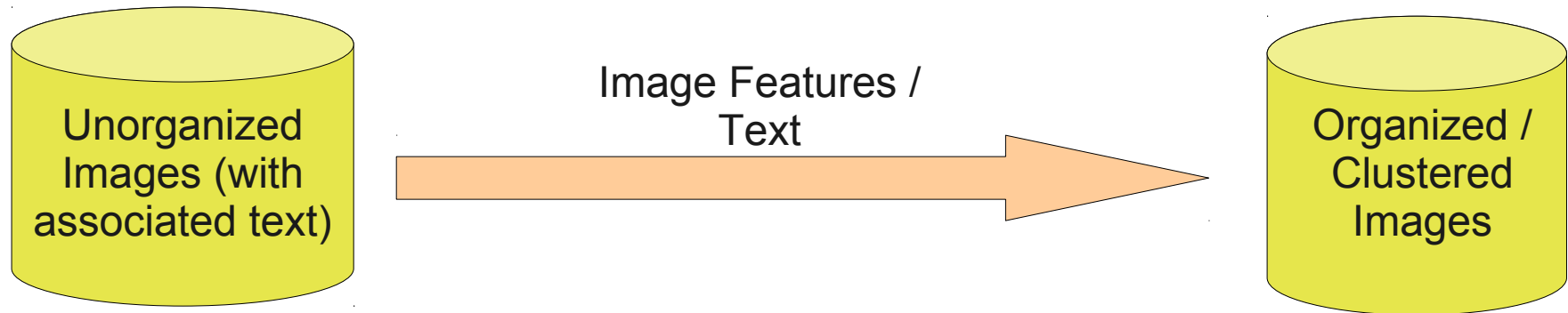


**Learning the semantics of
words and pictures
– Barnard, Forsyth**

**Presented By
Nikhil Patwardhan**

Overview



Why?

- Better browsing and searching (CBIR)
- Associating words and pictures
- Unsupervised learning

What's unique?

- **Combining image features and text** to organize the image database, instead of just one.
 - Blobworld is not sophisticated enough
- Using a **hierarchical** approach enables **browsing** through the collection.
 - General to specific

Model Overview

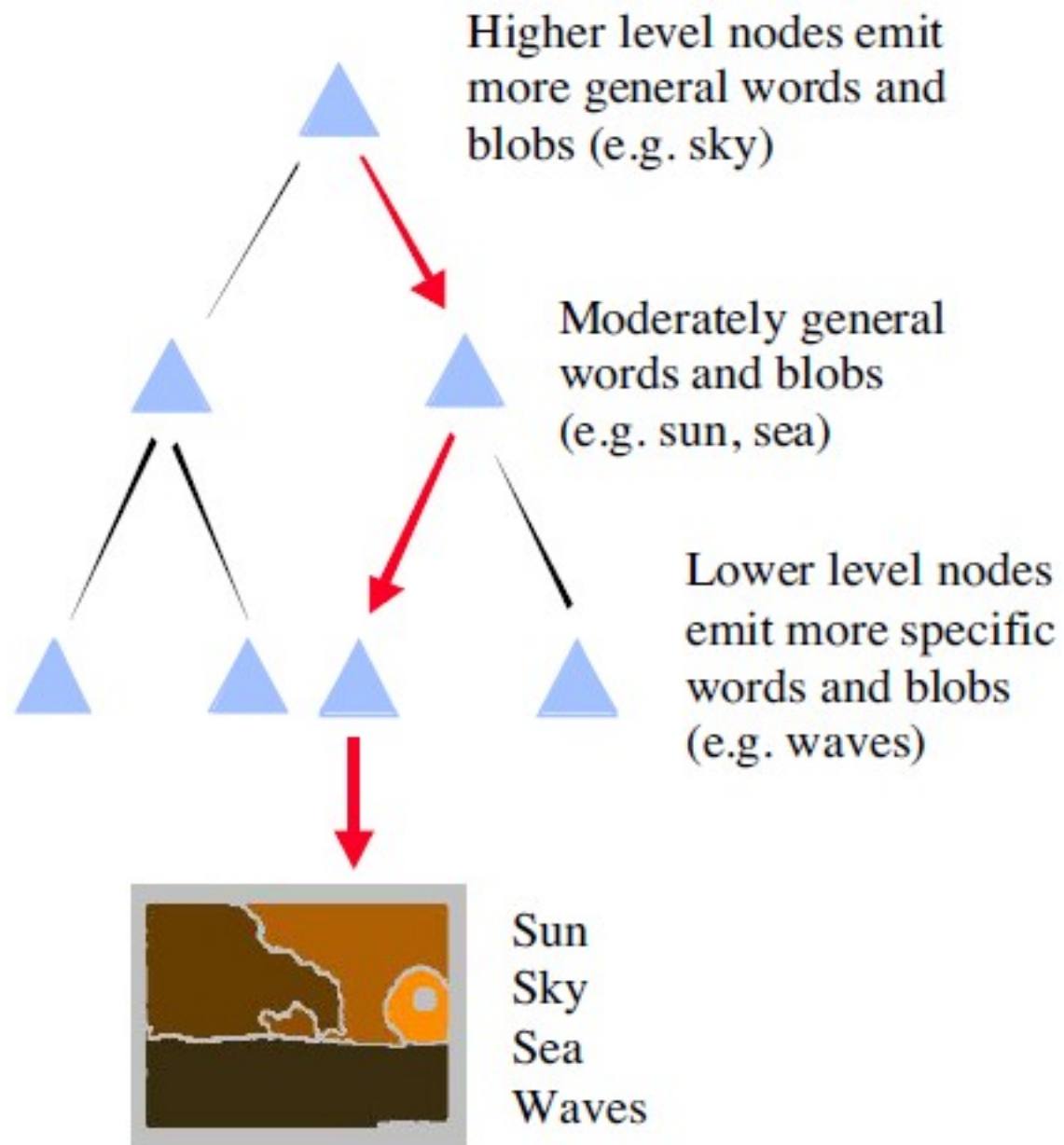
- Hierarchical model based on work by Thomas Hoffman
 - Generative Hierarchical Model
 - Uses Expectation-Maximization algorithm to train it
 - Ideal for browsing the images based on themes / sets of keywords
- Each node == $p(\text{word})$, $p(\text{segment})$

Expectation - Maximization

- Likelihood = function of the parameters of a statistical model, indicating $p(X|\Theta) \sim L(\Theta|X)$
- EM Step 1: Estimate hidden data
- EM Step 2: Maximize the likelihood function

Model Creation

- Divide all images into segments/items (i)
- Compute a single distribution histogram for all the items in all images (items v/s occurrences)
- Apply clustering on the histogram to get discrete “levels”
- Get a hierarchical tree by using probability threshold / fixed fan out.
- Result = most occurring items are at the top. The leaves are the clusters.



Considerations

- Clustering results depend on the starting point.
- Clustering based on image features and text is better.
 - Example on next page.
- Browsing – do visual clusters make sense?
- Searching the database
 - Using a formula that returns the probability of an image being associated with a certain query. Return high probability items to the user.

Text =
"Ocean"

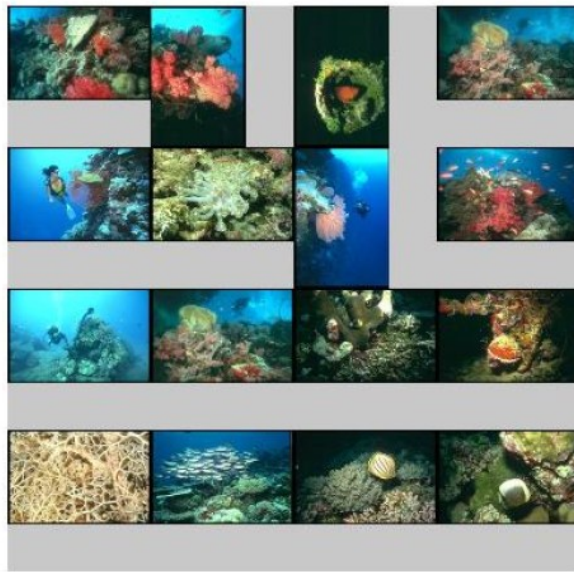


Figure 2. Some of the images from an ocean theme cluster found by clustering on text only. This cluster contains most the images in the two clusters in Figure 4, but the red corals are mixed in with the other more general ocean pictures.

Text =
"Ocean" +
Visual
Clustering

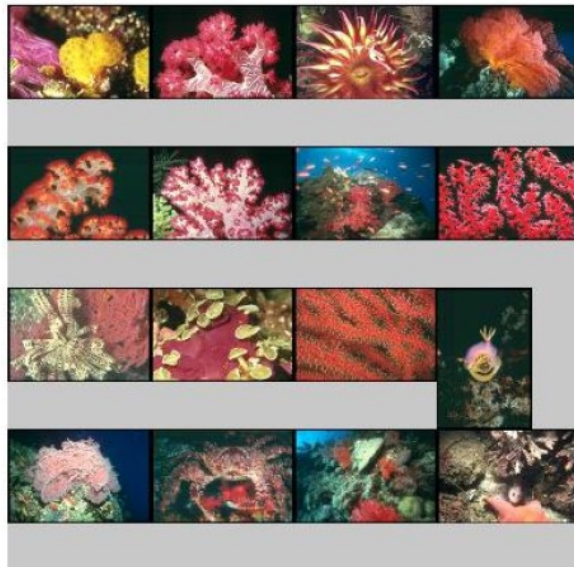


Figure 4. Two adjacent clusters computed using both text and image segments. The words clearly help ensure consistency in overall ocean/underwater themes, as well as making it so that the clusters are in neighboring leaves of the tree. Using image segment features in the clustering promotes visual similarity among the images in the clusters, and here ensures that the red coral is separated from the other ocean scenes which are generally quite blue.

Visually
clustering
of "red"
flowers and
corals



Figure 3. An example of a cluster found using image features alone. Here the coral images are found among visually similar flower images. Clearly some semantics are lost, although the grouping across semantic categories is interesting.

Model Use

- A document = sequence of words + segments
- $P(D|d)$ = probability of observations 'D' in a document 'd'

$$P(D|d) = \sum_c P(c) \prod_{i \in D} \left(\sum_l P(i|l,c) P(l|c,d) \right)$$

Pictures from Words and Words from Pictures

- Words → Pictures: Auto-illustrate text
- Pictures → Words: Auto-annotate
 - Machine learning (airplane → sky)
 - Ability to test performance (predict and check)

Limitations

- Some words just don't have any visual representations.
- Semantically meaningful segmentations are not always possible.