Correlated Predictors,
words
and scenes
or

What should we mean by “scene”?

D.A. Forsyth, UIUC,
with Derek Hoiem, Ian Endres, Ali Farhadi, Varsha Hedau,
Nicolas Loeff, all of UIUC

What are scenes?

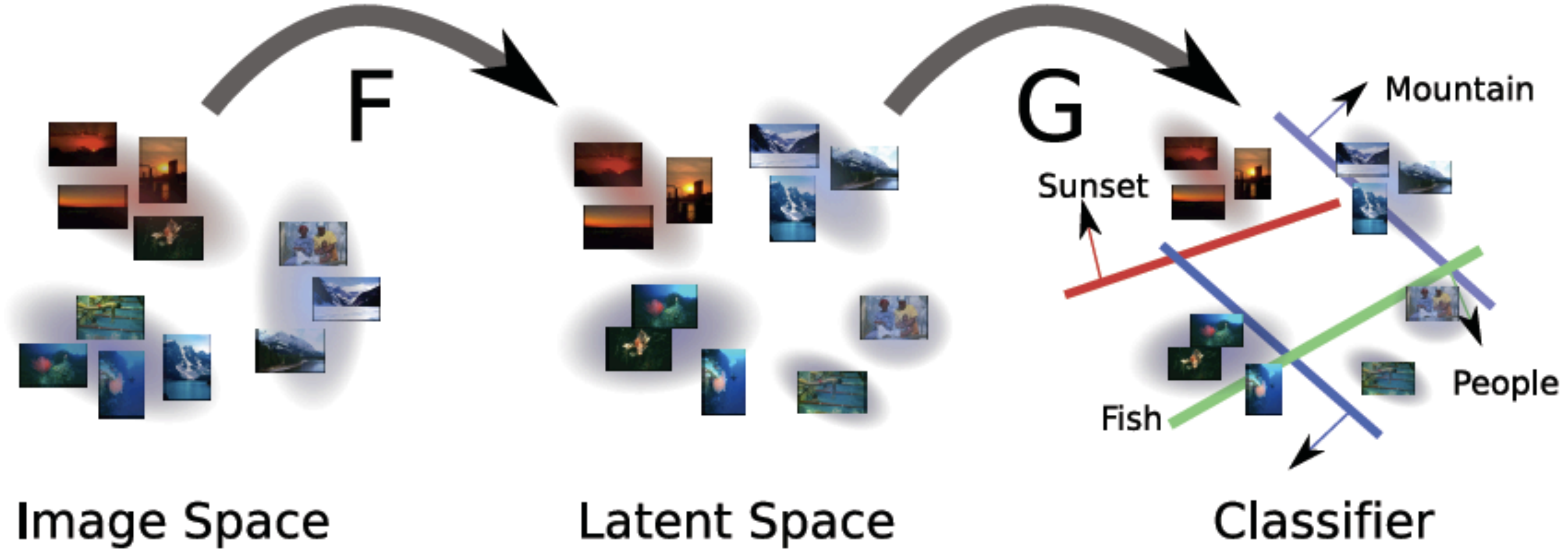
- A convenient bag in which helpful context lives
  - Objects that tend to want to appear together
  - Geometric “stages” on which objects appear
  - Illumination fields in which objects are immersed

Scenes as object bags

- We could build collections of labelled scene images
  - useful, but..
    - kitchen, bathroom, outdoor, and then?
- We could collect images of similar appearance
  - but...
    - might not really have similar objects in them
- Unsupervised bag discovery
  - Pictures of the same scene tend to contain similar objects
      - i.e. tend to attract the same image annotations

Word prediction

- Simple method:
  - rack up some features, build a bunch of linear classifiers one per word
  - works poorly
    - few examples per word
    - many features, only some are stable
- Idea
  - some features are not helpful
  - a low dimensional subspace is good at predicting most things (Ando +Zhang,)
  - We can find this space by penalizing rank in the matrix of linear classifiers

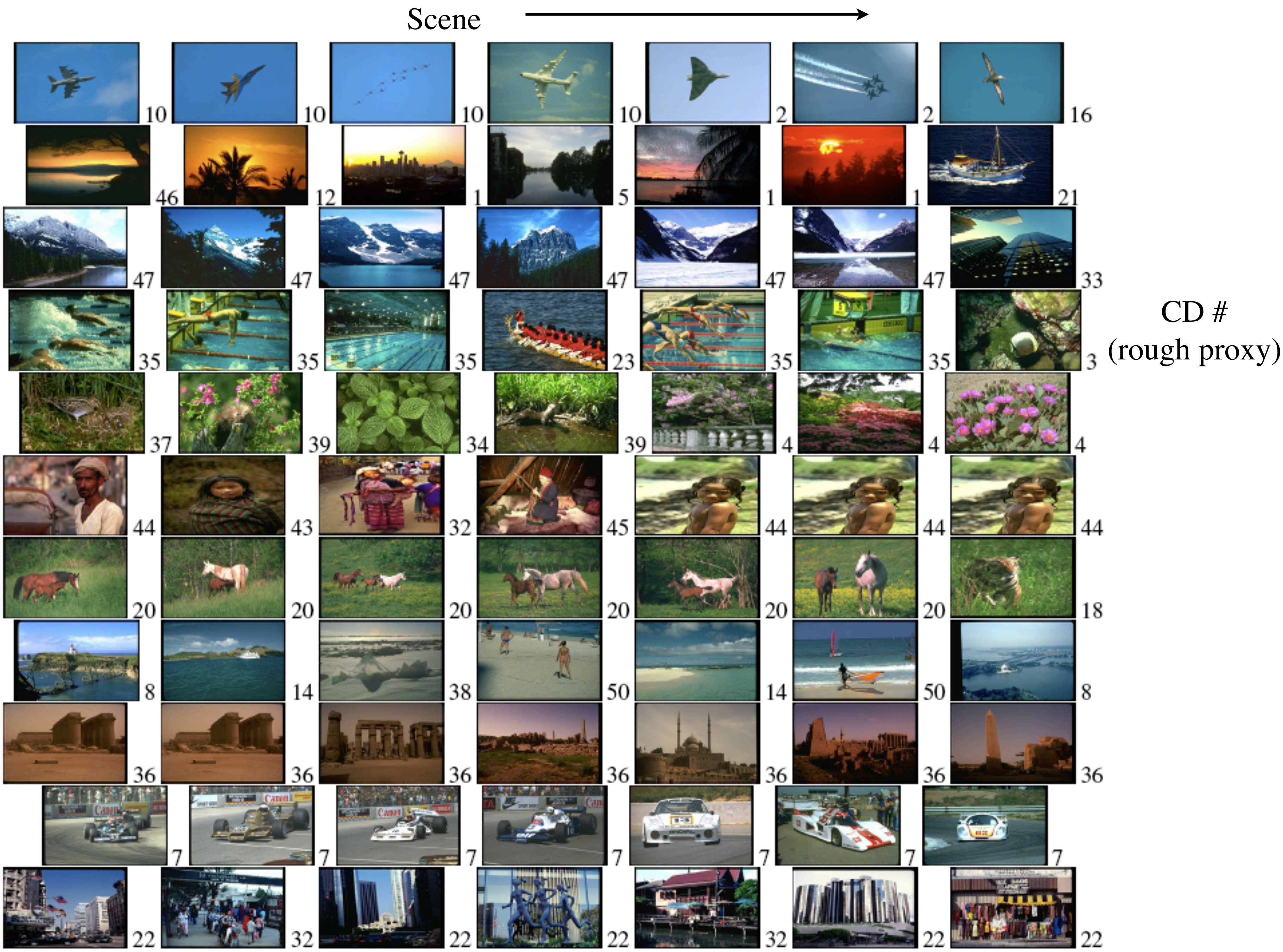


Cluster points in the latent space

Close together if they predict similar words

This means we get clusters of things that look similar in ways that support word prediction

Loeff Farhadi 08



Loeff Farhadi 08

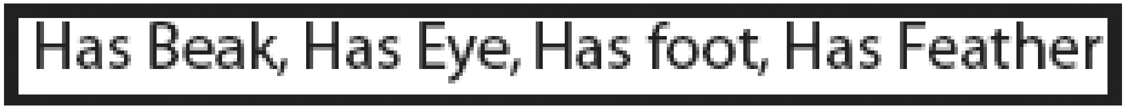
How should we describe scenes?

- We probably don’t know the names of each type
    - Can’t just
        - build a kitchen detector, bedroom detector, etc.
    - Like objects

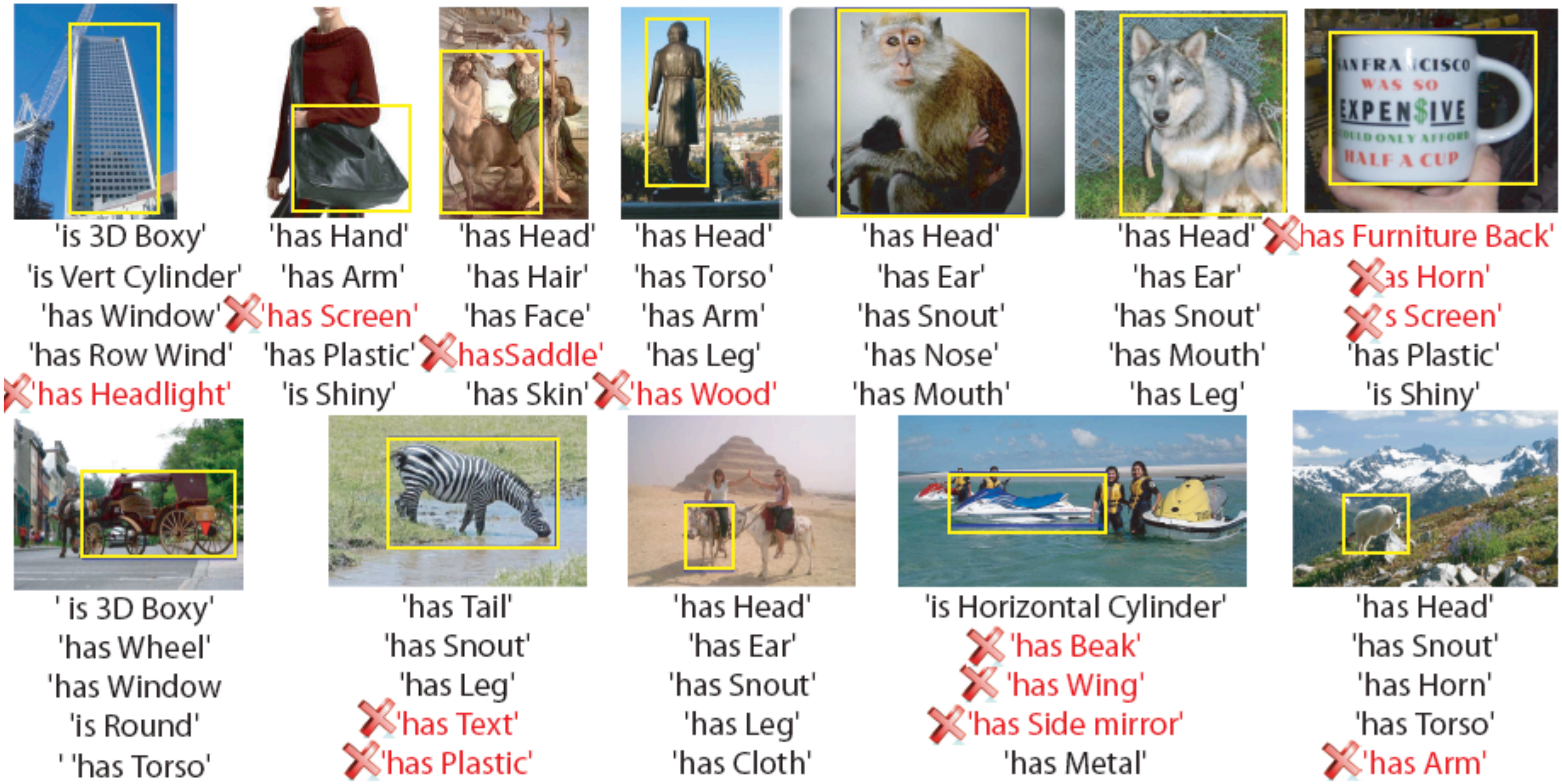
- Want to:
    - make useful statements about scenes whose names aren’t known
    - say how a scene of known type is different from others of that type

Attributes

- Desiderata suggest a fluid notion of scene category
  - in terms of useful properties rather than names
  - new scene types are new collections of properties
  - not all required to match (exemplar theory originated here)



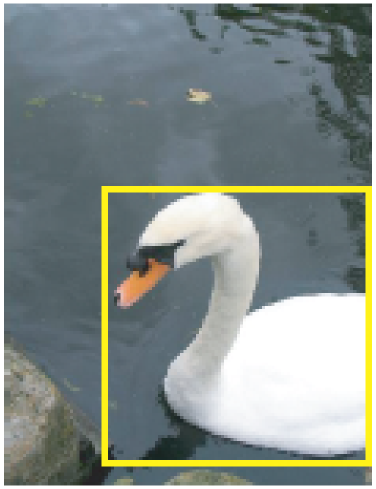
Has Beak, Has Eye, Has foot, Has Feather



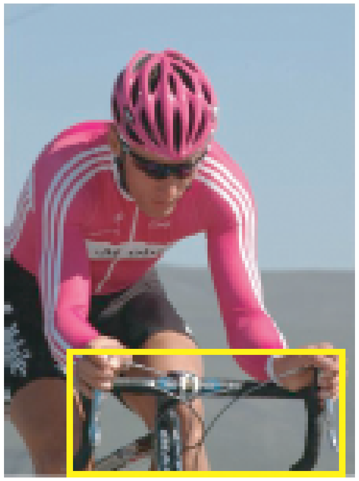
Farhadi et al, in review

Missing attributes

Boat
No "sail"



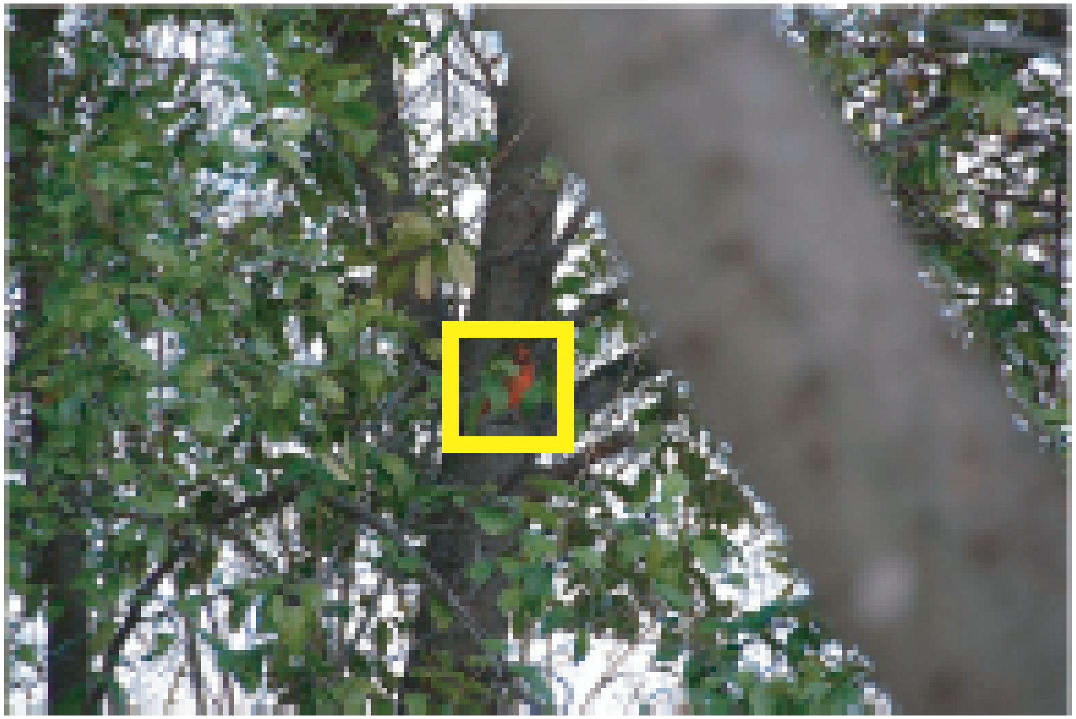
Bird
No "tail"



Bicycle
No "wheel"

Farhadi et al, in review

Extra attributes



But what should the properties be?

- Things that are good at helping predict where objects are
- And what they are like

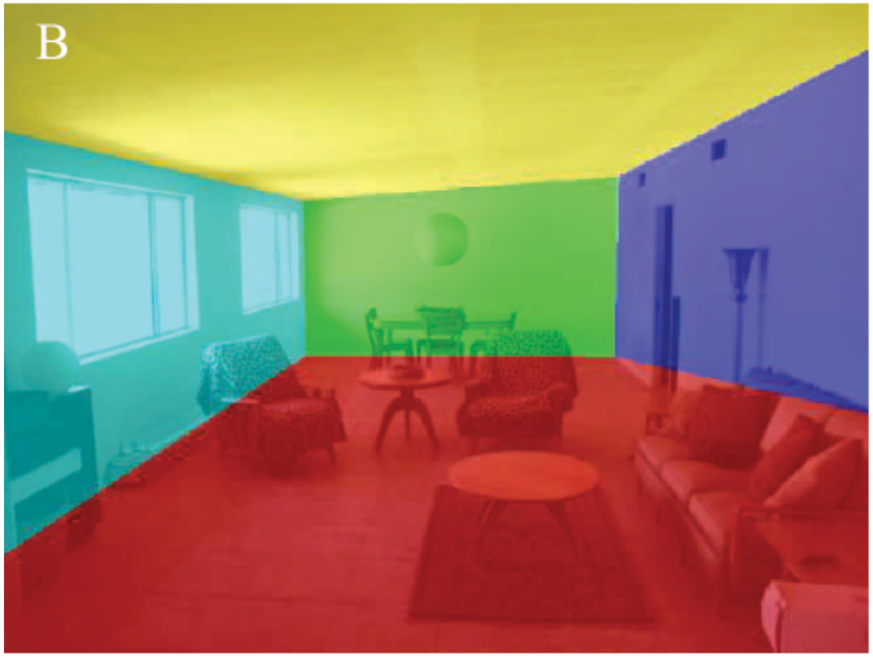
Scenes as stages

Hedau et al, in review

Estimating the box - I

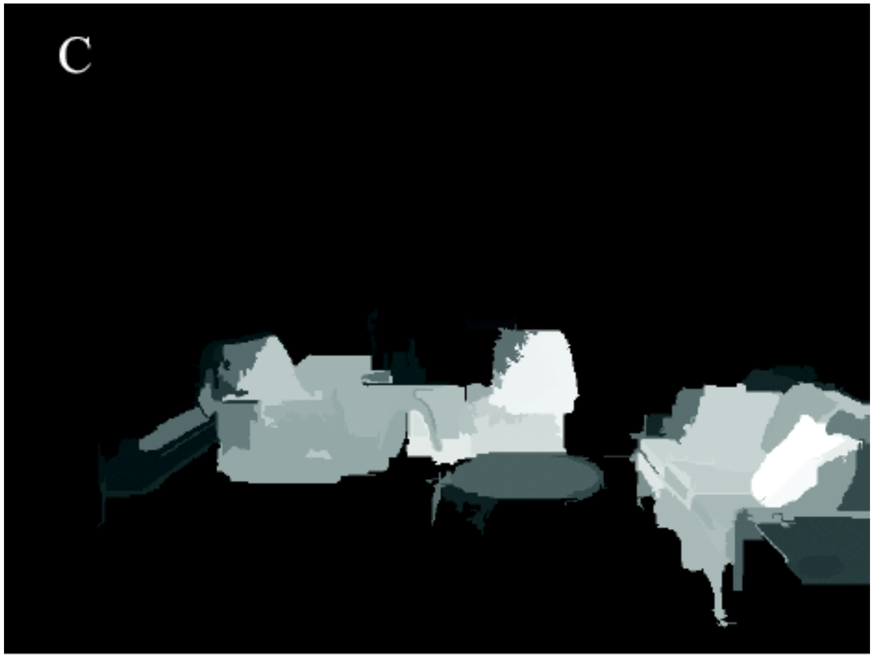
- Use vanishing points to estimate rotation
  - cf Coughlan Yuille 99, 03; Rother, 02; Yu et al ’08; Deutscher et al, ’02
- Use lines through vanishing points to get translation
  - box corner
  - using search, cost criterion learned using structure learning
  - yields first estimate of face appearance

The “stage”



Hedau et al, in review

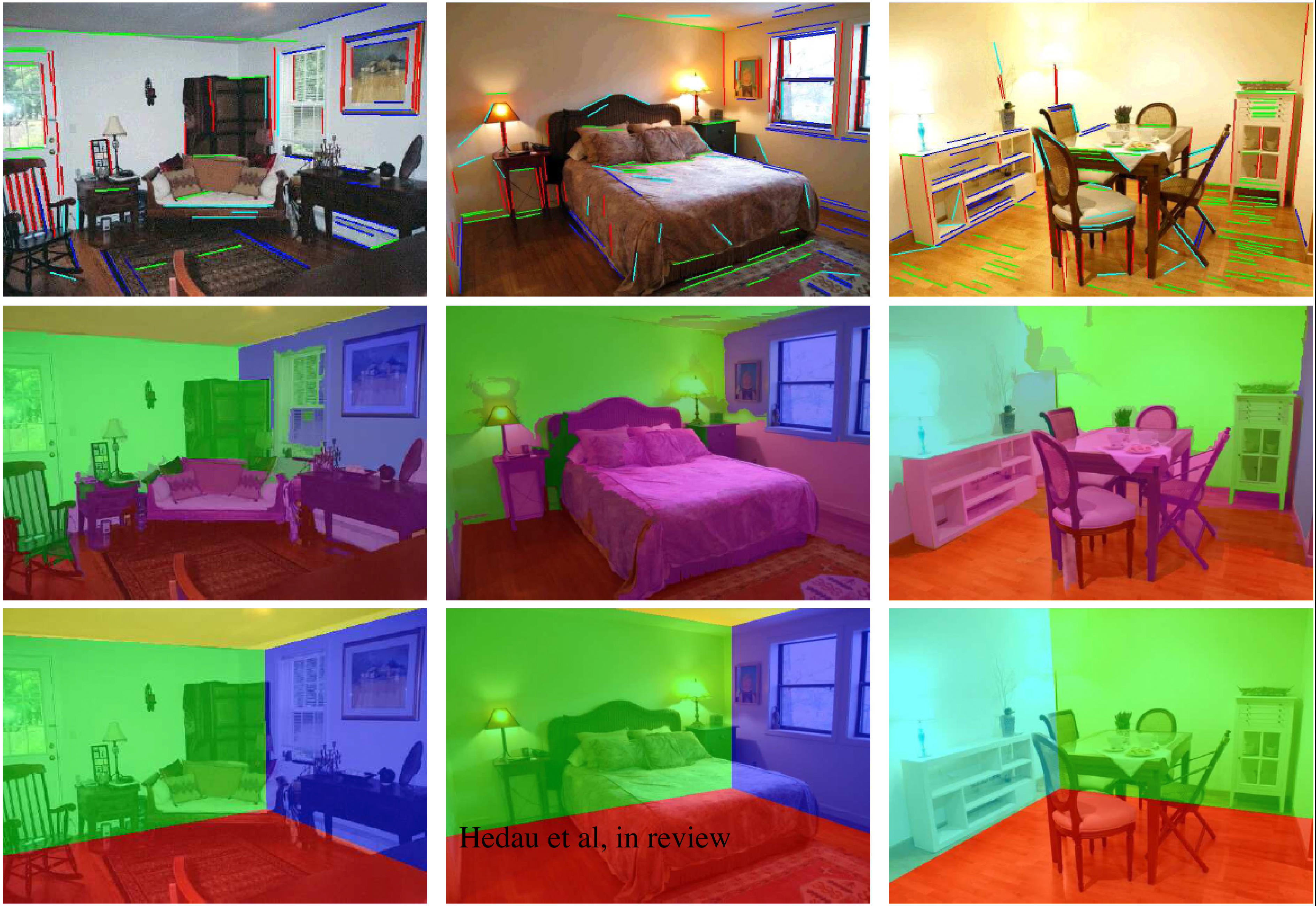
The "players"



Hedau et al, in review

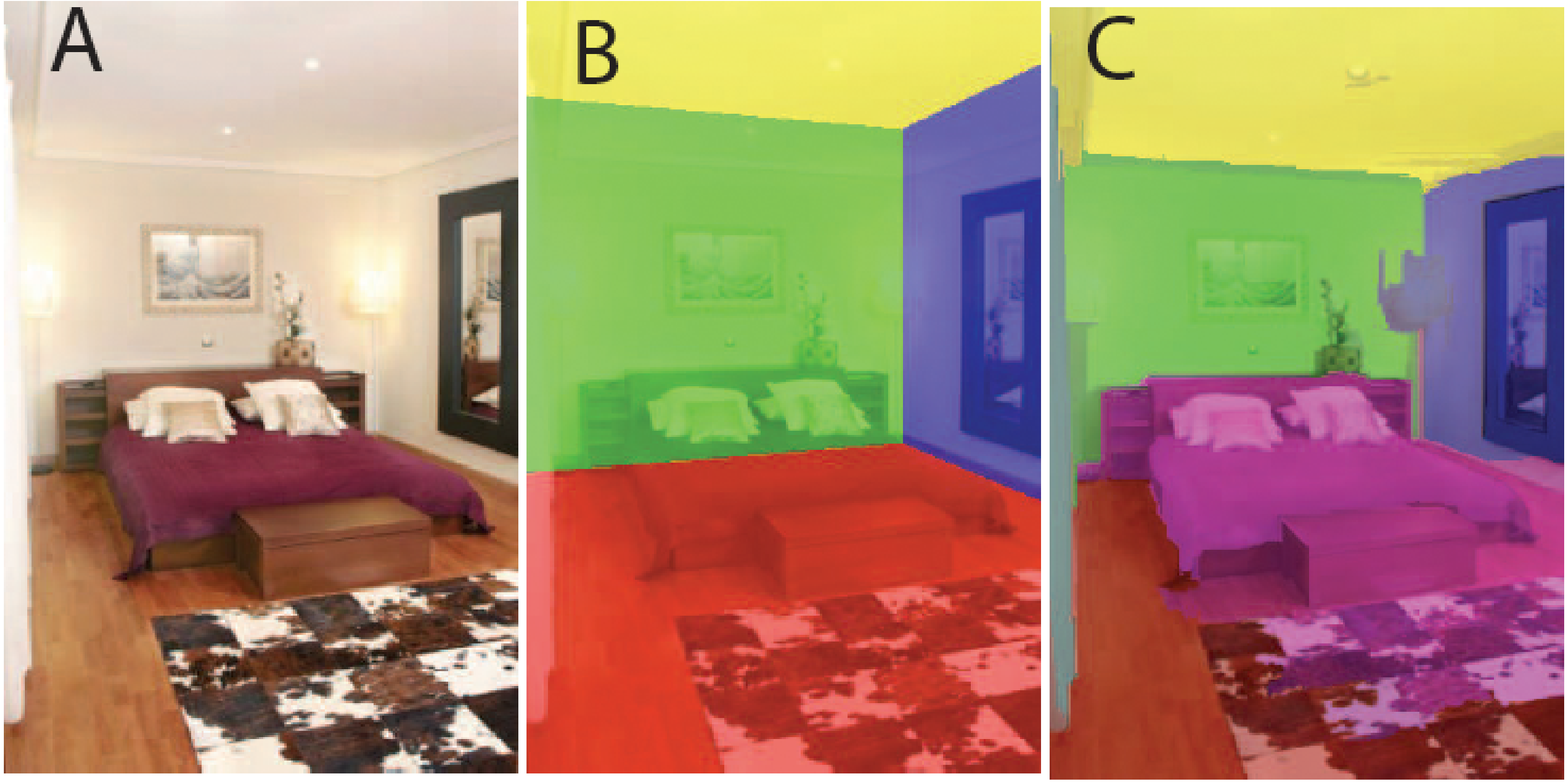
Process - II

- Refine translation estimate by
  - using “non-face” info to rule out some lines, features
  - reestimate with search, criterion learned via structure learning



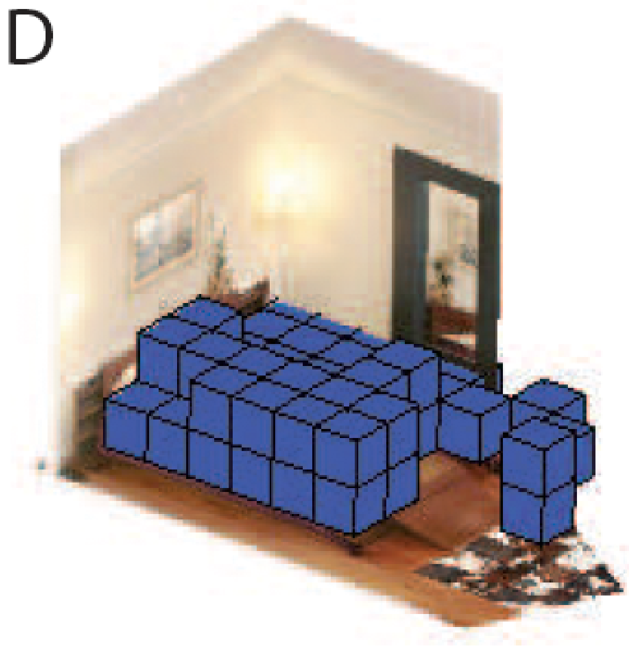
Hedau et al, in review

Stage and Players



Now, what can we do...

- Free space estimate
  - using standard SFM construction for camera given manhattan world
  - couple appearance model of objects to approximate geometric models?



Hedau et al, in review

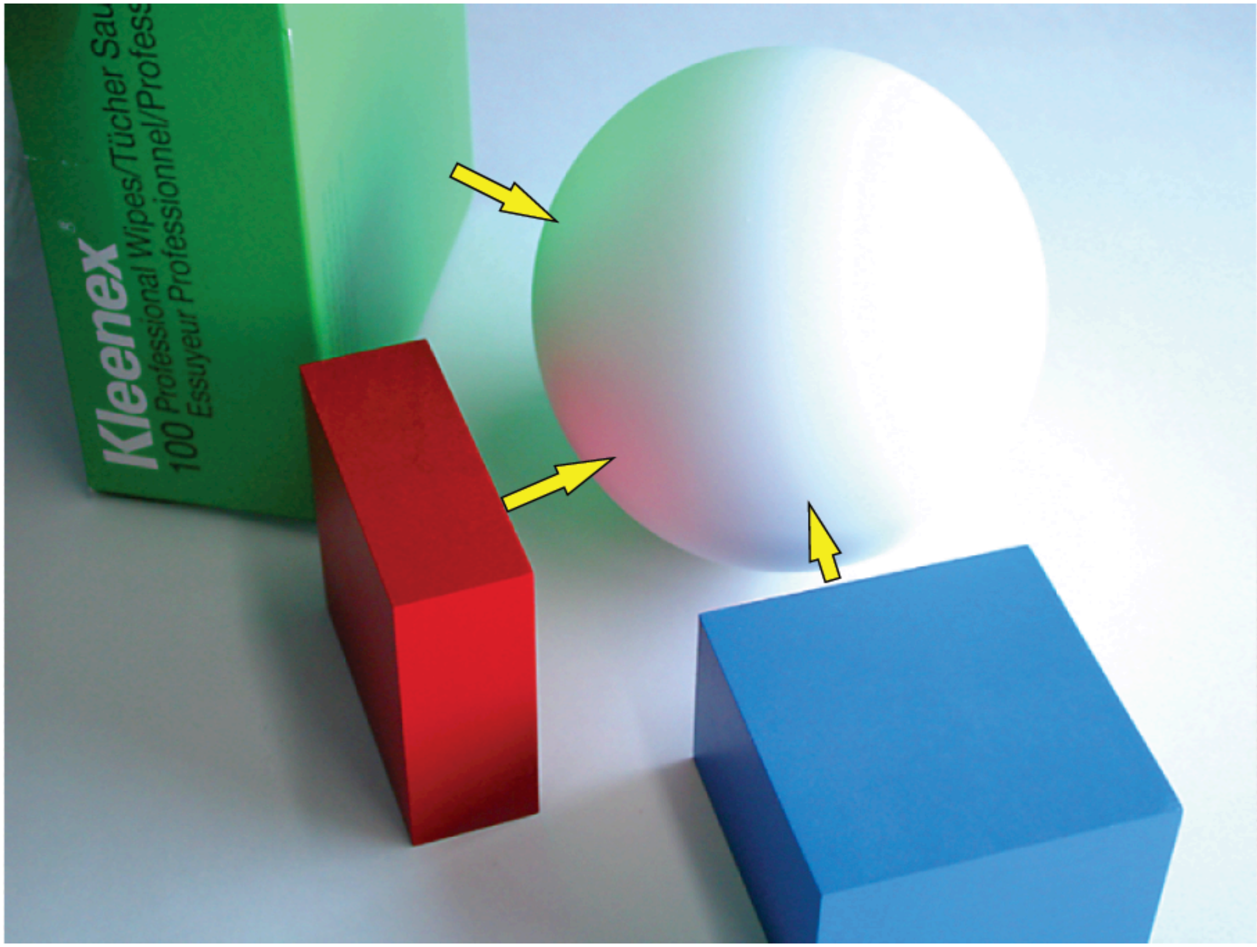
Stage lighting

From Koenderink slides on image texture and the flow of light

Stage lighting could tell us...

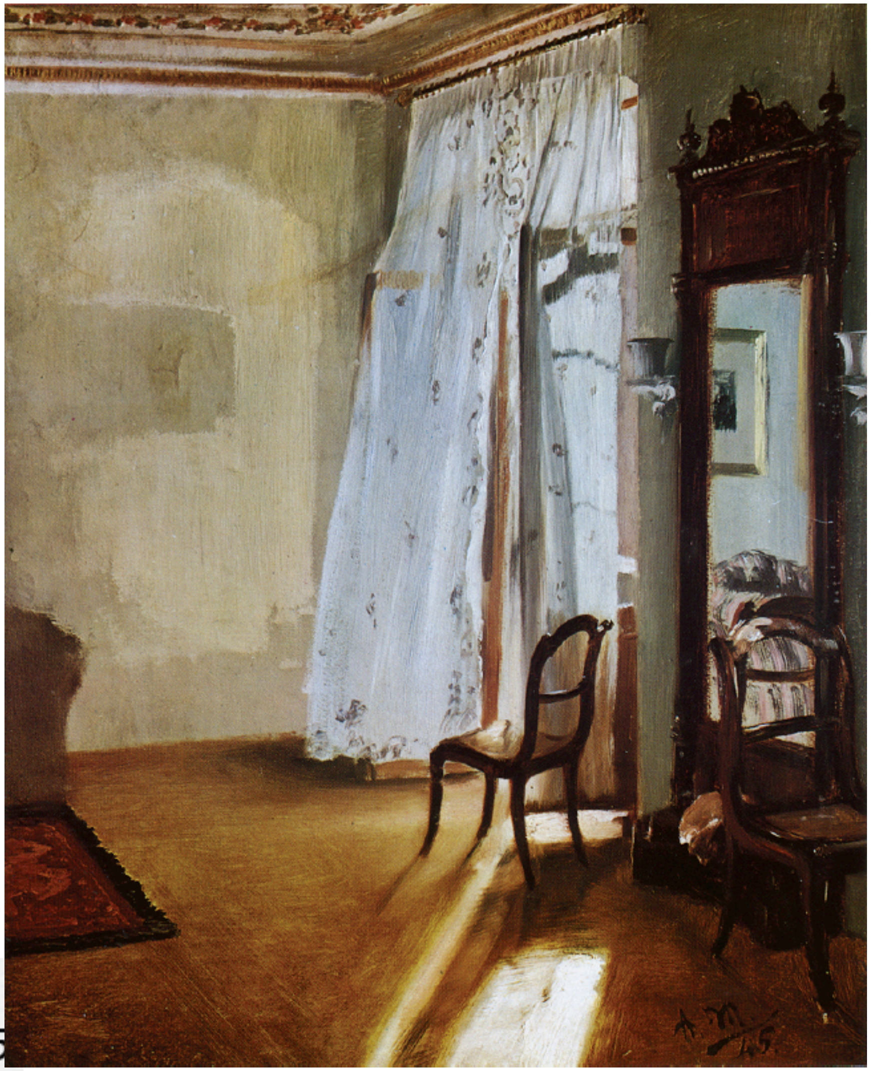
- Material properties
    - roughness
    - shininess
    - lightness
    - surface color
- Shape (!?!)

Stage lighting is hard


Kleenex
100 Professional Wipes/Tücher Sau
Essuyeur Professionnel/Profess

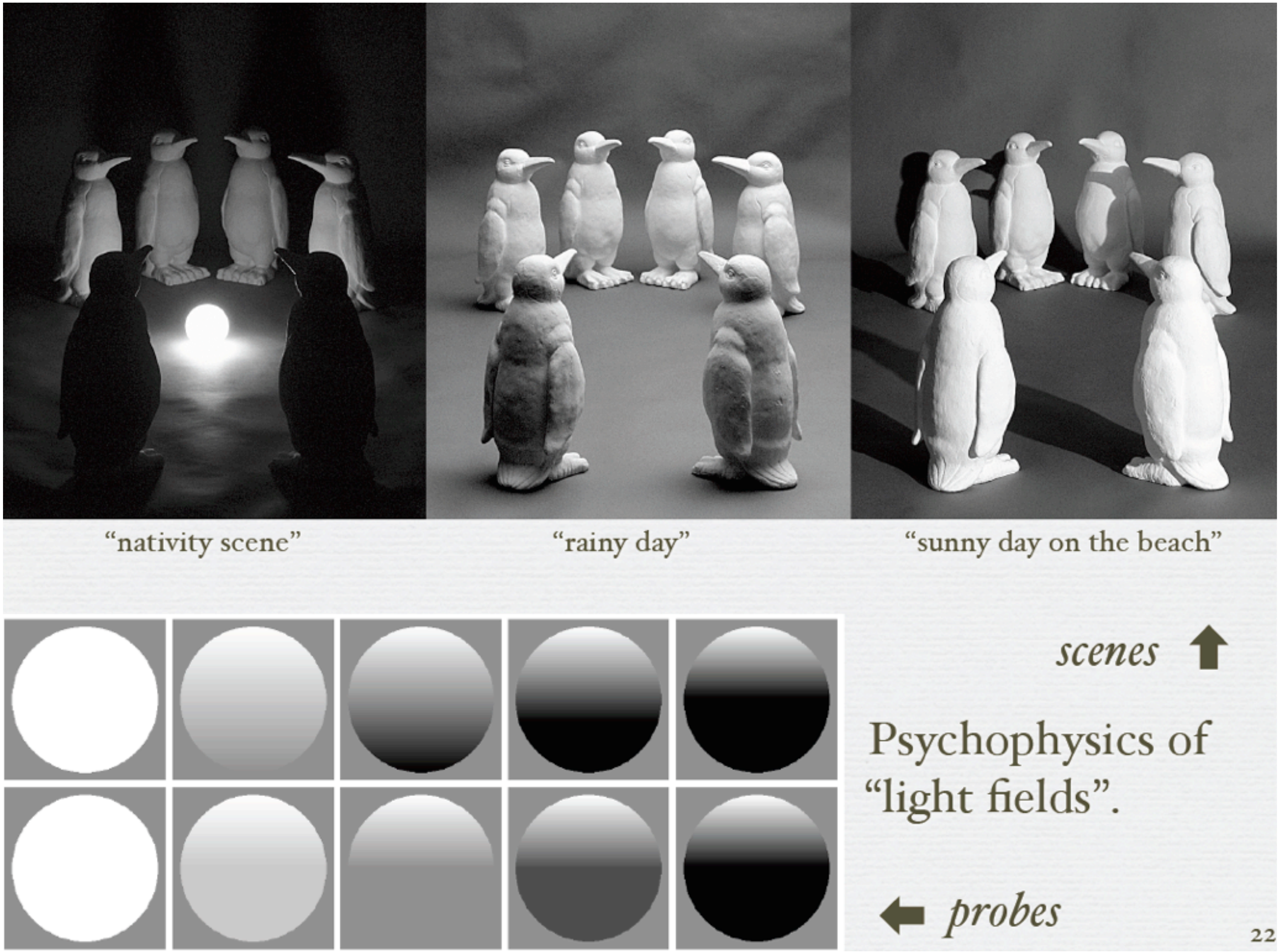
From Koenderink slides on image texture and the flow of light

Stage lighting is hard



Adolph von Menzel:
Das Balkonzimmer, 1845

From Koenderink slides on image texture and the flow of light



"nativity scene"

"rainy day"

"sunny day on the beach"

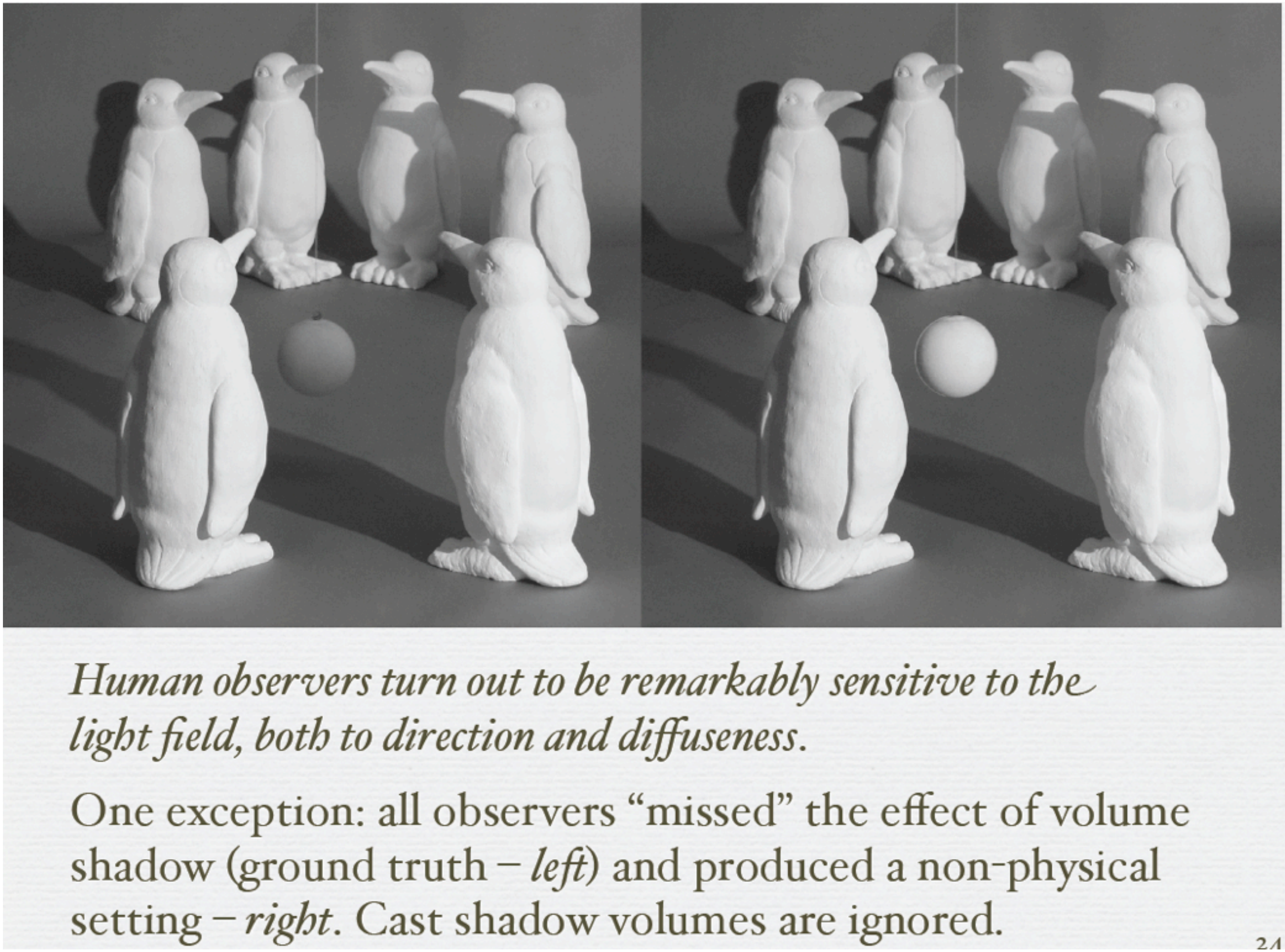
scenes ⬆

Psychophysics of "light fields".

⬅ *probes*

From Koenderink slides on image texture and the flow of light

Koenderink et al 07



Human observers turn out to be remarkably sensitive to the light field, both to direction and diffuseness.

One exception: all observers “missed” the effect of volume shadow (ground truth – *left*) and produced a non-physical setting – *right*. Cast shadow volumes are ignored.

From Koenderink slides on image texture and the flow of light

Koenderink et al 07

Summary points

- Scene categories should be fluid
  - which forces us to think about attributes rather than names
- Some attributes might be
  - what the picture looks like
  - what could be there
  - what the overall geometric stage is
  - how the stage is lit