Constructing Category Hierarchies for Visual Recognition

Marcin Marszałek (marszale@inrialpes.fr)
INRIA Grenoble, LEAR / LJK

Cordelia Schmid (schmid@inrialpes.fr)
INRIA Grenoble, LEAR / LJK

Abstract

We evaluate class hierarchies currently constructed for visual recognition. We show that top-down as well as bottom-up approaches, which are commonly used to automatically construct hierarchies, incorporate assumptions about the separability of classes. Those assumptions do not hold for visual recognition of a large number of object categories. We therefore propose a relaxation which postpones decisions in the presence of uncertainty. This results in higher recognition accuracies and retains the sublinear complexity of the hierarchical approach.

1. Introduction

Hierarchical classification scales well in the number of classes:

$O(n^2)$ one-against-one
$O(n)$ one-against-rest

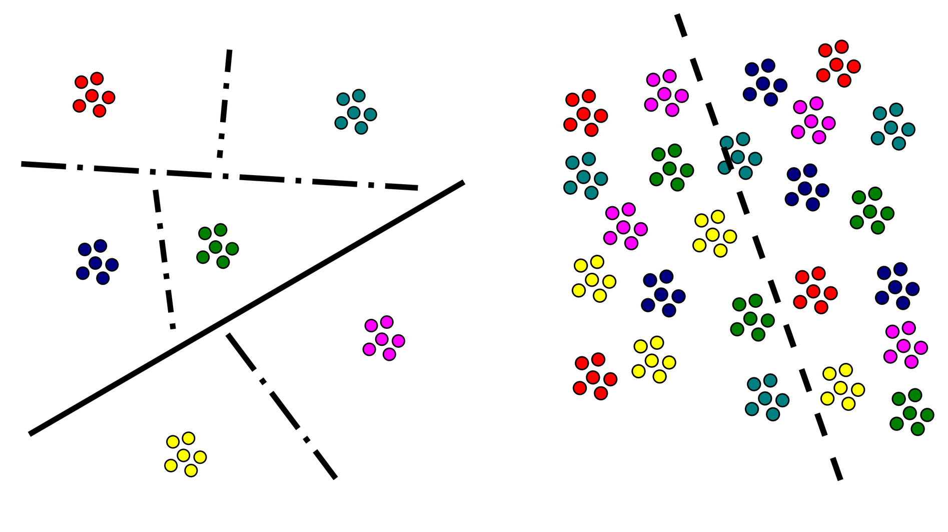
$O(\log n)$ classification tree

Various methods to construct class hierarchies exist:

- By hand [Zweig'07]
- From external sources [Marszałek'07]
- From visual similarities
  - Exhaustive [Yuan'06]
  - Top-down [Chen'04, Griffin'08]
  - Bottom-up [Zhigang'05, Griffin'08]

One common problem: disjoint trees

Disjoint partitioning of a set of classes might be easy for a few separated categories, but is increasingly difficult when the number of classes increases.



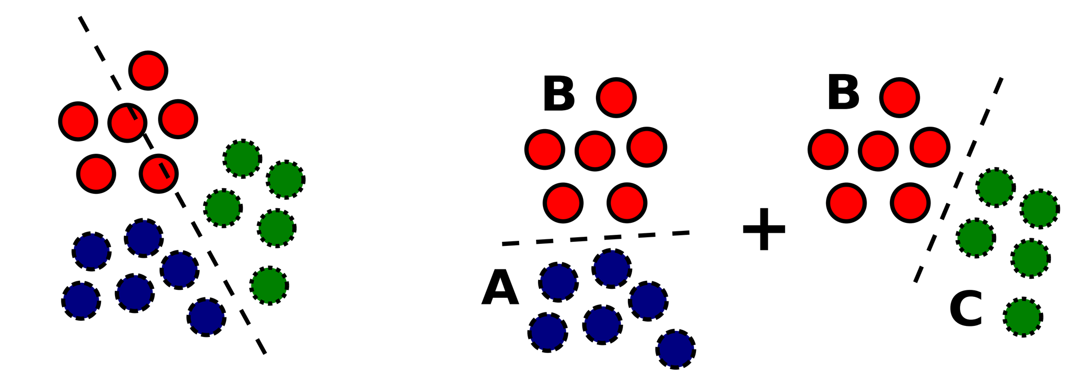
a) few classes b) many classes

Conflicting requirements for high-level partitioning:

- generalization (e.g., *animals* vs *man-made objects*)
- precision (e.g., *bear* vs *teddy bear*)

2. Method

We propose a relaxation of the disjoint separation constraint:



a) separation problem b) relaxed solution

A disjoint bi-partitioning of samples $S = A \sqcup B$ leads to a disjoint tri-partitioning of classes $\mathcal{C} = \mathcal{A} \sqcup \mathcal{X} \sqcup \mathcal{B}$ such that:

- $\mathcal{A}$ have all samples in A
- $\mathcal{B}$ have all samples in B
- $\mathcal{X}$ have samples in both partitions

We propose to split the set of classes $\mathcal{C} = \mathcal{L} \cup \mathcal{R}$ such that:

$$\mathcal{L} = \mathcal{A} \cup \mathcal{X} = \{C : \exists_{s \in A}[s] = C\}$$
$$\mathcal{R} = \mathcal{B} \cup \mathcal{X} = \{C : \exists_{s \in B}[s] = C\}$$

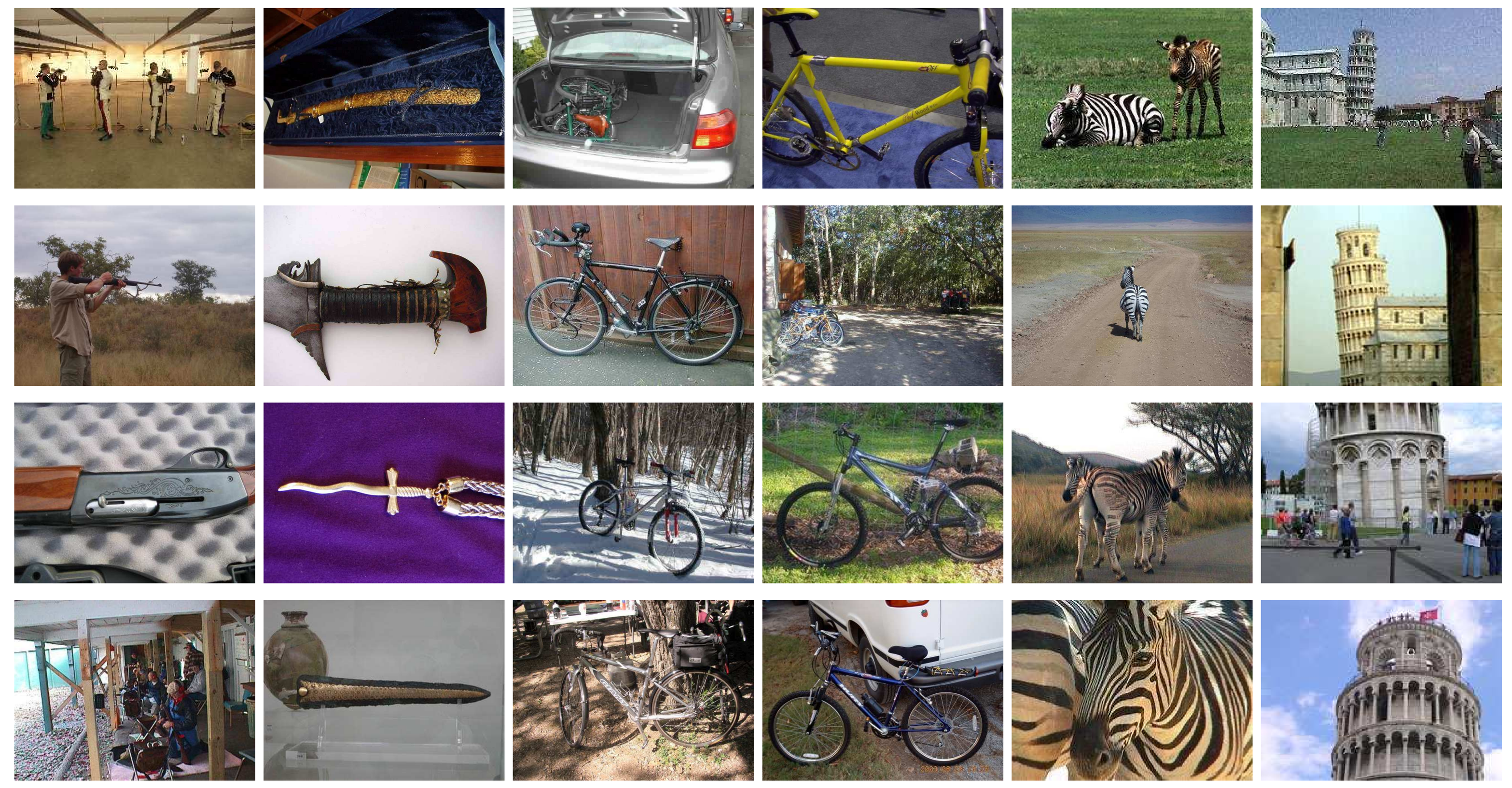
Furthermore, we note that a partitioning $p : S \to \{-1, 1\}$ determines a soft assignment of classes $q : \mathcal{C} \to [-1, 1]$:

$$q(C) = \frac{1}{|C|} \sum_{s \in C} p(s)$$

This allows to control the strength of the relaxation by defining:

$$\mathcal{L} = q^{-1}\big([-1, 1-\alpha)\big)$$
$$\mathcal{R} = q^{-1}\big((-1+\alpha, 1]\big)$$

3. Results



rifle	sword	touring-bike	mountain-bike	zebra	tower-pisa
(0%)	(0%)	(37%)	(45%)	(86%)	(91%)

a) most difficult b) most confused c) easiest

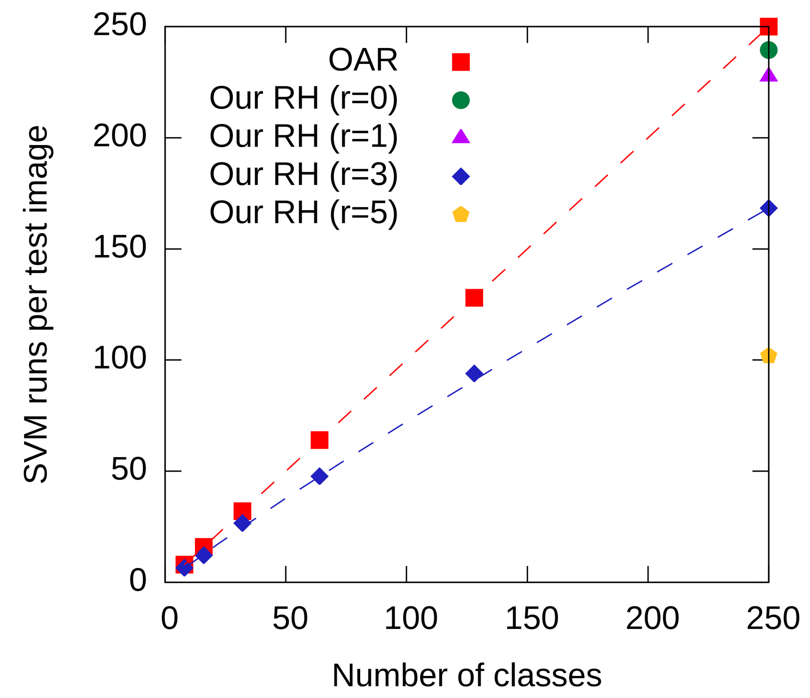
Sample images of Caltech-256 categories.

Zhang et al. [IJCV'07]	23.6%
Our RH ($\alpha = 0$, sparse IPs)	23.4%

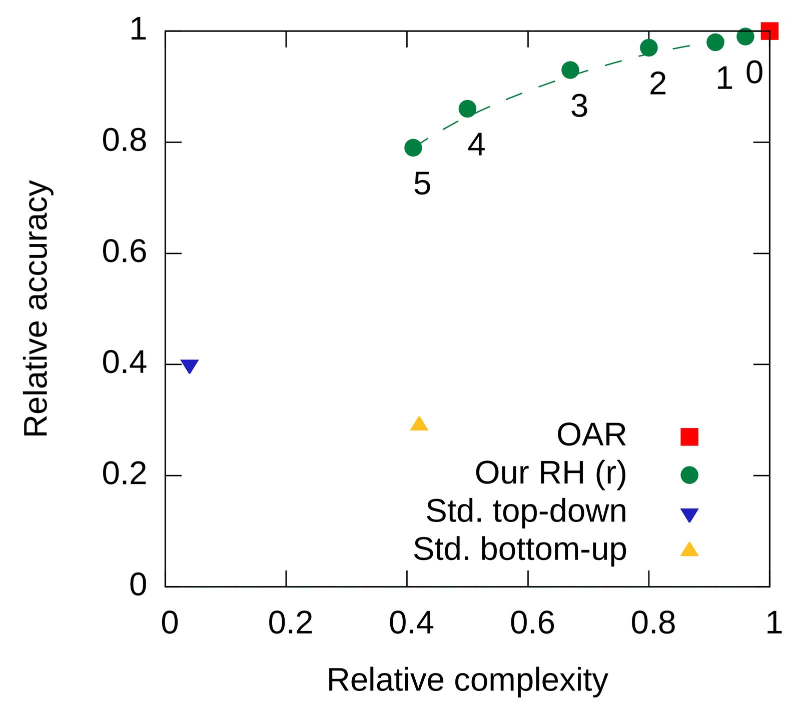
Lazebnik at al. [CVPR'06]	28%
Our RH ($\alpha = 0$, dense/grid)	27.9%

Average per-class accuracy on Caltech-256.

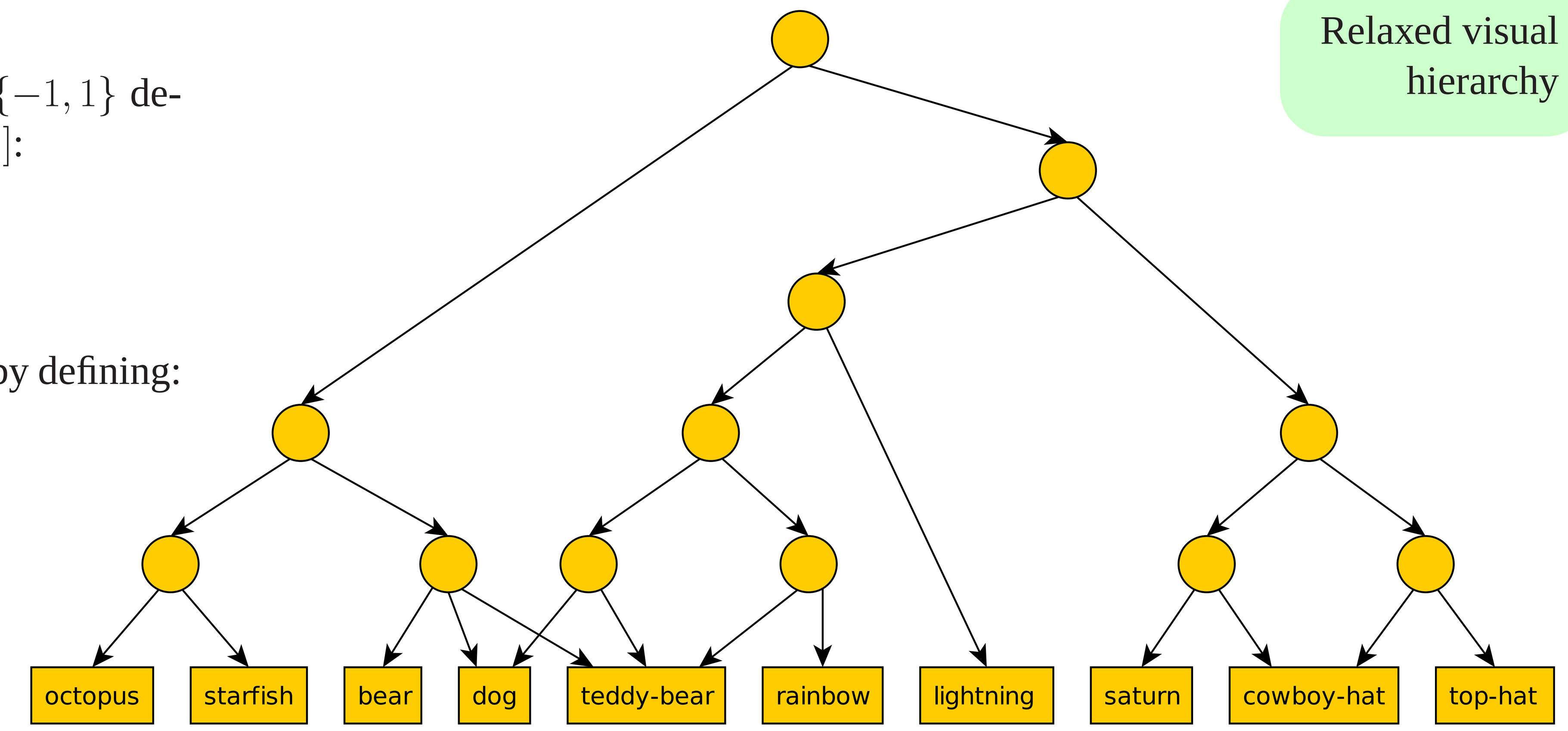
Complexity in the number of classes.



Speed-for-accuracy trade-off.



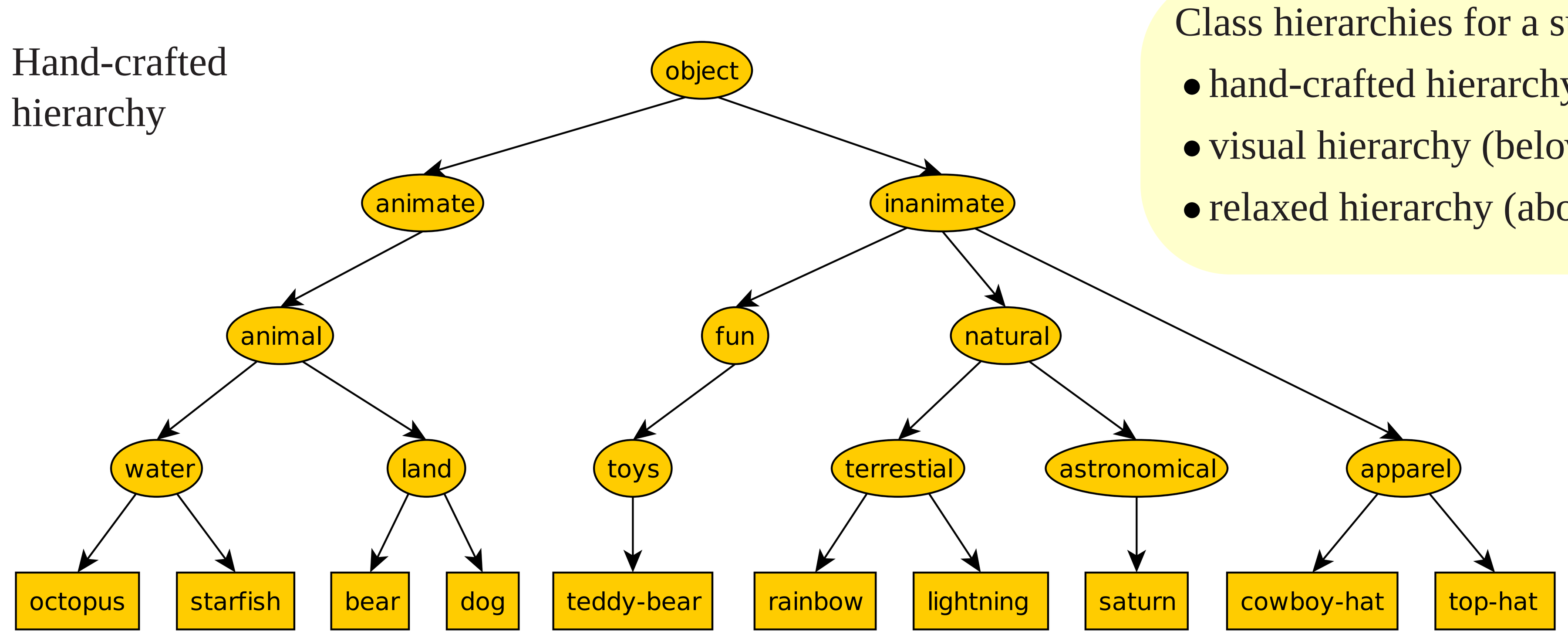
Relaxed visual hierarchy



Class hierarchies for a subset of 10 classes:

- hand-crafted hierarchy (below, left)
- visual hierarchy (below, right)
- relaxed hierarchy (above)

Hand-crafted hierarchy



Disjoint visual hierarchy