Description factors

Which factors affect the things that people will describe in an image

e.g.,
- Nameable scene
- Unusualness

Nameable scenes: MTurk
Probing people for description factors

- Mechanical Turk (MTurk)
  - ○ scenes vs. non-scenes

non-
scene:

scene:

  - ○ Will people describe an office or will they describe a chair, a desk and a computer?

Nameable scenes: SUN09 database

- segmented and labeled images
- associated scenes

Examples:

non-scene   *scene*

Scene:
Labels: gorilla, wall, plant

youth hostel
wall, window, bunk bed, blanket

Nameable scenes: MTurk

Experiment and results

- 10 images from SUN09 database
  - half "typical scenes", half not a coherent scene
- Asked 9 MTurk workers to describe the images

Results:
- p(mention scene | scene) = 0.8611
- p(enumerate objects | scene) = 0.5
- p(mention background | non-scene) = 0.2407

Scene-ness: Labels vs. scenes

- Certain labels
  common across scenes
- Certain scenes have
  many labels

- Some labels
  are particularly
frequent, possibly because of
how labels were assigned

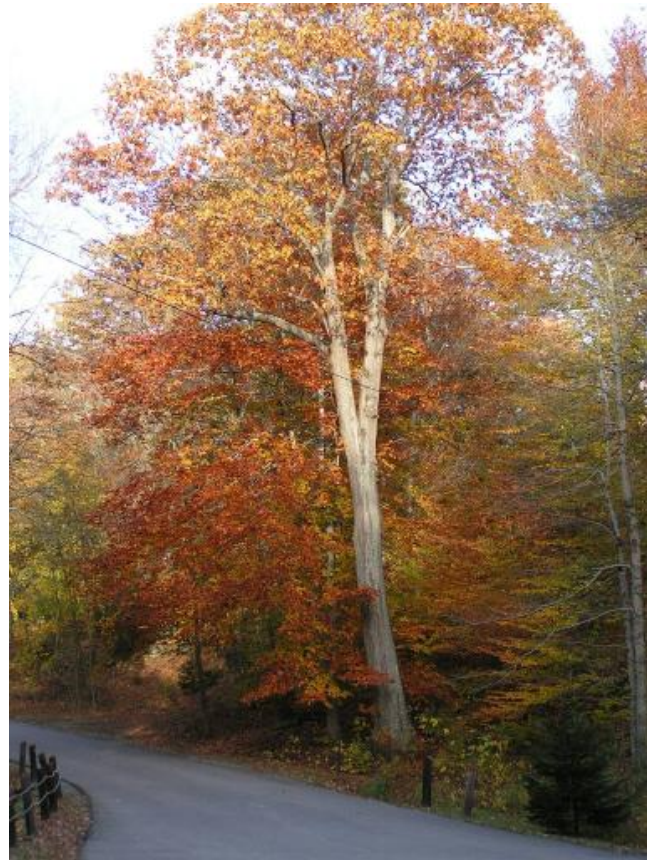| wall	10276
window	8929
sky	5309
building	5044
trees	4520
floor	3962
tree	3558
ceiling lamp	3408

scenes
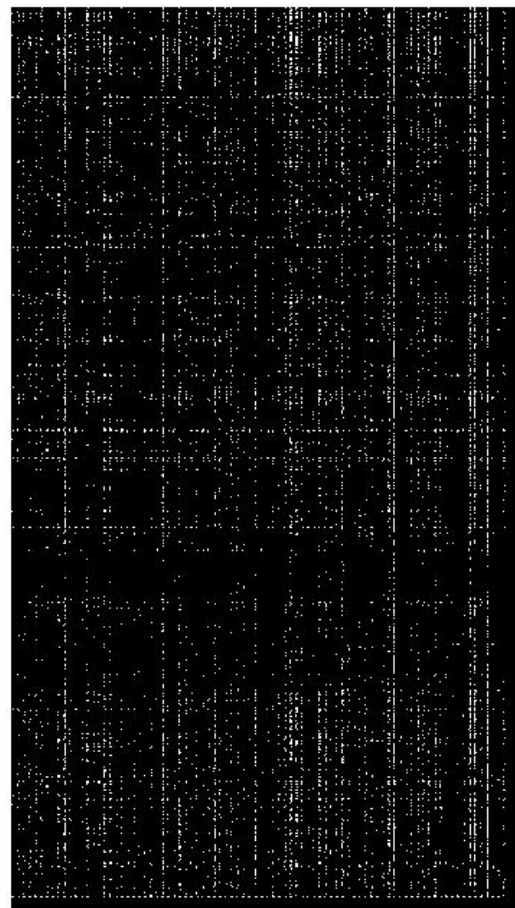
labels

Unusualness: Trees

- What are the most likely colors for trees on flickr?
- Hypothesis:
  - Flickr descriptions more likely to have unusual colors

- 413 tree images
  - (from 708k flickr images)
- MTurk:
  - up to five colors to describe trees in an image

Unusualness: Trees

- Results--

flickr	MTurk	probability
none	yellow	0.049
none	grey	0.055
none	black	0.097
none	brown	0.248
none	green	0.366
green	green	0.167
green	white	0.167
green	brown	0.333
green	red	0.333
white	yellow	1.0
red	black	0.1
red	grey	0.1
red	yellow	0.1
red	brown	0.3
red	green	0.4

Unusualness: Shirts and hats

- How likely are people to be wearing hats or shirts in images?
- ~3100 images with word "man" in caption (from flickr)

Ask MTurkers:

- *Hat/no hat?*
- *Shirt/no shirt?*
- *Is/is not a picture of a man?*

Unusualness: Shirts and hats

● Results--

Shirts vs. hats:
    p(shirt | man) = 0.861
    p(hat | man) = 0.404

In flickr descriptions:
    p(shirt mentioned | man) = 0.075
    p(hat mentioned | man) = 0.083

Approximation:
    p(shirt mentioned | shirt, man) = 0.087
    p(hat mentioned | hat, man) = 0.205