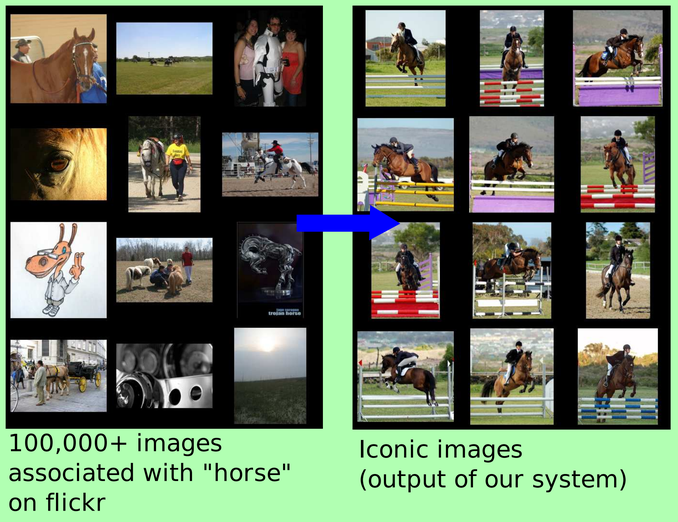
Finding Iconic Images

Tamara L Berg, Alexander C Berg
Yahoo! Research

Goal



100,000+ images associated with "horse" on flickr

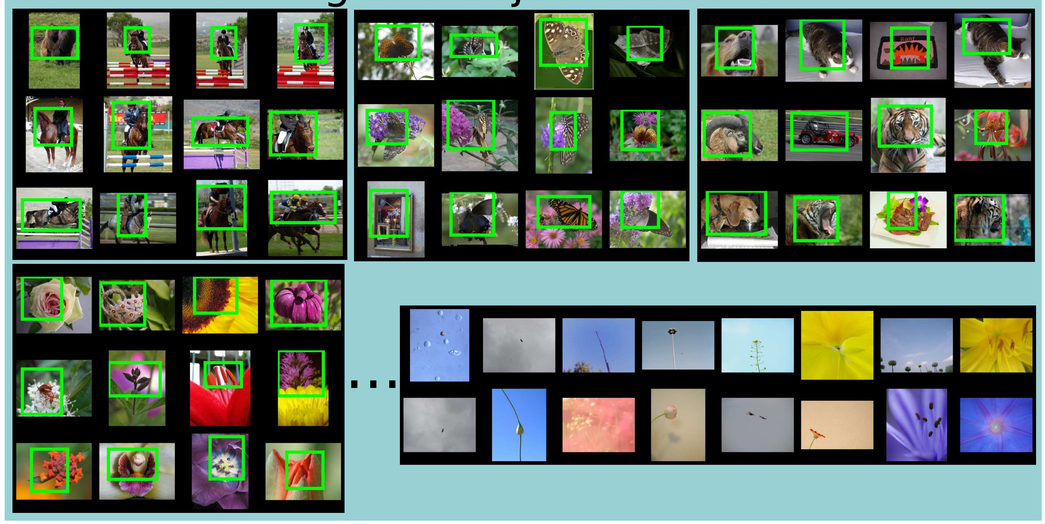
Iconic images (output of our system)

Iconic?

1. Large salient object clearly delineated from the background.

2. Object looks like many other objects in the collection.

Salient Ranking and object detection



Iconic Images

1. Ranking by Saliency

* Percolates good/interesting images up in the ranking
* Provides a rough division of the image into object/ background
* Eliminates junk images that can confuse clustering
* Generic with respect to object class
* Linear in the number of images, so can be performed efficiently on web-scale datasets

Naive Bayes Classifier

trained on a general set of images with and without salient objects

$$P(L|F) = \frac{P(L)\prod_i P(F_i|L)}{P(F_1, F_2, ...F_n)}$$

$$= \frac{P(L)\prod_i P(F_i|L)}{P(L)\prod_i P(F_i|L) + P(\overline{L})\prod_i P(F_i|\overline{L})}$$

For each image:

* Compute saliency measure of all possible layouts, divisions into fg rectangle and bg.
* Select the best layout.

Rank images by probability of selected layouts and keep the top 1000.

Features

Contrast (chi-squared dist) of fg to bg cue values

Cues:

Computed efficiently using summed-area tables.

Hue, Saturation, Value - histograms with 11 bins

Texture - histogram of total response to a set of 5 oriented filters and one center-surround filter.

Focus - ratio of high pass to low pass energy

Size/Location - 4d histogram

2. Appearance Clustering

Features

Geometric Blur - local shape descriptor (A.C. Berg & J. Malik, CVPR '01)

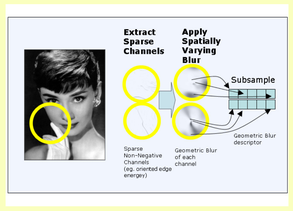


Image Similarity

mean best feature match (spatially restricted)

$$S(i,j) = \frac{1}{n}\sum_k max_l(sim(f_i^k, f_j^l))$$

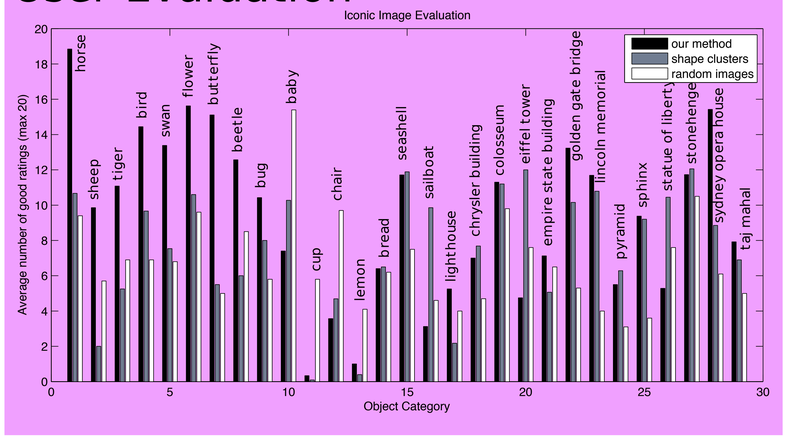
Remove Junk Images

Find 20 nearest neighbors in set of images containing in class and out of class items. Remove images with more neighbors out of class than in.

Cluster

k-medoids clustering on remainder of images.

User Evaluation



Clusters

Appearance should be relatively coherent.

Multiple senses are displayed.