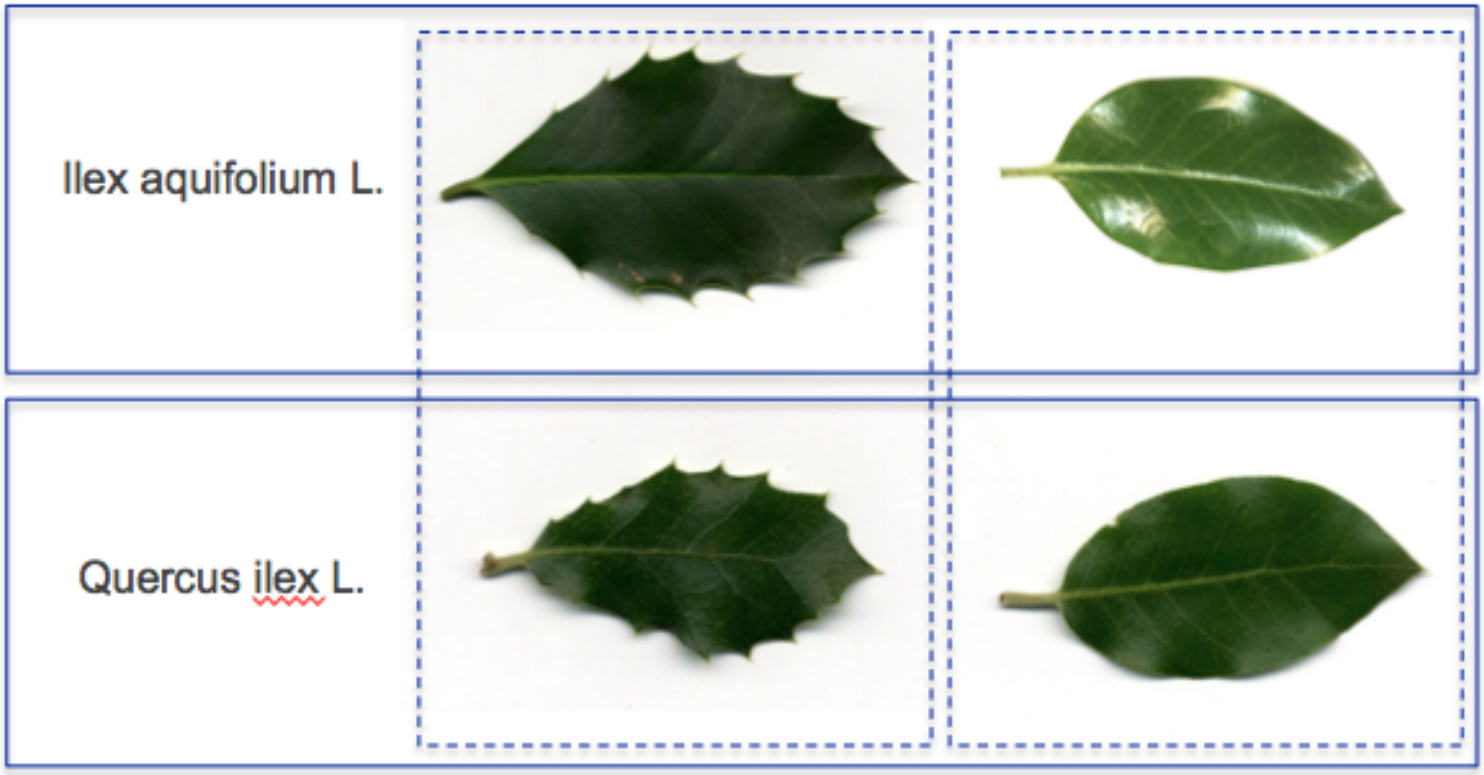
- Botanists generally proceed sequentially and adaptively.
- They compare observed characteristics around botanical landmarks from one or more samples.

Difficulties (I)

- Large number of biologically relevant plant categories (more than 300,000 *known* species).
- Large variation of patterns among fundamental features.
- Ongoing shortage of skilled taxonomists.
- However, botanical identification keys are much too complex for most non-specialists.
- Hence, an *automated* or even *partially automated* system would have considerable value.

DIFFICULTIES (II)

- Large intra-species variability and inter-species similarity.



DIFFICULTIES (III)

- *Heteroblastic* leaf development.

Related work

- Bird world
  - Part localization with humans in the loop: Wah et al.*ICCV 2011*
  - Poseltes: Farrell et al. *ICCV 2011* and Zhang et al. *CVPR 2012*
  - Bubbles game: Deng et al. *CVPR 2013*
- Dog world
  - Face part localization: Liu et al. ECCV 2012
- Insect world
  - Stacked evidence trees: Martnez-Muoz et al. CVPR 2009

LEAF WORLD

- Shape-tree matching algorithm: Felzenszwalb and Schwartz *CVPR 2007*
- Inner Distance Shape Context (IDSC): Ling and Jacobs *PAMI 2007*
- Shape and Venation features: Park et al. *Journal of System and Software 2008*
- Multi-scale curvature histograms: Kumar et al. *ECCV 2012*
- Multi-scale triangular representation: Mouine et al. *ICMR 2013*

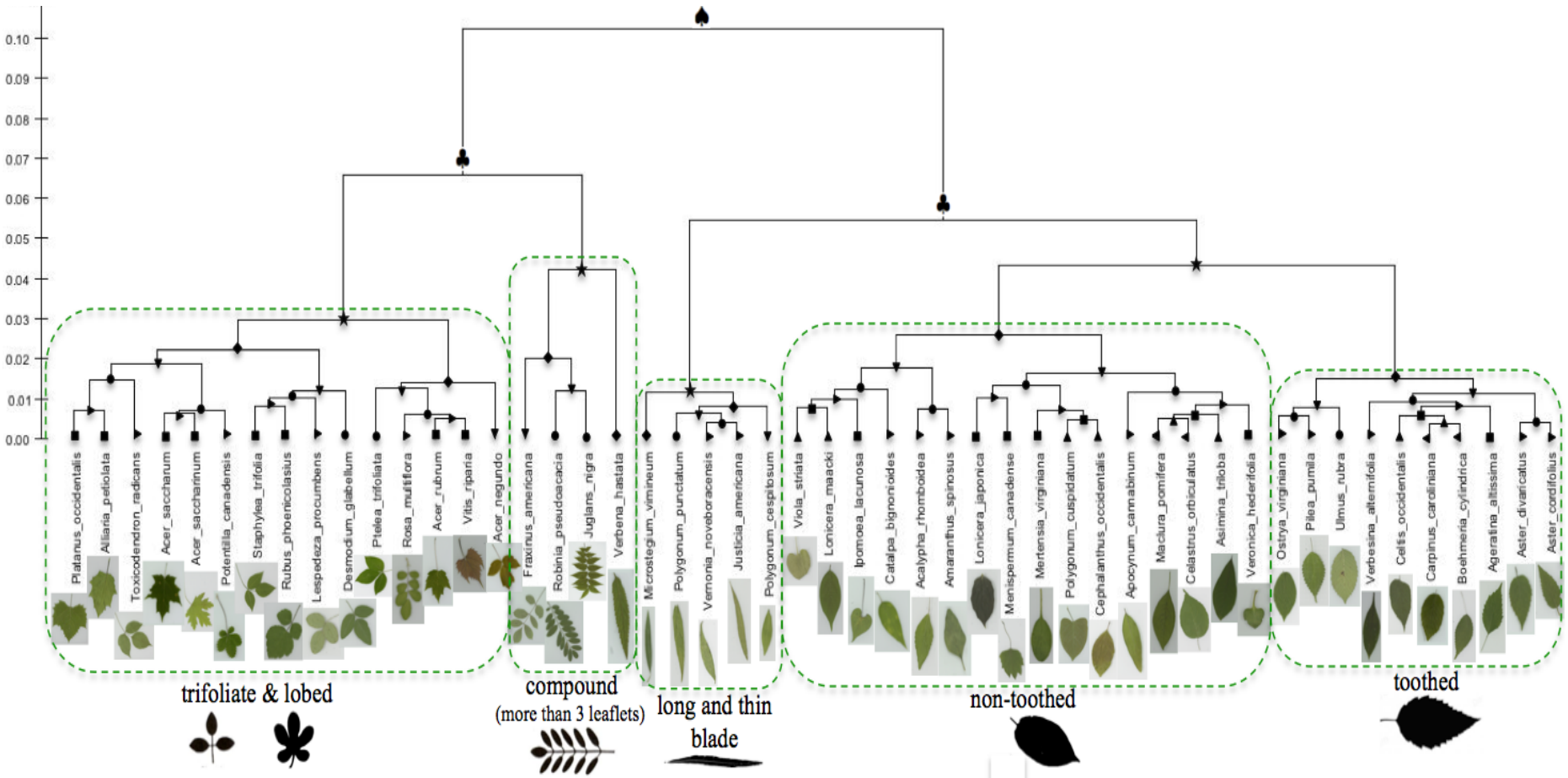
Outline

- Introduction
- Representation
- Learning
- Classification
- Conclusion

COARSE-TO-FINE

- Pre-defined species hierarchy, either
  - Handcrafted (e.g., taxonomic), or
  - By shape-based hierarchical clustering.
- A novel object representation is used to build efficient local classifiers.

EXAMPLE OF TREE-STRUCTURED HIERARCHY



Notation

- $\mathcal{Y}$: complete set of hypotheses (species).
- $Y = Y(I) \in \mathcal{Y}$: true species of image I.
- $\mathcal{T}$: tree graph.
- t: node of $\mathcal{T}$.
- $C_t \subset \mathcal{Y}$: a set of categories associated with t.
- X_t: classifier score for $Y \in C_t$ versus $Y \notin C_t$.

VANTAGE FEATURE FRAMES (I)

- A representation for building X_t.
- Motivated by the strategy used by botanists.
- Landmarks in the sense of *vantage points*.
- What to look for in a neighborhood of the landmarks may be category-dependent.
- Hence,
  - Where to look, and
  - What to compute.

VANTAGE FEATURE FRAMES (II)

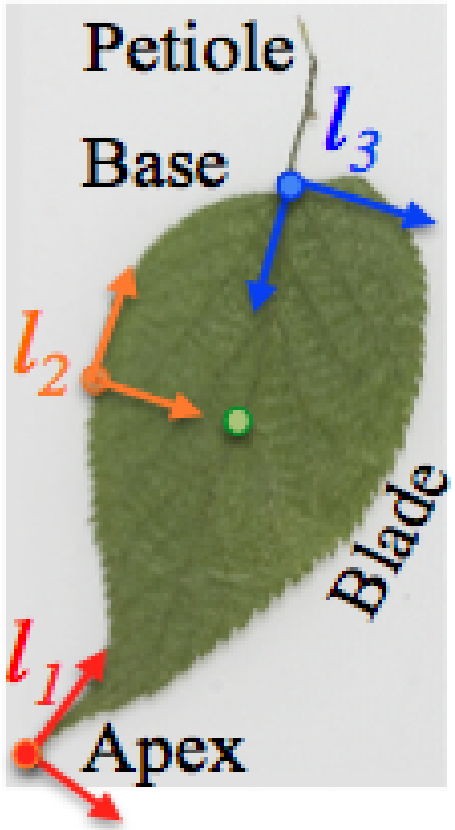
- A frame $\mathcal{F}$ has two components
  - Geometric component Θ: a *category-independent* and local coordinate system.
  - Appearance-based component $\mathcal{Z}$: a *category-dependent* family of pose-indexed features.
  - $\mathcal{Z} = \{\mathcal{Z}_1, \ldots, \mathcal{Z}_N\}$, where $\mathcal{Z}_t$ is the set of local features to compute in frame $\mathcal{F}$.
- $\mathcal{F}$ must be both *detectable* and *discriminating*.

Outline

- Introduction
- Representation
- Learning
- Classification
- Conclusion

LEARNING THE FRAMES (I)

- Leverage domain knowledge.
- Given a list of candidate origins $\{l_1, ..., l_K\}$, we associate frames with a subset of these.



LEARNING THE FRAMES (II)

- The orientation of the frame is determined by the centroid of the object (the landmark points to the centroid).
- The unit distance is the approximative scale of the object.

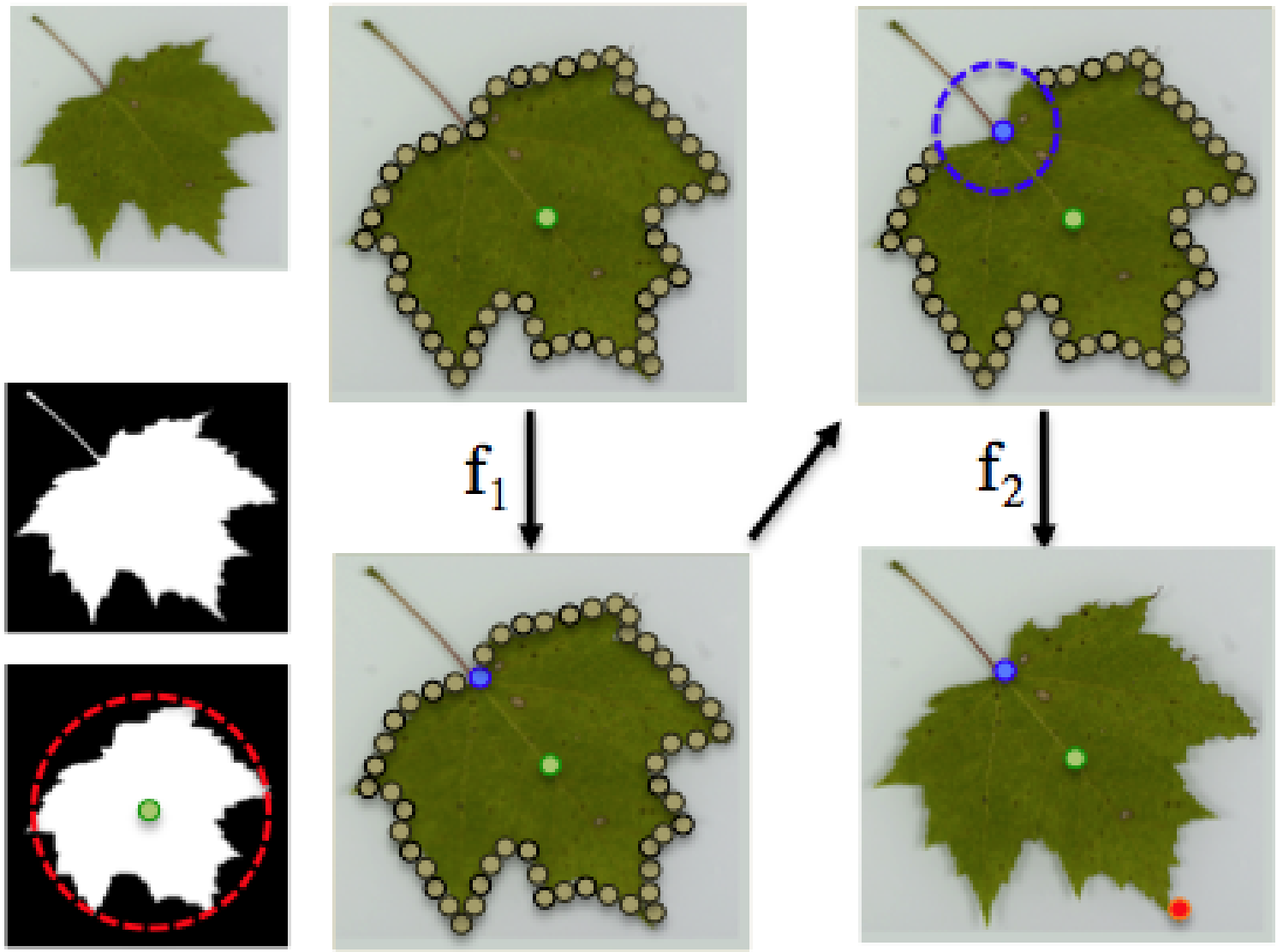
LEARNING THE FRAMES (III)

- The choice of landmarks is performance-based.
- During the learning, the locations of the landmarks are manually annotated.
- The errors in automatically detecting the landmarks are not considered in choosing the representation.
- The best performance is obtained with two frames corresponding to *apex* and *base* of the leaf.

DETECTING THE FRAMES (I)

- The orientation is determined by the centroid, which is directly computed from the raw image data after a segmentation process.
- The scale is taken to be the radius of the bounding circle.
- Each landmark is detected by a binary SVM classifier trained on manually annotated images.

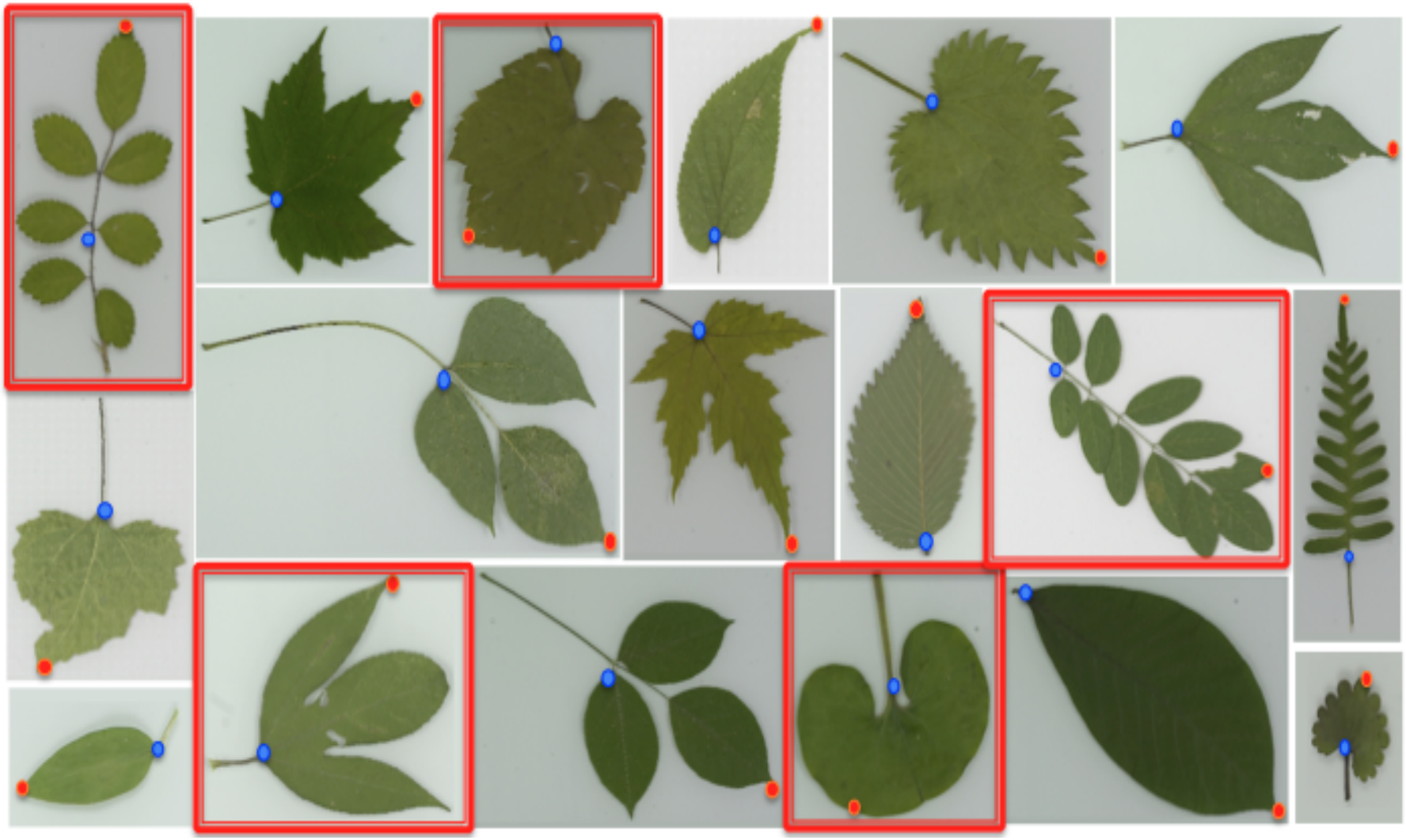
EXAMPLE



DETECTING THE FRAMES (II)

- The features for SVM learning are defined in the local coordinate system centered on the candidate landmarks.
- Invariant focusing of this nature is enabled by the type of *pose-indexed* features Z introduced first for detecting cats.
- Given a frame, there is a candidate feature $Z = Z(w, j)$ for each (local) window w in frame coordinates and image property j.
- Shape and texture were used as properties (i.e., Hough, EOH and Fourier histograms were used as base features).

SOME LANDMARK DETECTION RESULTS



LEARNING THE FEATURES (I)

- A separate binary SVM score is built for each t (i.e., for each C_t).
- Each SVM employs a learned, category-dependent subset of features $\mathcal{Z}_t$.
- Category-dependent features increase recognition performance.

LEARNING THE FEATURES (II)

- The probability distribution of each feature is estimated under both hypotheses $Y \in C_t$ and $Y \notin C_t$ from the positive and negative examples.
- For feature $Z = Z(w, j)$, denote the two distributions by $p^+_{w,j}$ and $p^-_{w,j}$.
- $d_{w,j} = |p^+_{w,j} - p^-_{w,j}|$
- $\mathcal{Z}_t$ consists of the features with the M largest differences.

Outline

- Introduction
- Representation
- Learning
- Classification
- Conclusion

CLASSIFICATION

- Given a pre-defined tree hierarchy $\mathcal{T}$ along with scores $X = \{X_t, t \in \mathcal{T}\}$ how can we estimate the species $Y(l)$ of the leaf in l as accurately as possible?
- Standard method:
  - Report a single species $\hat{Y}$.
  - Utilize Coarse-grained to fine-grained category identification.
  - Build a binary classifier for each t using X_t (we use SVMs).
  - Process the hierarchy breath-first coarse-to-fine: at each level, all the children of a *positive* node t are retained and tested at the next level.

LIKELIHOOD RATIOS

- Likelihood ratio:

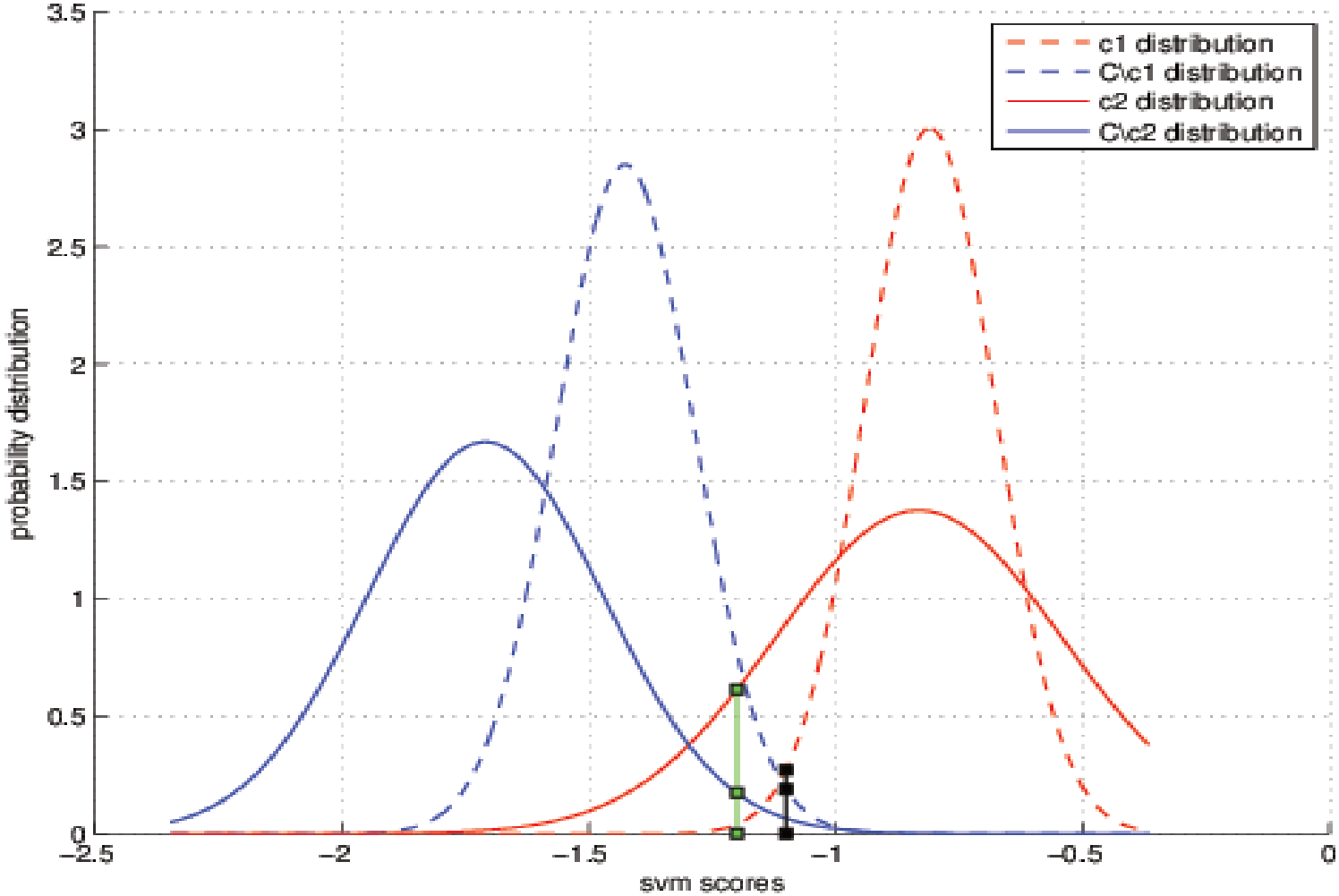
$$L_t(I) = \frac{P(X_t|Y \in C_t)}{P(X_t|Y \notin C_t)}$$

- Now threshold $L_t(I)$.
- *Positive* leaf-nodes (i.e., species) are then sorted according their likelihood ratios.

WHY LIKELIHOOD RATIOS? (I)

- The advantage of mapping the SVM score to a likelihood ratio is that it takes into account the distribution under both hypotheses.
- This mapping is *not* monotone, i.e, does not preserve the ordering of SVM scores across a level, which might naturally occur on different scales.

WHY LIKELIHOOD RATIOS (II)



Results (Swedish)

- Data: 15 Swedish species (1125 leaf images).

- One-third of images for training (375 images)
- Two-third for testing (750 images)

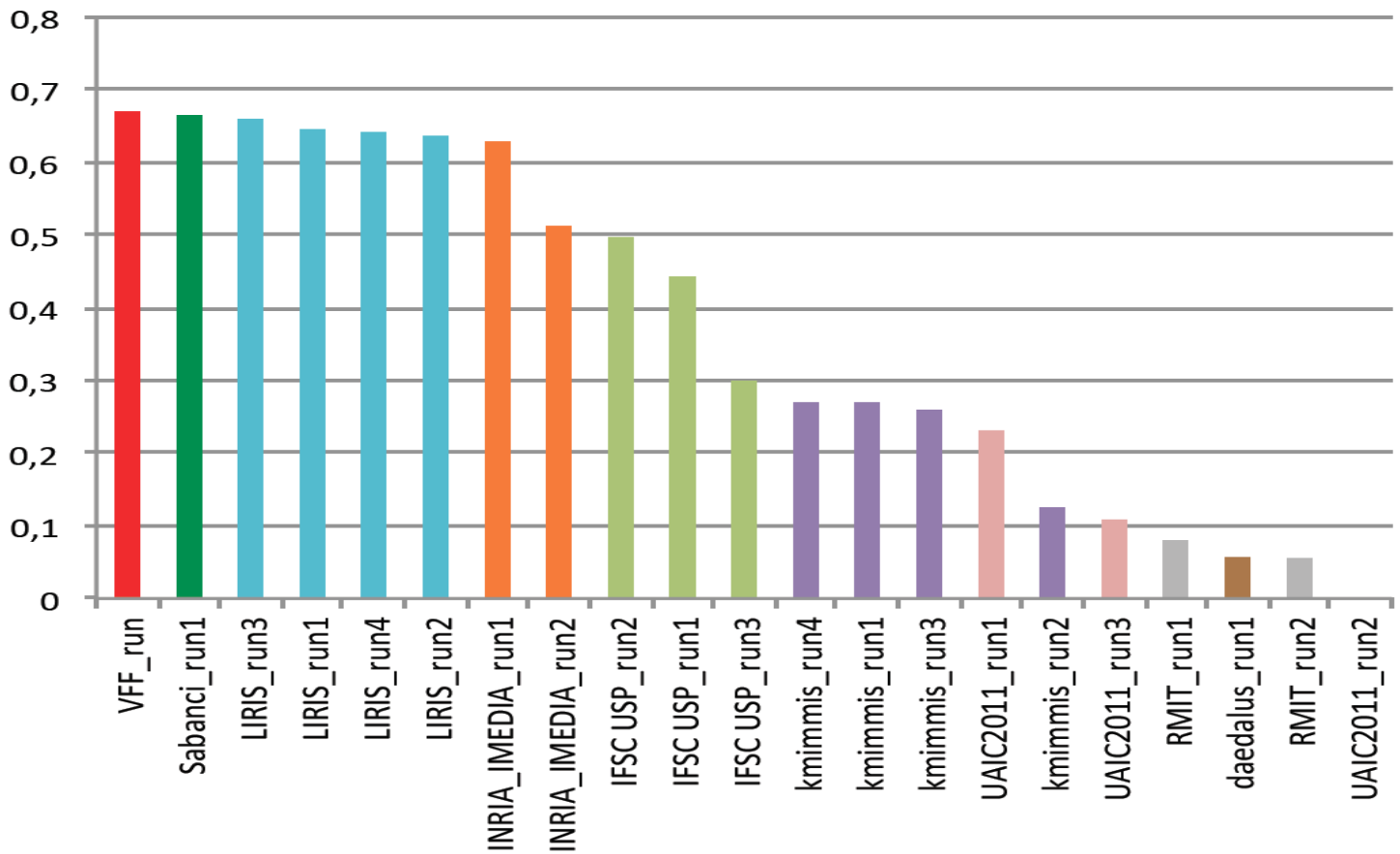
Results (Swedish)

Table: Different results on the Swedish data

Methods	Perf. (top-1)
Ours (taxonomic hierarchy)	**98.4%**
sPACT	97.92%
TSLA (triangular representation)	96.53%
Shape-Tree	96.28%
IDSC	94.13%
Shape Context	88.12%
Söderkvist	82.40%

RESULTS (IMAGECLEF)

- Data: 46 species (scanned simple ImageClef2011 leaves).
- Comparison between the ImageClef2011 score of our method using a taxonomic hierarchy and those of the 8 participants at ImageClef2011.



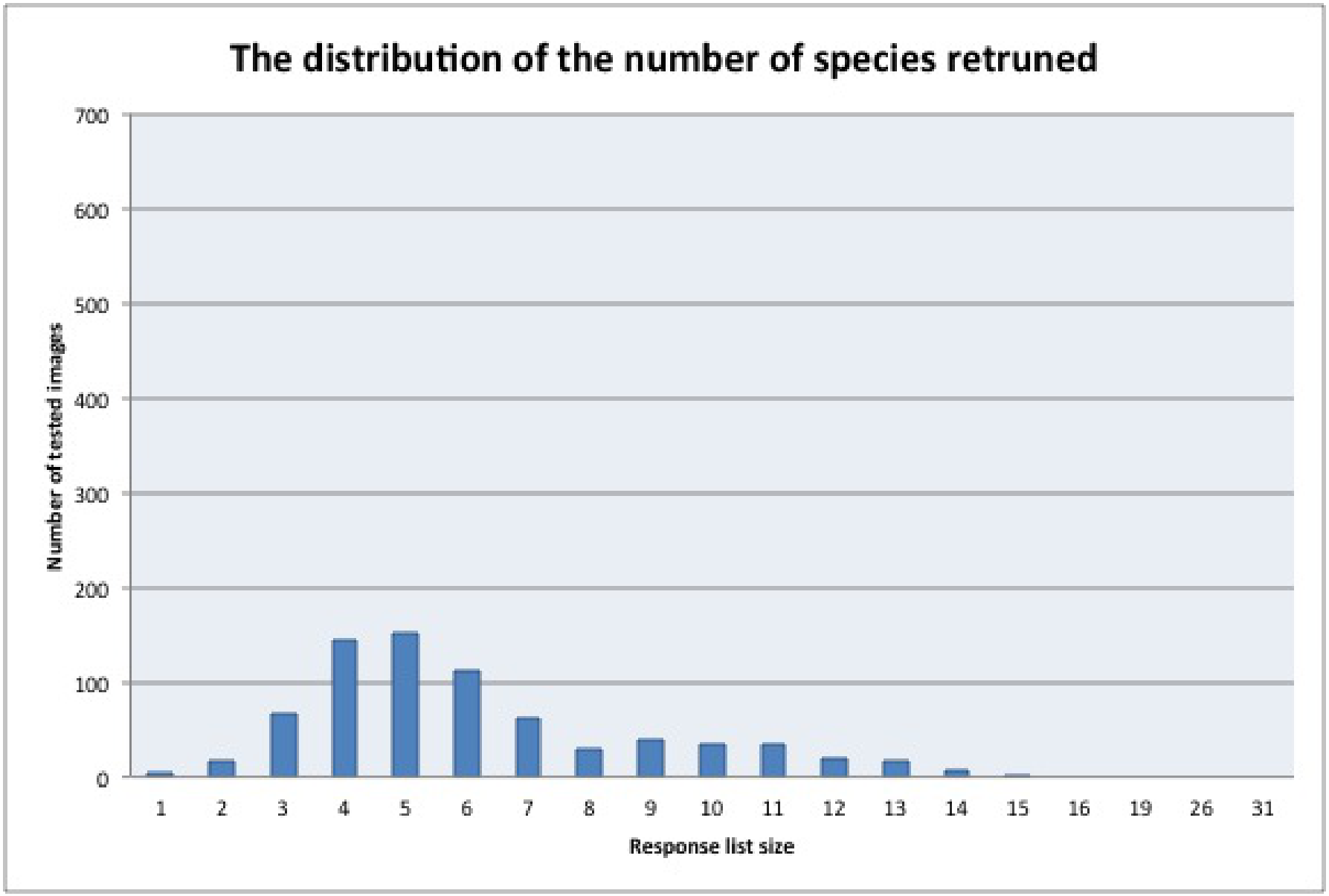
RESULTS (SMITHSONIAN)

- Data: 50 Smithsonian species (2160 leaf images).

- Two-third of images for training (1426)
- One-third for testing (734 images)
- Performance of shape-based hierarchy:

top-1	top-2	top-3	top-5	top-10	top-15
66%	81%	87%	92%	94%	94.5%

CLASSIFICATION RESULTS (CONT)



SET-VALUED CLASSIFICATION (I)

- Motivated by applications, we consider another scenario:
- Suppose a specialist is perfect without any assistance from a computer vision, i.e., could determine the true species from one or more images from the same plant.
- Instead of providing a single predicted species $\hat{Y}$, we report a set $\hat{C} \subset \mathcal{Y}$ of species and the human then examines $\hat{C}$.
- The key constraint is then $P(Y \in \hat{C}) \geq 1 - \epsilon$
- Performance criterion: minimize $E|\hat{C}|$ subject to the constraint.
- Context: constructing $\hat{C}$ will take into account *all the local scores* simultanesously?

SET-VALUED CLASSIFICATION (II)

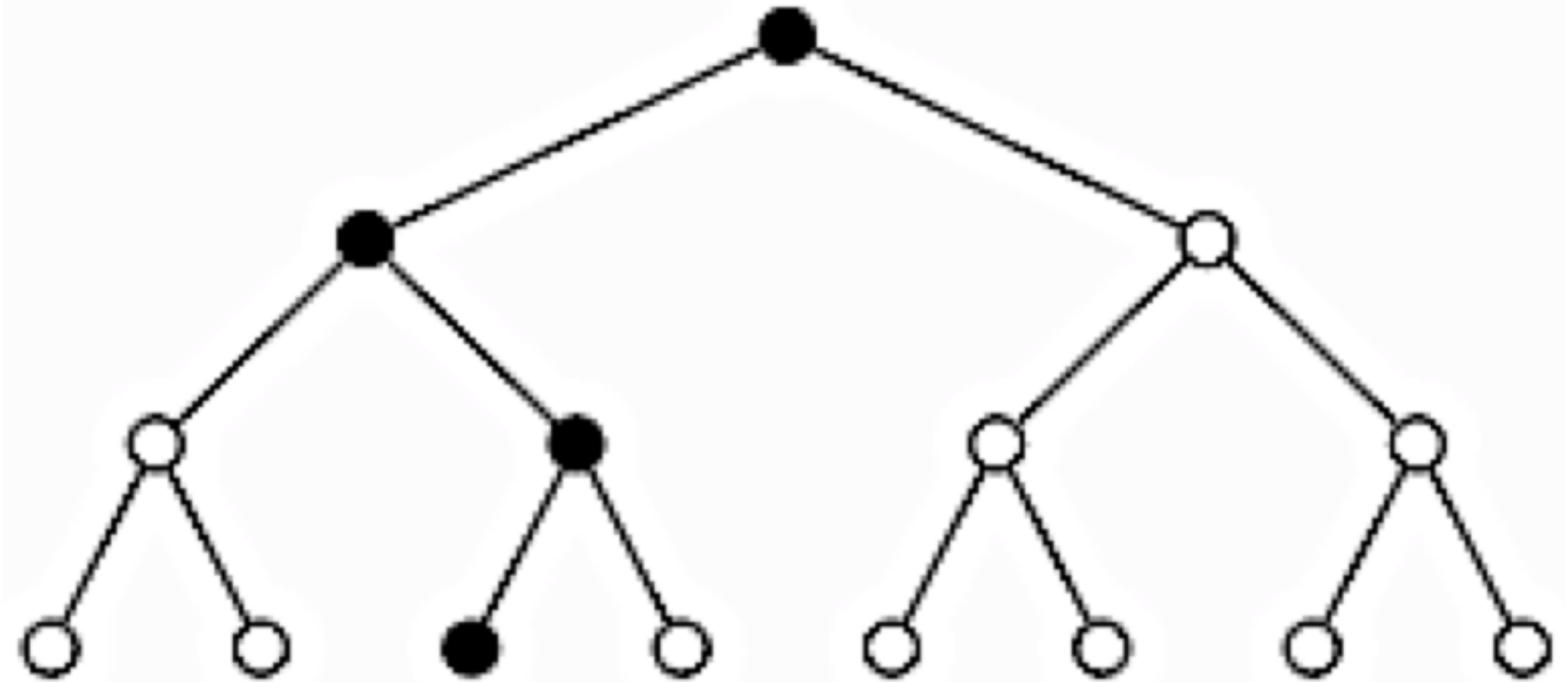
- Restrict $\hat{C}$ to $\{C_t, t \in \mathcal{T}\}$.
- Obviously we need to compute $P(Y \in C_t|X)$ where $X = \{X_s\}_s$.
- $f(X|Y = s)$: density of X given $Y = s$, estimated from data.
- Assuming conditional independence of scores Y:

$$P(Y \in C_t|X = x) = \frac{\sum_{s \in C_t} f(X|Y = s)}{\sum_{s \in \mathcal{T}} f(X|Y = s)} = \frac{\sum_{s \in C_t} \prod_{r \in \mathcal{T}} f(x_r|Y = s)}{\sum_{s \in \mathcal{T}} \prod_{r \in \mathcal{T}} f(x_r|Y = s)}$$

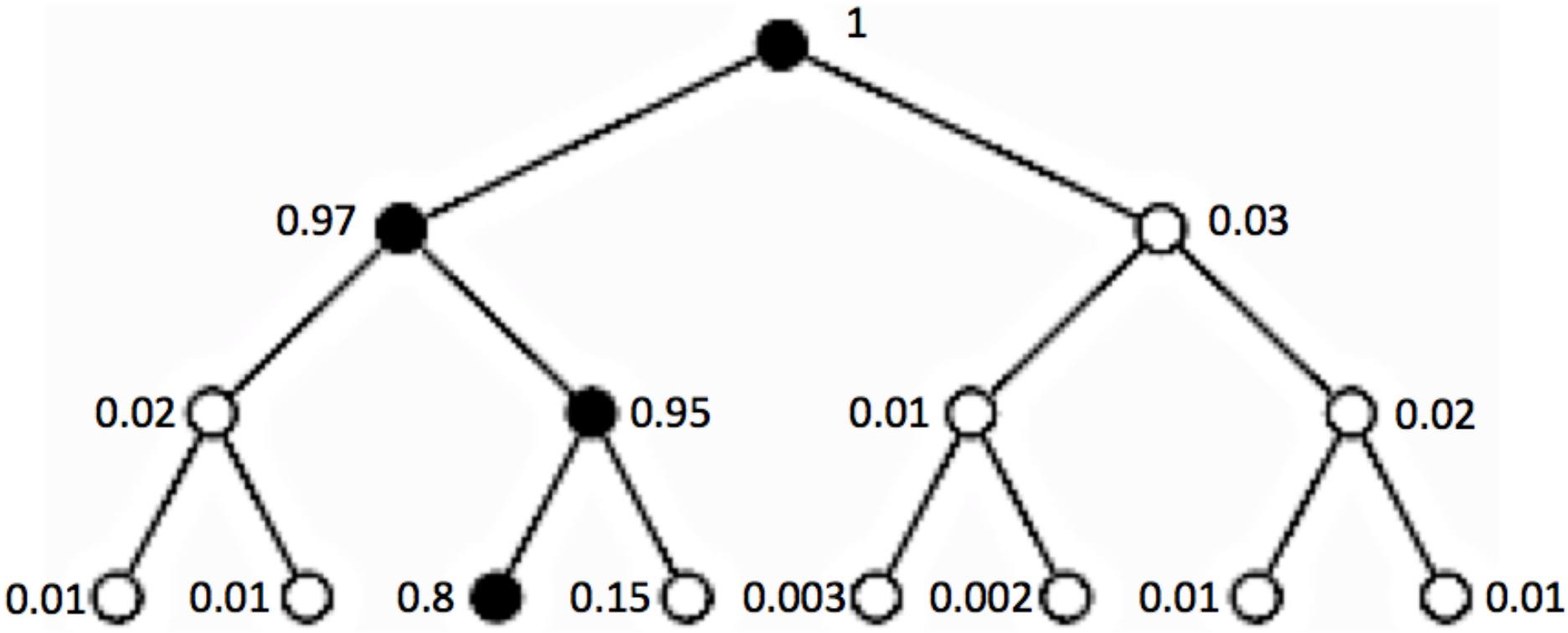
SET-VALUED CLASSIFICATION (III)

- Define $B(x) = \{t \in \mathcal{T} : P(Y \in C_t | X = x) \geq 1 - \epsilon\}$.
- Note: $B(x)$ is necessarily a sub-path in $\mathcal{T}$ originating at the root.
- $X \rightarrow t(X) = argmin_{t \in B(X)} |C_t|$

Toy example (I)

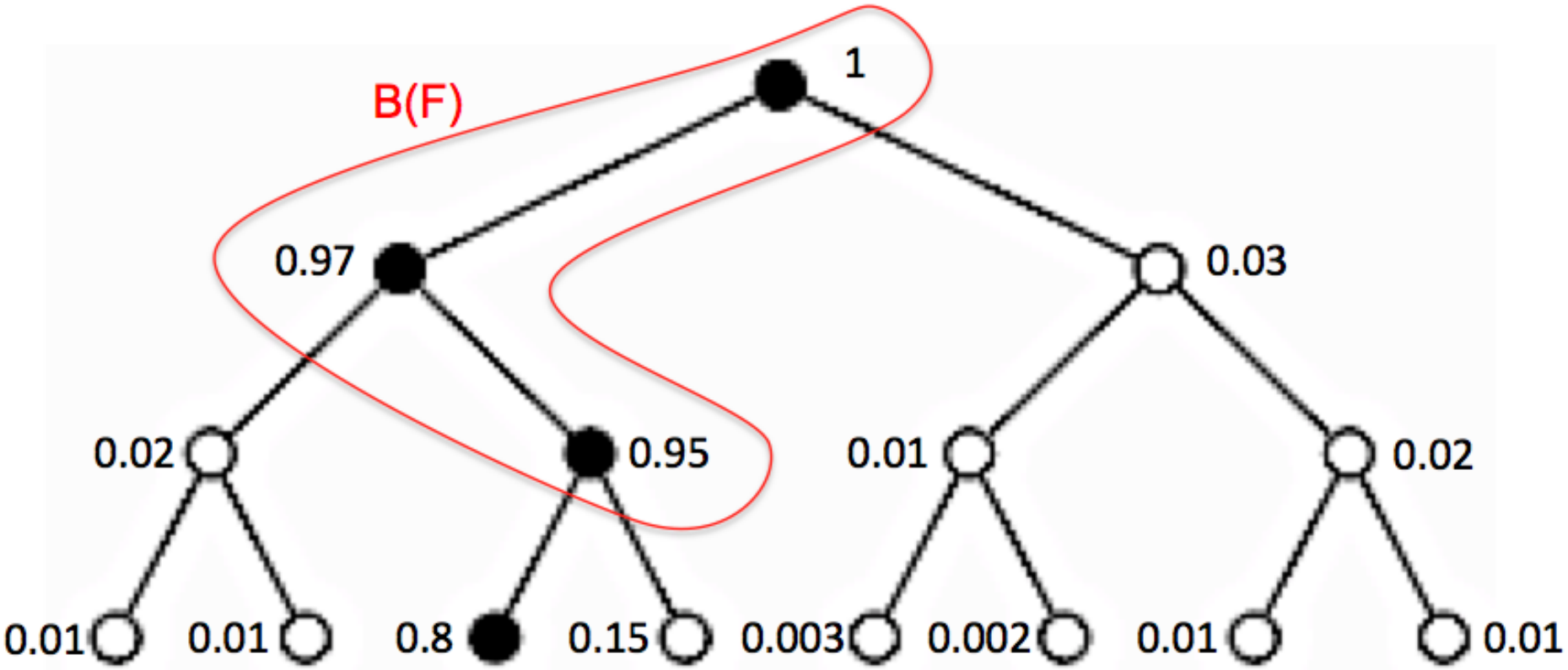


Toy example (II)



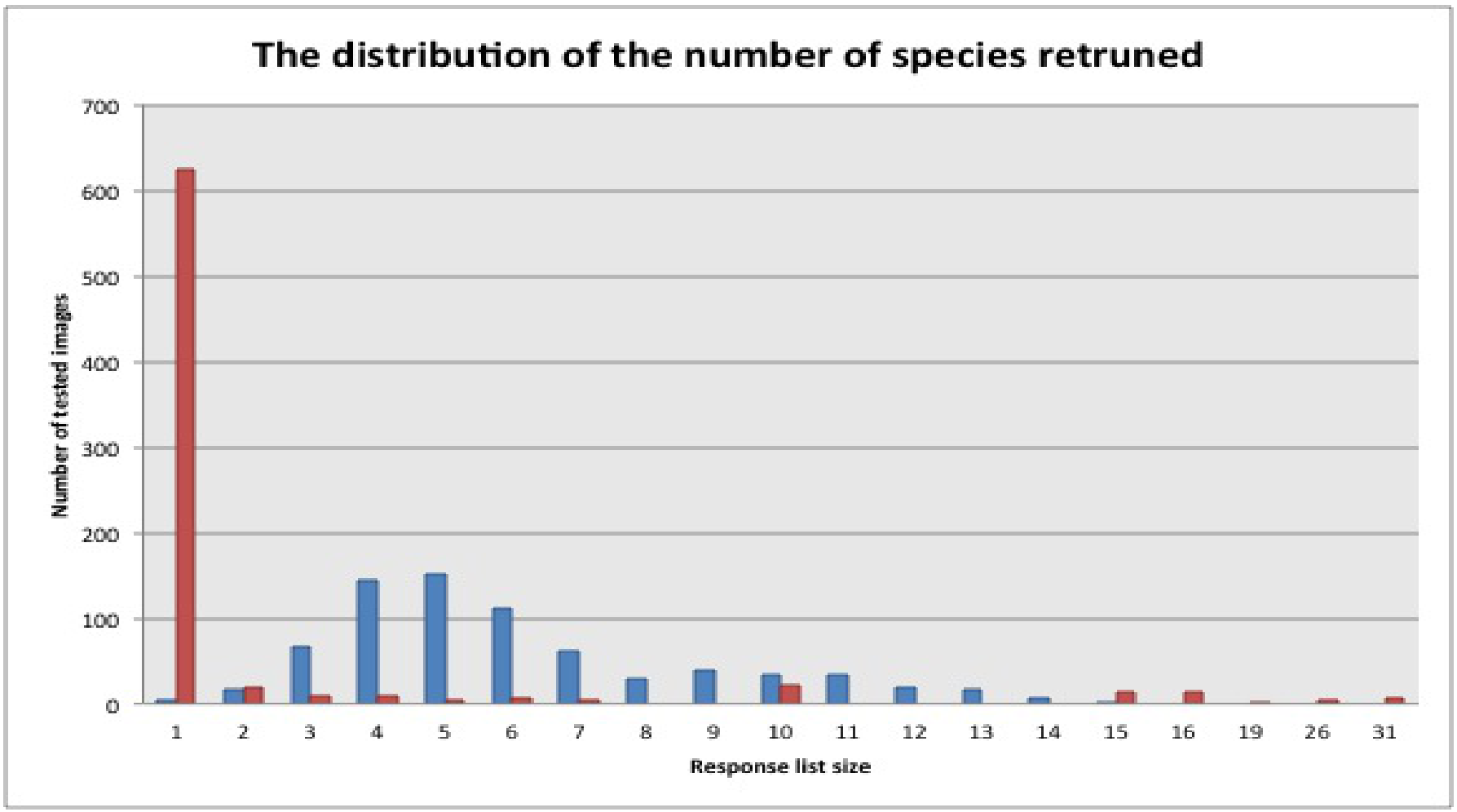
Toy example (III)

$\epsilon = 0.05$



RESULTS (I)

- Data: 50 Smithsonian species.



Results (II)

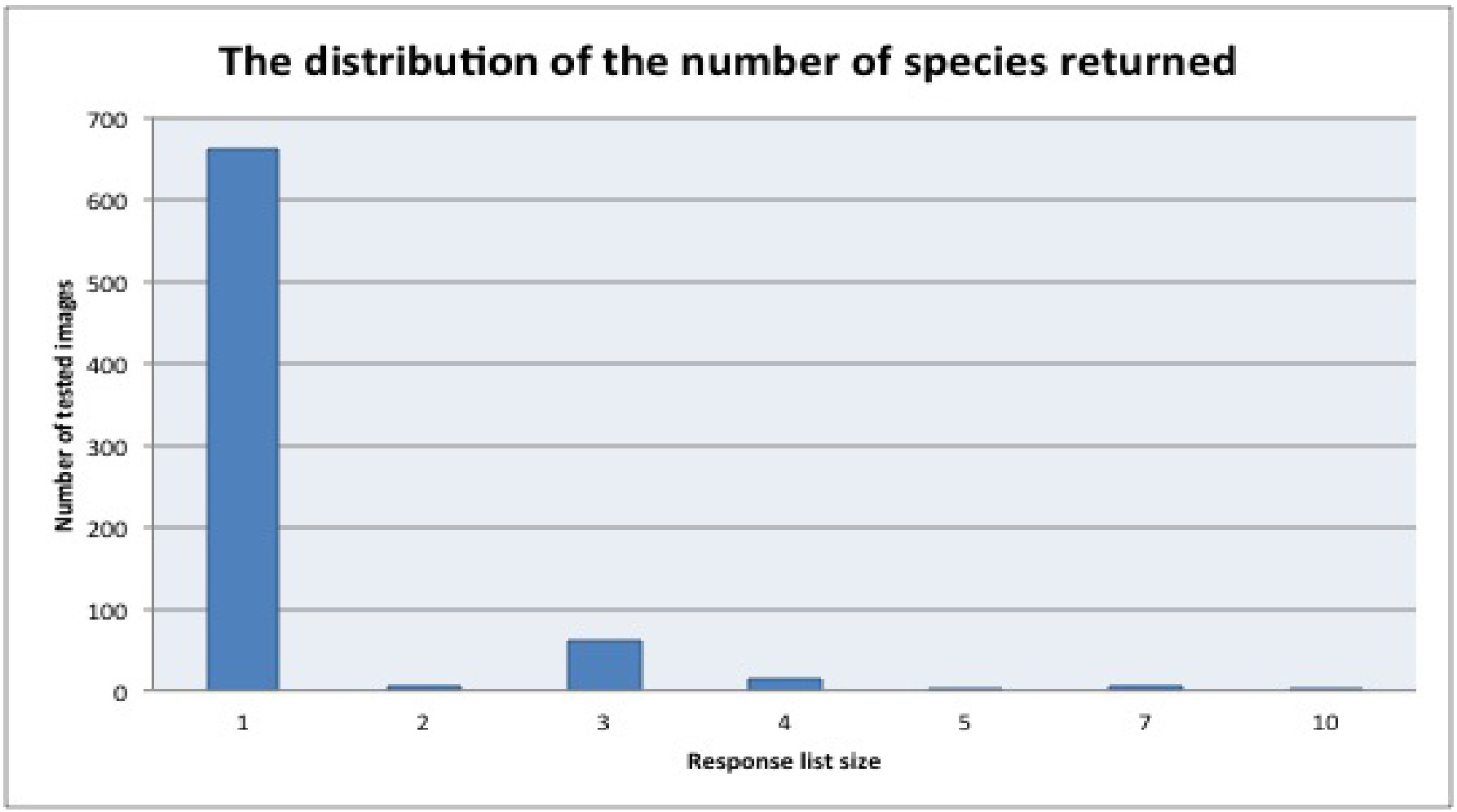
Accuracy	Average size of the response
94%	2.4

- If the species are simply sorted according their SVM scores:

top-1	top-2	top-3	top-5	top-10	top-15
89%	93%	93.5%	94%	94%	94%

RESULTS (III)

- Data: 15 Swedish species.



Results (IV)

Accuracy	Average size of the response
99.6%	1.3

- If the species returned are sorted according their SVM scores, we have the following performances

top-1	top-2	top-3	top-5
98.8%	99.6%	99.6%	99.6%

Outline

- Introduction
- Representation
- Learning
- Classification
- Conclusion

CONCLUSION

- Vantage feature frames provide the cues needed to distinguish between closely-related categories such as plant species.
- Works as well as detailed boundary analysis for standard classification.
- For applications (e.g., for botanists), reporting Y along with other species is more valuable than reporting $\hat{Y} \neq Y$.
- The same framework could be extended to be able to consider several leaf images of the same plant.
- The more ambitious problem is to classify from more challenging images.

NATURAL PHOTOS