Stages as Models of Scene Geometry

Vladimir Nedović[1], Arnold Smeulders[1], Jan-Mark Geusebroek[1], and André Redert[2]

[1] *Intelligent Systems Lab Amsterdam (ISLA), University of Amsterdam, Kruislaan 403, 1098 SJ Amsterdam, The Netherlands*

[2] *Philips Research Laboratories Eindhoven, High Tech Campus 36, 5656 AE Eindhoven, The Netherlands*

Problem Statement

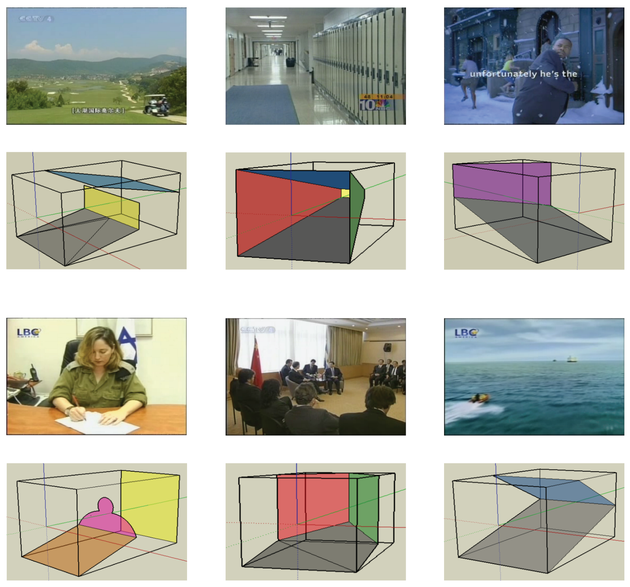
- **GOAL:**
  - exploit the inherent constraints of the 3D world to reduce the problem of scene geometry estimation from single images
- **APPROACH:**
  - separate the scene into *a stage* and the objects
  - observe the structure in real-world images to define few stage categories
  - to derive scene geometry of an image, classify it first into one of the stages
  - the stage label provides rough background depth profile, used as a prior for further estimation

The Stages

We observe the structure present in real-world images in order to arrive at a limited number of geometric scene types [Nedovic et al. ICCV 2007].

The structure in visual space is imposed by three crucial constraints:

- natural image statistics results in statistical regularities
- 3D viewpoint constraints limit the perspective possibilities [Hoiem et al. ICCV 2005]
- 'modal' configurations [Richards et al. in *Perception as Bayesian Inference*, 1996.] ensure for the orthogonality of relevant lines and angles

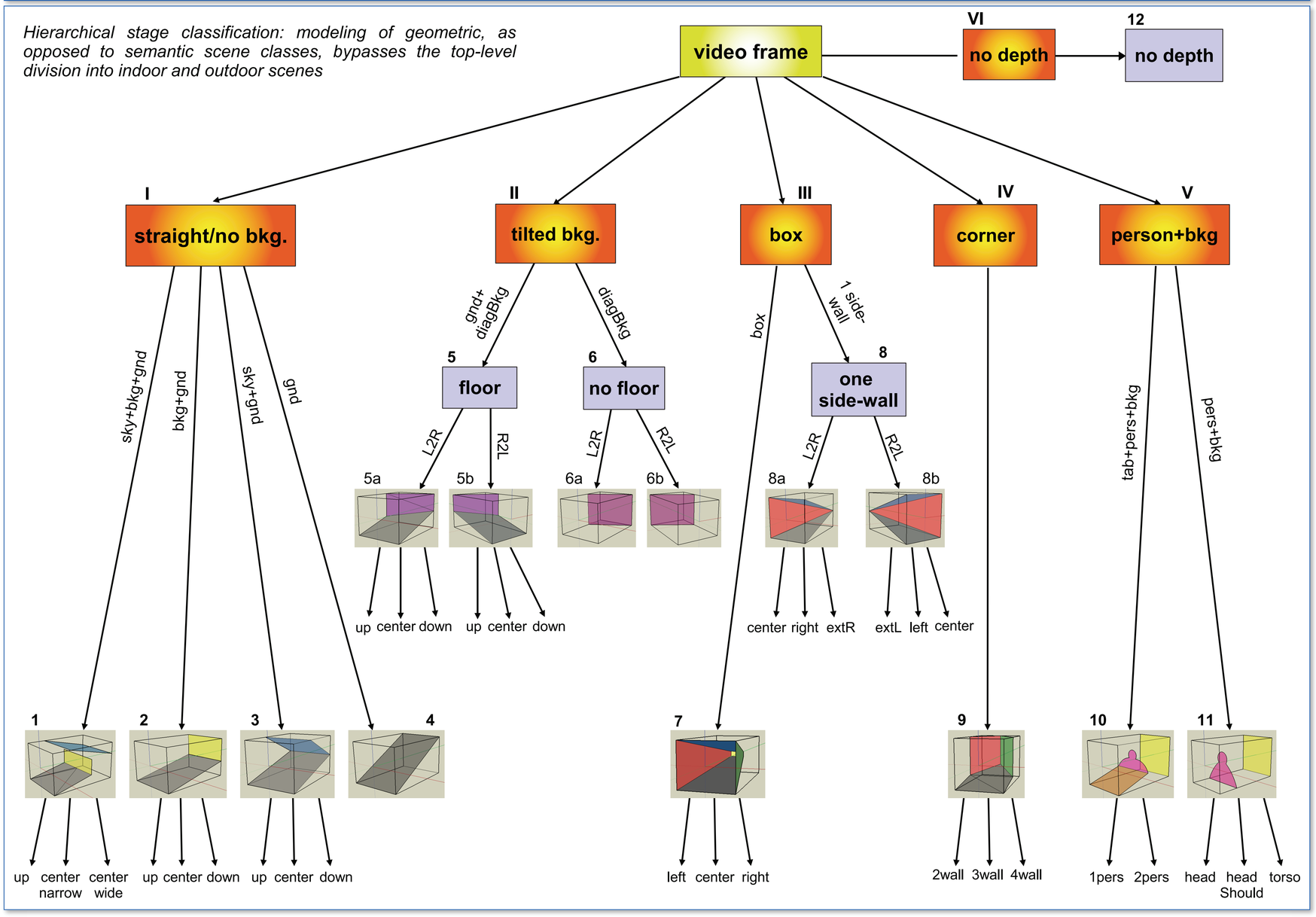


Example frames and their stage models;
top two rows, from left to right:
sky+background+ground, box, ground+diag. Background (RL);

bottom two rows:
table+person+background, corner, sky+ground.

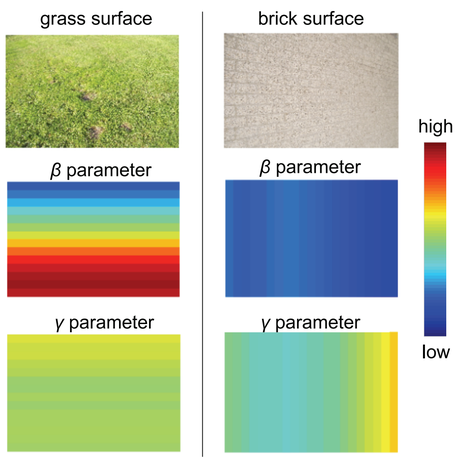
Stage Classification

Hierarchical stage classification: modeling of geometric, as opposed to semantic scene classes, bypasses the top-level division into indoor and outdoor scenes



Natural Image Statistics

- There exists a direct relation between image statistics, scene structure and depth pattern [Torralba and Oliva, PAMI 2002]
- With a single visual surface observed, gradient histogram typically follows a decaying power-law distribution
- With increased depth and multiple structures present, integration over various power-laws results in a Weibull distribution [Geusebroek and Smeulders, IJCV 2005]
- Spatial image statistics will conform to Weibull *pdf* until depth increases to the point at which the observed samples become completely uncorrelated, resulting in a Gaussian histogram
- Natural image statistics are captured by parameterized edge histograms



Weibull parameters as a function of depth for textures of grass and bricks:
β decreases from the point of fixation, whereas γ increases with depth.

Weibull Distribution

We use a Gaussian scale-space framework to extract texture features. Histograms of gradient magnitude are modeled by an integrated Weibull distribution,

$$f(x) = \frac{\gamma}{2\gamma^{\frac{1}{\gamma}}\beta\Gamma\left(\frac{1}{\gamma}\right)} e^{-\frac{1}{\gamma}\left|\frac{(x-\mu)}{\beta}\right|^{\gamma}}$$

The parameters μ, β, and γ represent the center, width and shape (i.e. peakness) of the distribution, respectively, Changes in scene depth are directly transposed into the parameters of the distribution.

Preliminary Classification

For evaluation [Nedovic et al. ICCV 2007], we have used the key-frames of the 2006 TRECVID video benchmark [Smeaton et al. ACM MIR, 2006].

stages

class	name	% in dataset	% correct
1	sky+bkg+gnd	6.3%	16.7%
2	gnd+bkg	7.1%	8.2%
3	sky+gnd	8.7%	60.7%
4	gnd+bkg	7.4%	44.7%
5	gnd+diagBkg	10.8%	26.9%
6	diagBkg	6.4%	14.3%
7	box	5.5%	8.1%
8	1 side-wall	9.0%	13.6%
9	corner	10.8%	34.3%
10	tab+pers+bkg	7.4%	48.0%
11	pers+bkg	13.1%	42.5%
12	no depth	7.4%	22.4%
			AVG: 28.4%

stage groups

group	name	% in dataset	% correct
I	straight/no bkg.	29.5%	69.5%
II	tilted bkg.	17.2%	35.2%
III	box	14.5%	19.6%
IV	corner	10.8%	13.2%
V	person+bkg	20.5%	63.1%
			AVG: 40.1%

Correct classification is given by the total number of correctly classified (true positives + true negatives) divided by the total number of images.

Conclusions & Future Work

- We describe how the problem of scene geometry estimation from single images can be approached by first performing scene classification
- Inherent structure of the visual world, resulting from natural image statistics, viewpoint constraints and modal configurations, leads to only 15 typical 3D scene geometries – stages - each with a unique depth pattern
- Three-level classification proposed - geometry at the bottom is constrained sufficiently, such that pre-defined crude depth models are already possible
- Preliminary scene classification performed with a single feature type, on a challenging video dataset; the results indicate that stages without much variation or object clutter can be detected with up to 60% success rate
- Stage information is a prior that reduces the search space:
  - needs to be determined robustly => more features necessary (e.g. perspective lines, horizon and vanishing point location, etc.)
  - can be used for more precise depth estimation, object localization, etc.

vnedovic@science.uva.nl