ELL 788
Computational Perception & Cognition

July – November 2015

Module 4

Perceiving visual space

What is space

Space is the theatre where events take place

- Visual space = Visual environment around a person

- Objects: What we act upon
- Environment: Wherein we act / navigate
  - *Described by floor, ceiling, walls, openings (doors / windows), partitions, etc.*

Perception of space

- Observer-centric
  - *Viewpoint: The space as seen by the observer*
- Two levels of description
  - Structural (geometric)
  - Semantic (what it represents)
    - *e.g. street scene, natural landscape, etc.*
- Perceived in
  - In real 3D world
  - From 2D images

Isovists

What is an 'Isovist' ?

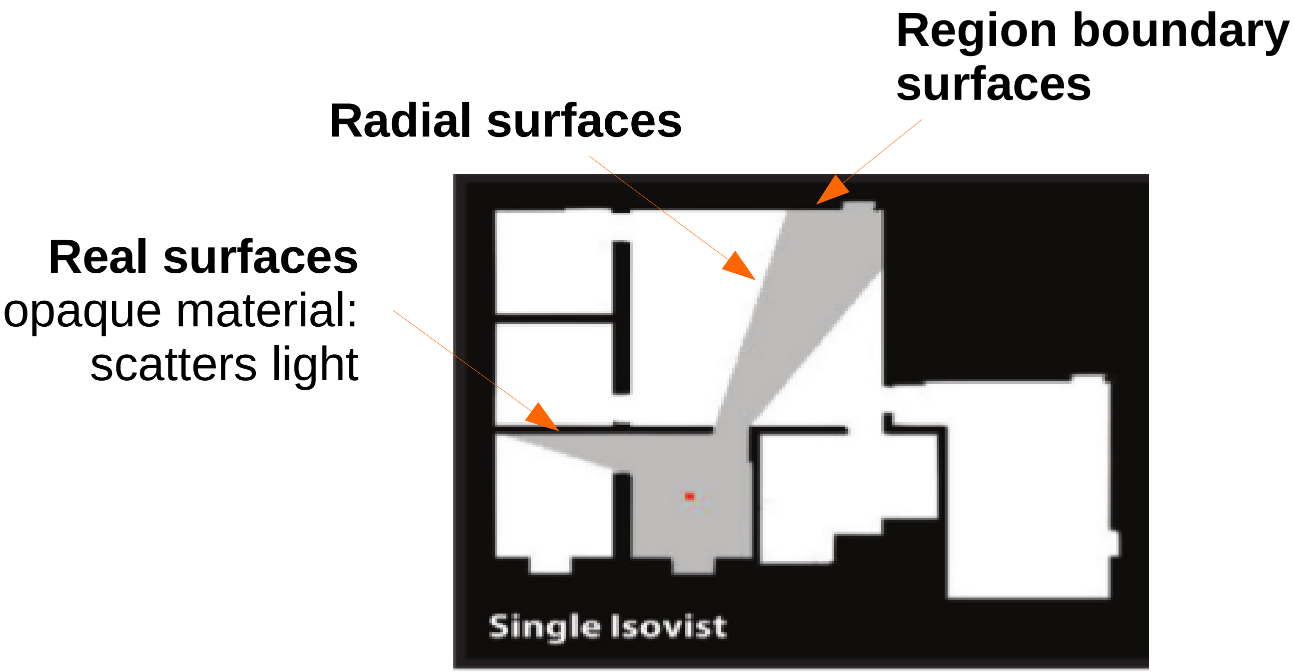
- **View / Viewpoint**: a cone of visible space as seen from an observer’s vantage point
  - Orientation
  - Aperture
- **Isovist**: The set of surfaces visible from that location if the observer rotates through 360 degrees

Source: Oliva, et al. 2010

Isovist: Formal definition

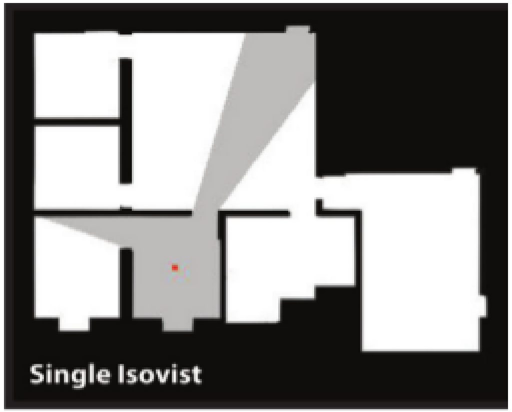
Two alternate definition of isovist: V_x (x: Vantage point)

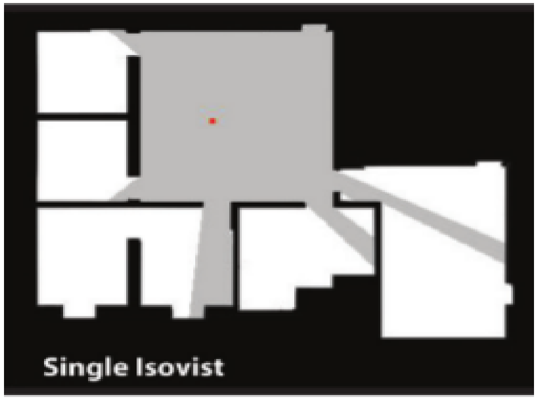
- Set of all points in the environment that are visible from x (3D)
- Set of surfaces that are visible from x (2D)



Isovist map

- A collection (superimposition) of all isovists visible from all possible locations in a space





Source: Oliva, et al. 2010

Sufficient set and Cardinality number

How many vantage points are required to see an entire environment E?

- A finite (usually small) number of vantage points are sufficient
  - The set of isovists is called a **sufficient set**
- The cardinality of smallest sufficient set(s) is called the **cardinality number** for an environment

<u>Application</u>

- *How many security cameras do you need to cover a retail outlet?*
- *Where do you deploy them ?*

Minimal path

- Paths connecting vantage points in a sufficient set is called a **sufficient path**
- The sufficient path that is shortest is called a **minimal path**

Application

What should be the path taken by a sentry in a museum / residential layout ?

Concealment

- Let D1 and D2 be two sub-regions in D, such that $D1 \cap D2 = \varnothing$
- Let I(D1) be the isovist map from D1

- If $I(D1) \cap D2 = \varnothing$,
  - D2 is totally invisible (concealed) from D1
- If $I(D1) \cap D2 \neq \varnothing$, and $I(D1) \cap D2 \neq D2$
  - D2 is partially visible (concealed) from D1
- Assymetrical relation:
  - In general, $\text{Volume}(I(D1) \cap D2) \neq \text{Volume}(I(D2) \cap D1)$

__Privacy of a space__

- *How much you can see without showing your own space*

Isolation

- Let D1 and D2 be two sub-regions in D, such that D1 ∩ D2 = ∅
- Let I(D1) be the isovist map from D1
- Let I(D2) be the isovist map from D2

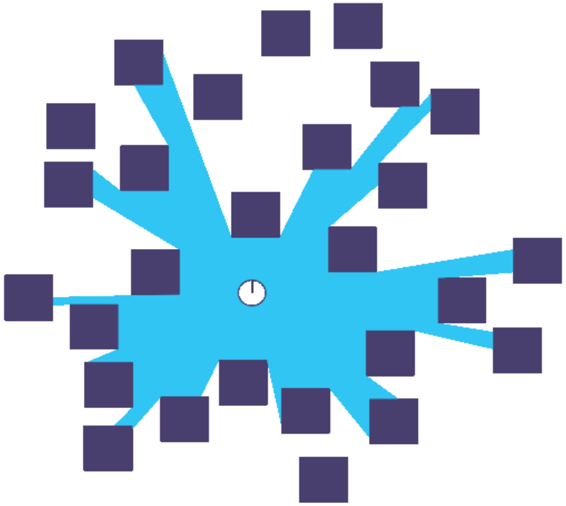
- If I(D1) ∩ I(D2) = ∅,
  - D1 and D2 are fully isolated
- If I(D1) ∩ I(D2) ≠ ∅, and I(D1) ∩ I(D2) ≠ D1 ∪ D2
  - D1 and D2 are partially isolated

Architectural design

- *Should the living room be isolated from the kitchen or the bedrooms?*

Shape of space

Simple isovist



Complex isovist

Application in architecture

- *The characteristics of an isovist (volume, shape) determines our psychological perception about a closed space*

Isovist measures

- Consider a planar slice of an isovist
- Consider all radial lines from vantage point (x) to the boundary of the environment
  - $l(x,\theta)$ *: length of a radial line from x in direction* θ
- L_x (θ) is a distribution function of lengths over θ
  - One way to describe an isovist V_x quantitatively
- Statistical measures $m(V_x)$ with can be derived
  - Have perceptual significance

- Area (A)
  - How much space can be seen from x
  - From how much space, x can be seen
- Perimeter (P):
  - How much wall (real surface area) can be seen from x
- Occlusivity (Q):
  - The sum of lengths of the “radial surfaces”
  - Indicator of how much of the environment may be occluded
- Variance and Skewness of $L_x(\theta)$:
  - Dispersion and assymetry of dispersion of the perimeter from the vantage point x
- Circulartity (N):
  - Ratio between perimeter-square and 4π times area

How does these measurements relate to human perception and behavior?

- Simple descriptors (area, occusivity, perimeter, ...) relates to people's impressions of the spaciousness of hotel lobbies and urban spaces ...
  - *Narrow, cramped, spacious, ample*
- Relates to architectural designs and aesthetics of space
- Guides pedestrian behavior in narrow roads, seating preferences in restaurants ...

[See Franz, et al.]

- Isovists do not account for surface properties, such as illumination, color, texture, that are important for perception of space

Spatial envelop representation

Describing shape of scenes






- Can we characterize a scene without refering to the individual objects ?
- Can we capture the overall spatial layout and geometry ?

Spatial envelop representation

- Spatial envelop [architecture]
  - A description of the whole space that provides an “instant impression of the volume of a room or an urban site”

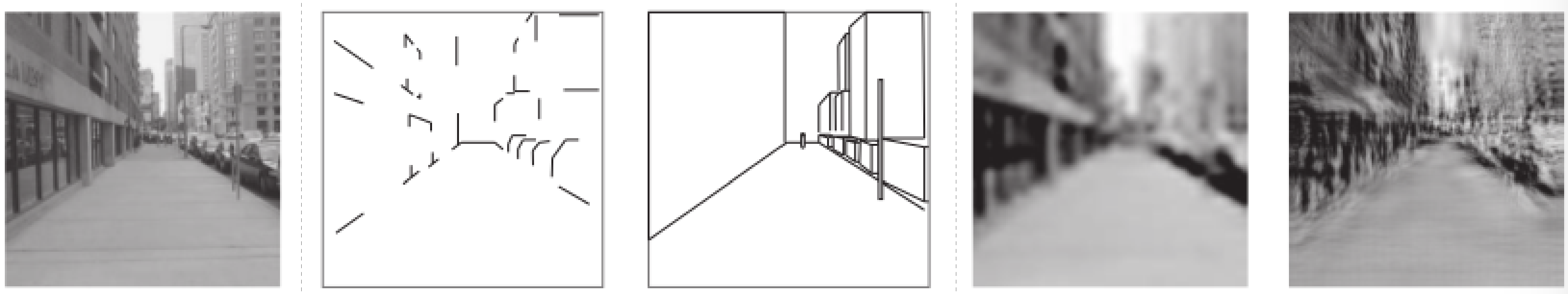
- Spatial envelop representation
  - A formal, computational approach to the capture of the shape of space, as it would be perceived from an observers vantage
  - Can be described by a collection of properties
    - perspective, size, dominant depth, openness, naturalness, ...

Levels of scene model

- **Subordinate level**
  - Scene is treated as a composition of objects, e.g.
  - Cars + People + Buildings $\rightarrow$ City Street,
  - Sand + Sea + Sky $\rightarrow$ Beach
- **Basic level**
  - Scenes having similar <u>global</u> visual properties
  - e.g., forest, mountain, street
- **Superordinate level**
  - Highest level of abstraction, little visual similarity in a class
  - e.g., an urban environment, natural landscape

Spatial envelop Representation

Properties of scene

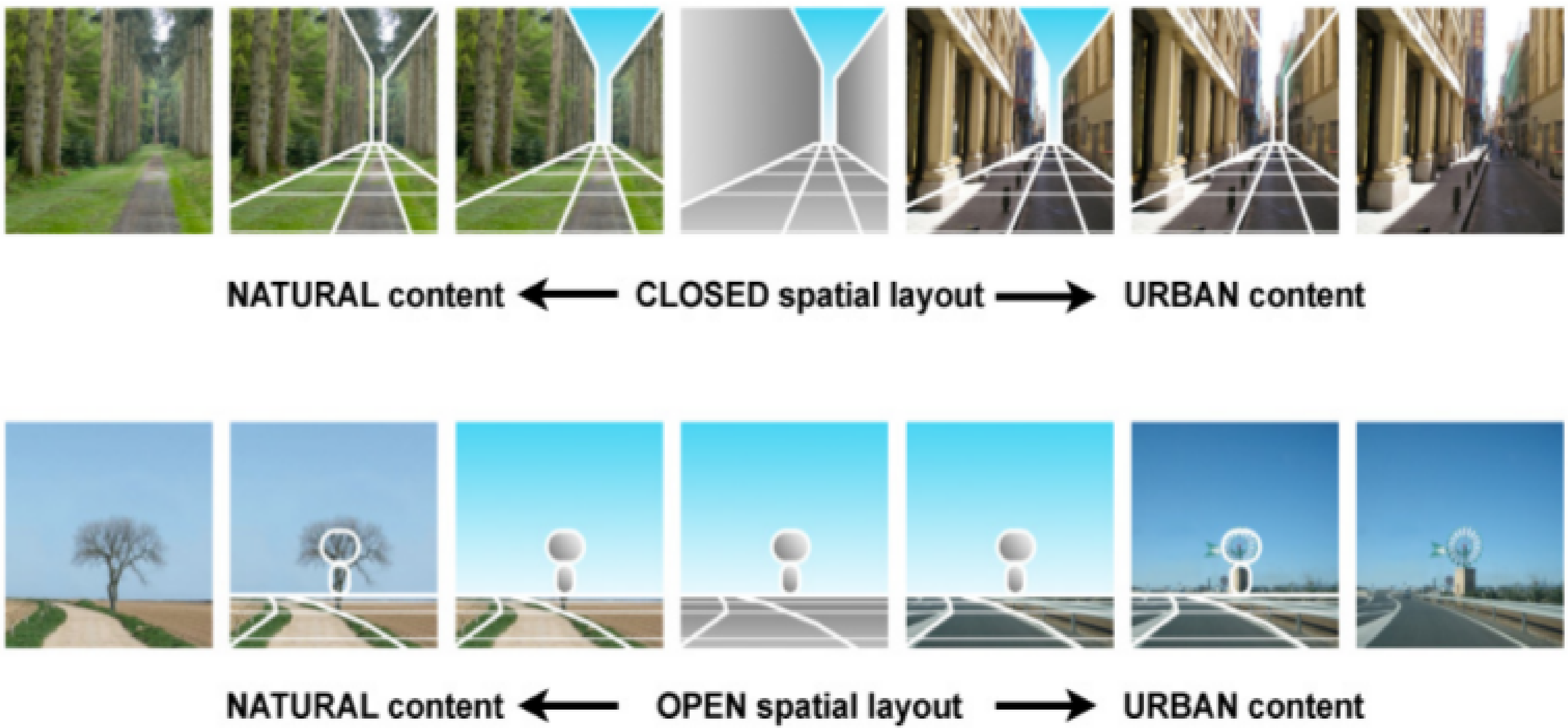


Original scene "a street" | Suggestive contours (Biederman, 1981) | Geometric forms (Biederman, 1981) | Blobs in relations (Schyns & Oliva, 1994) | Sketch of textures (Oliva & Torralba, 2001)

Boundaries
(size, shape)

Contents
(Texture, color)

Properties of scene ... more



Source: Oliva, et al. 2010

Some properties of space
[Boundaries]

- Openness
  - A horizon line and the lack of visual references
- Expansion (Man-made scene)
  - The convergence of parallel lines gives the perception of the depth gradient of the space.
  - A flat view of a building → low degree of Expansion.
  - A street with long vanishing lines → high degree of Expansion
- Ruggedness (terrain)
  - Deviation of the ground with respect to the horizon
  - Open environments with a flat horizontal ground level vs. mountainous landscapes with a rugged ground
  - Rugged environment produces oblique contours in the picture and hides the horizon line.

Some properties of space
[Contents]

- Naturalness
  - Man-made structures: Straight horizontal and vertical lines
  - Natural landscapes: textured zones and undulating contours
- Roughness (for man-made scenes)
  - Size of basic components constituting the scene
  - correlated with the fractal dimension of the scene / its complexity

2D image and properties of space

- Spatial envelope properties $\rightarrow$ visual features of 2D surfaces.
- Thus, the parameters of the spatial envelop can be computed from the statistical distribution of (local) image features
- Spatial envelop is a compressed representation of an image capturing the scene properties

Computing spatial envelop

Global and local Texture descriptors

DFT:
(Global description)

$$I(f_x, f_y) = \sum_{x,y=0}^{N-1} i(x,y)h(x,y)e^{-j2\pi(f_x x + f_y y)} = A(f_x, f_y)\, e^{j\Phi(f_x, f_y)}$$

WFT:
(Local Decsription)

$$I(x, y, f_x, f_y) = \sum_{x',y'=0}^{N-1} i(x', y')\, h_r(x' - x, y' - y)e^{-j2\pi(f_x x' + f_y y')}$$

- *H(.,.) is a cicular Hanning window to take care of the boundary effect*
- ***<A*^{2}*, Φ >*** *represents energy spectrum of an image*
  - ***A*^{2}** *represents energy*
  - **Φ** *represents phase*

Use KLT for decomposition and PCA for dimensionality reduction

Energy Spectrum (global):

$$A(f_x, f_y)^2 \simeq \sum_{i=1}^{N_G} v_i \, \psi_i(f_x, f_y)$$

Spectogram (local):

$$A(x, y, f_x, f_y)^2 \simeq \sum_{i=1}^{N_L} w_i \, \Psi_i(x, y, f_x, f_y)$$

$N_G, N_L < 50$

$\mathbf{v} = \{v_i\}_{i=1,N_G}$: Provides a low-resolution global description of an image

$\mathbf{w} = \{w_i\}_{i=1,N_L}$: Provides a low-resolution description of of an image with spatial arrangement

Spectral signature

In real-world images, the energy spectra fall in average with a form

$1 / f^{\alpha}$ ($\alpha \sim 2$)

The energy spectrum of different types of images can be approximated by a function

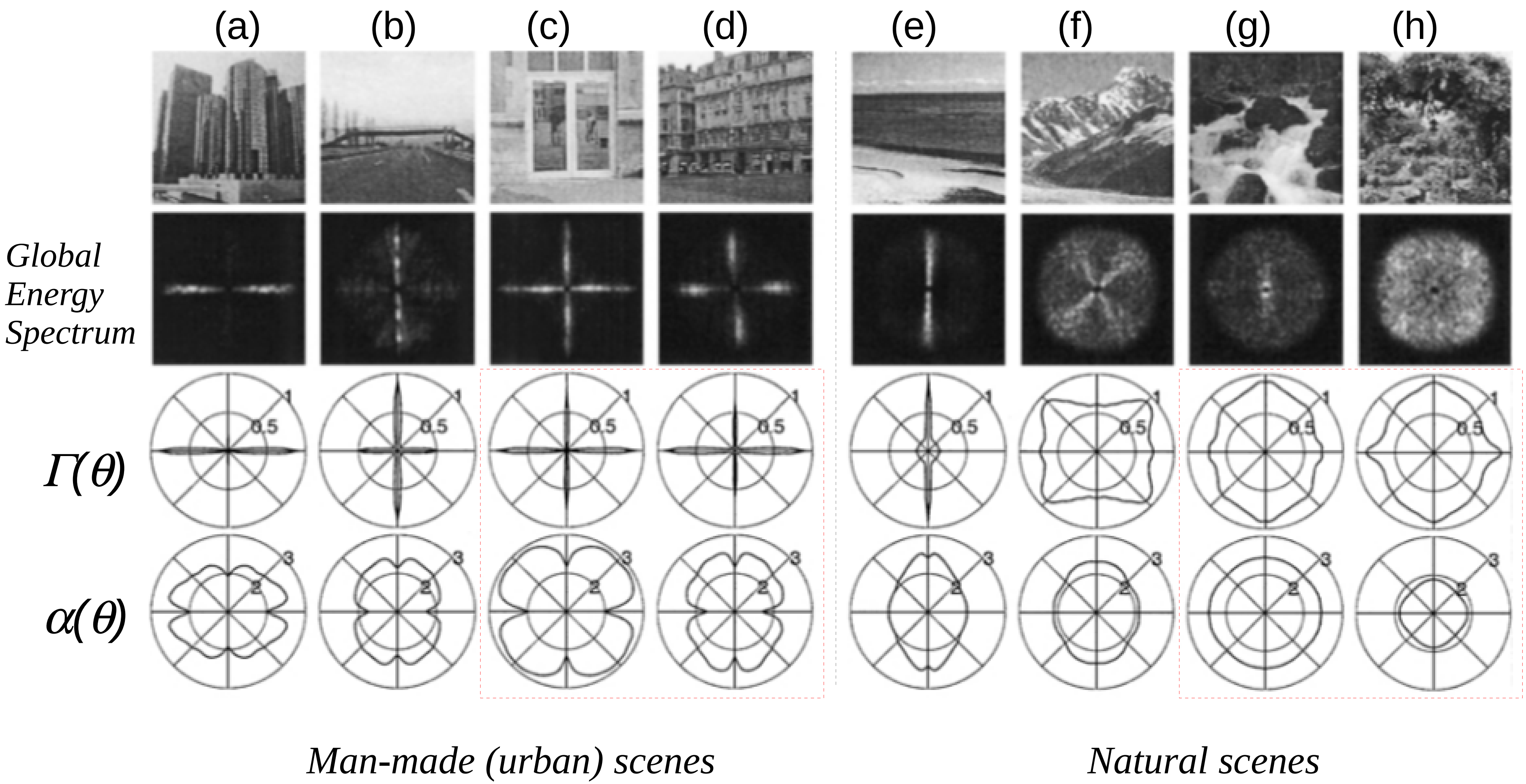
$$E[A(f,\theta)^2 \mid S] \simeq \Gamma_s(\theta)/f^{-\alpha_s(\theta)}$$

Expected energy spectrum given a category of image S

Parameters estimated through linear regression over a large number of images

- $\Gamma_s(\theta)$: Dominant orientations of a scene category
- $\alpha_s(\theta)$: Slope of the decreasing energy spectrum values
  -- related to roughness or complexity of the scene

Spectral signatures for scene categories



Source: Oliva & Torralba 2001

Estimation of spatial envelop properties

$\Gamma(\theta)$ and $\alpha(\theta)$ cannot be easily used to discriminate between properties of scene, e.g. naturalness, openness, etc.

A scene attribute can be expressed as

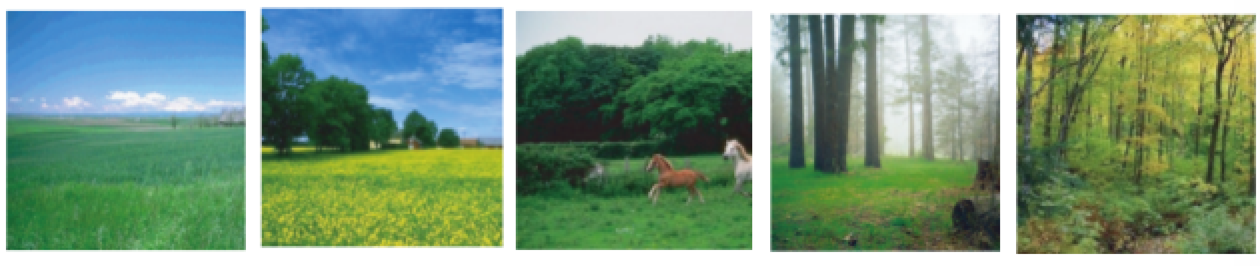
$$\hat{s} = \mathbf{v}^T \mathbf{d} = \sum^{N_G} v_i \, d_i$$

The parameters d_i can be estimated through simple linear regression after arranging many (~500) pictures in order of that attribute

... And similarly for spectogram (local descriptors)

DST: Discriminating Spatial Template (Global)
WDST: Windowed Discriminating Spatial Template (Local)

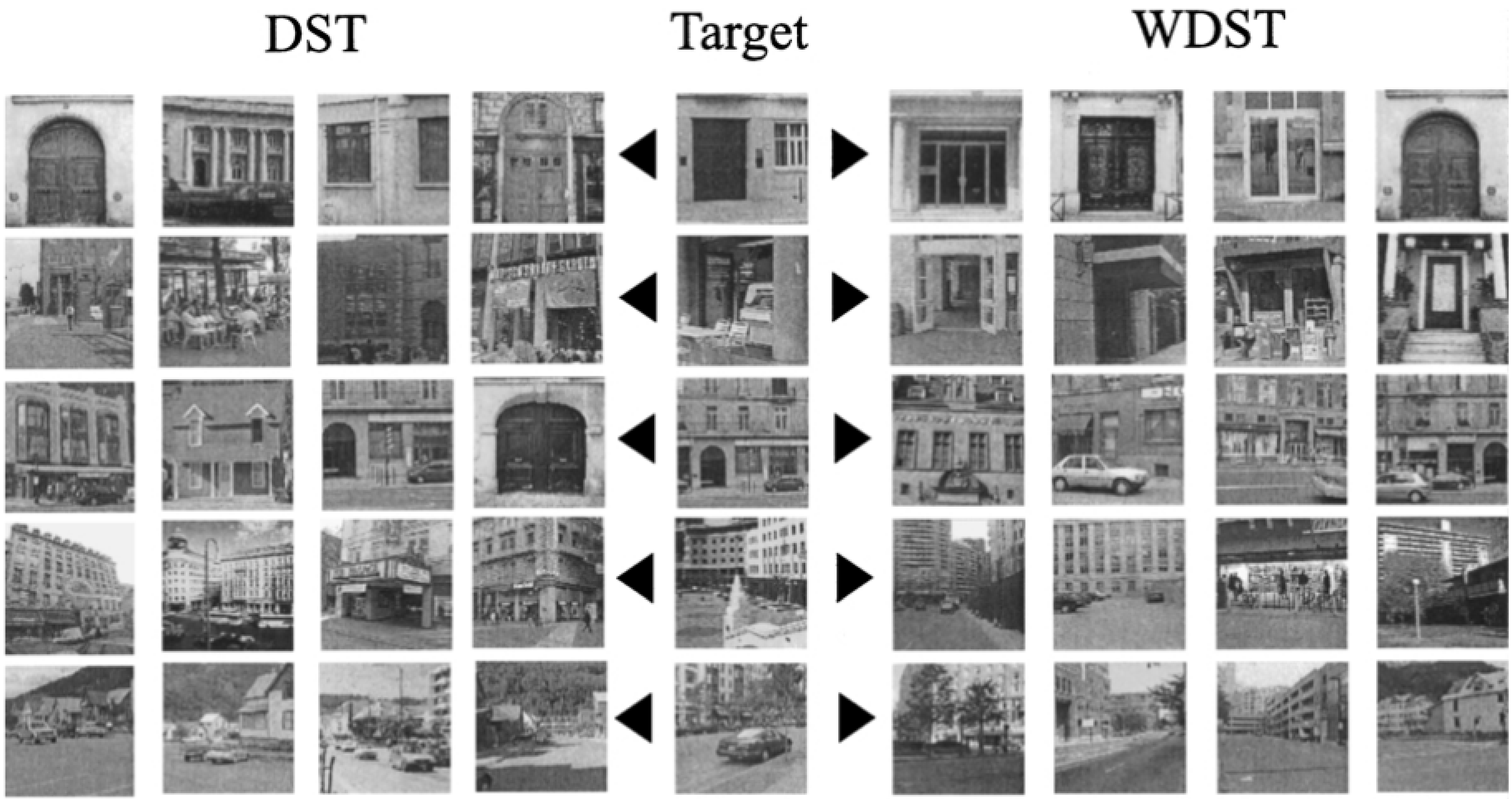
Example: Openness



Field (Open) ⟷ Forest (Closed)

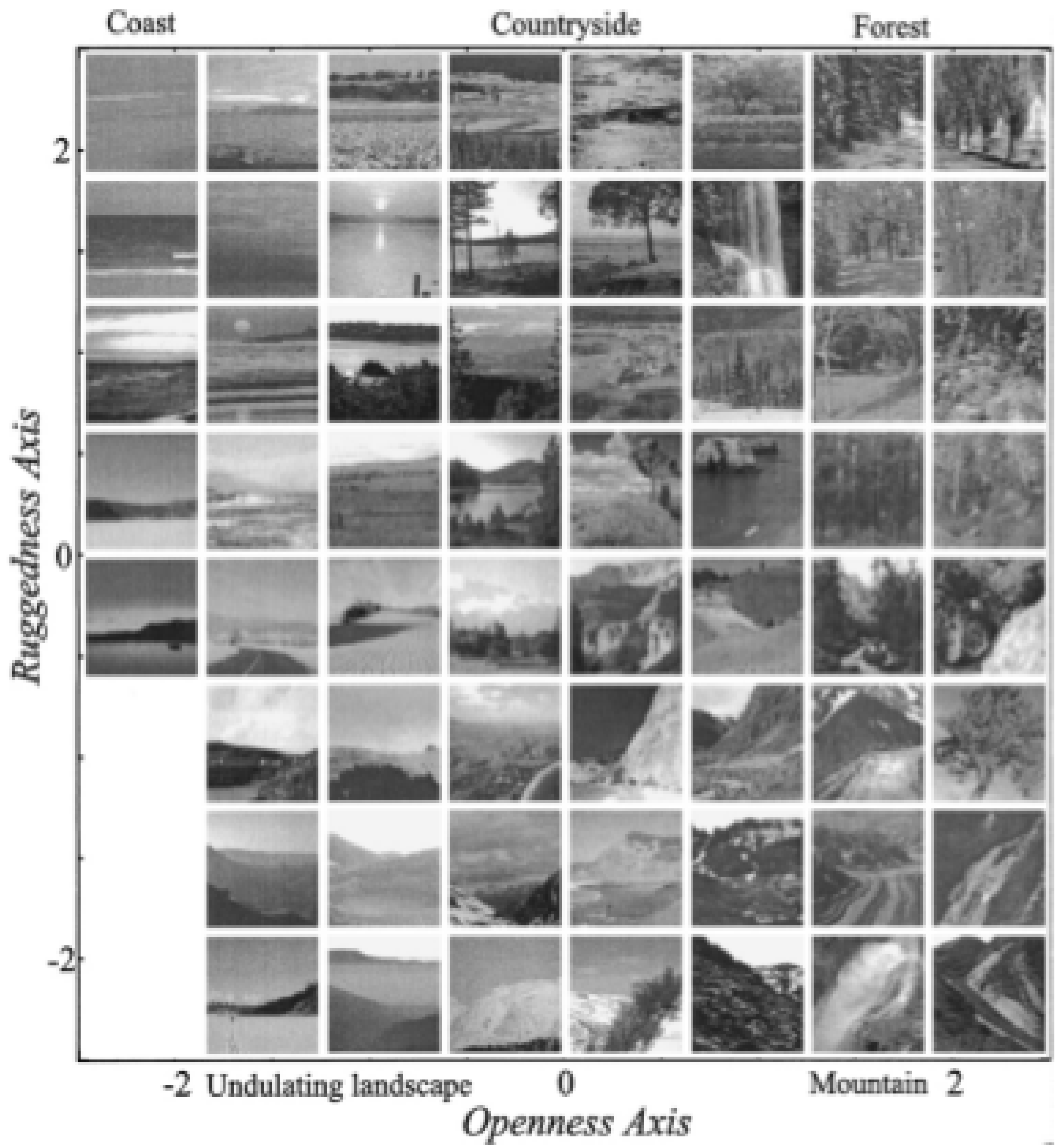
Source: Oliva, et al. 2010

Nearest neighbors (man-made scene) – with DST and WDST



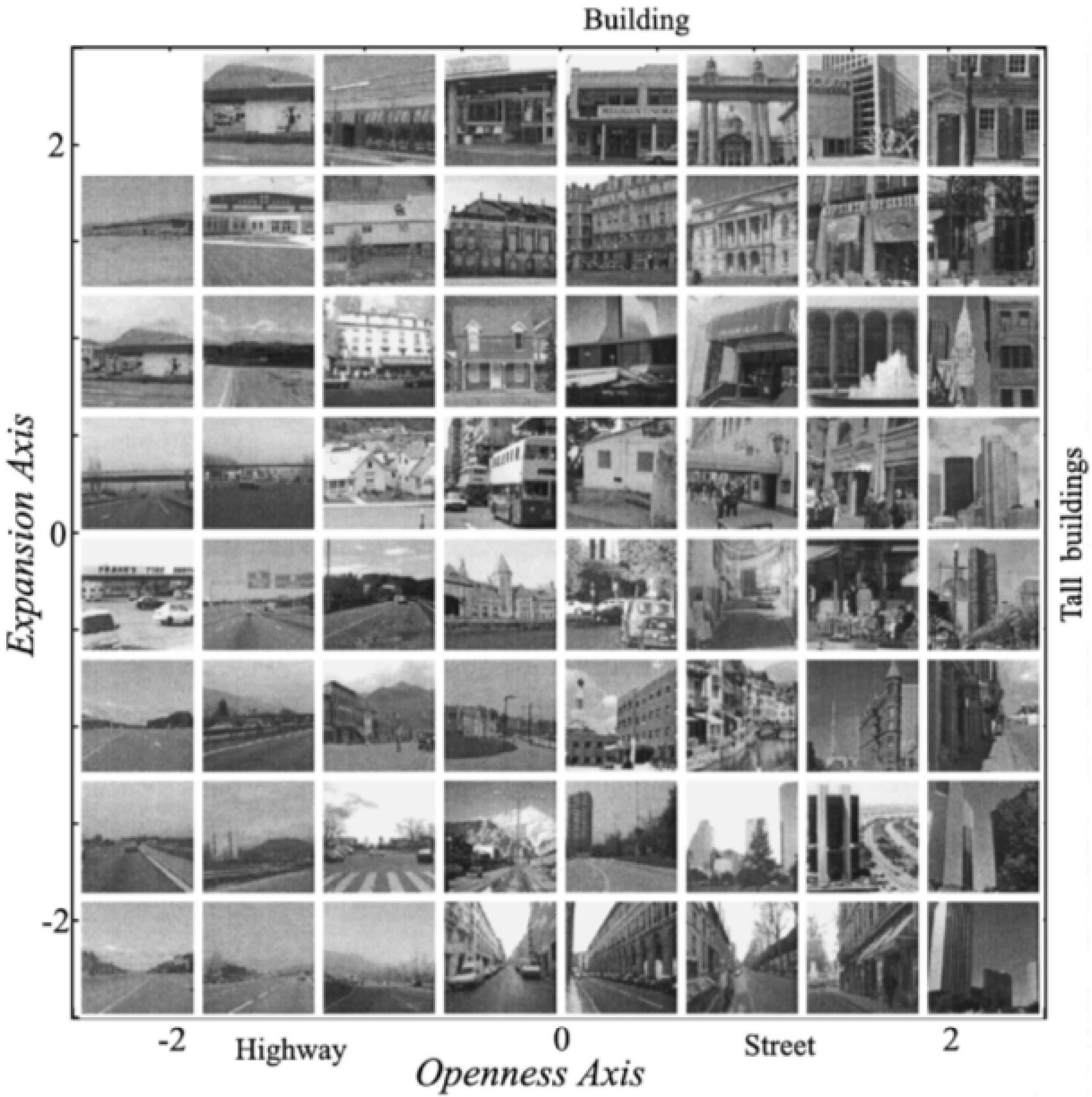
... WDST generally works better

Source: Oliva & Torralba, 2001



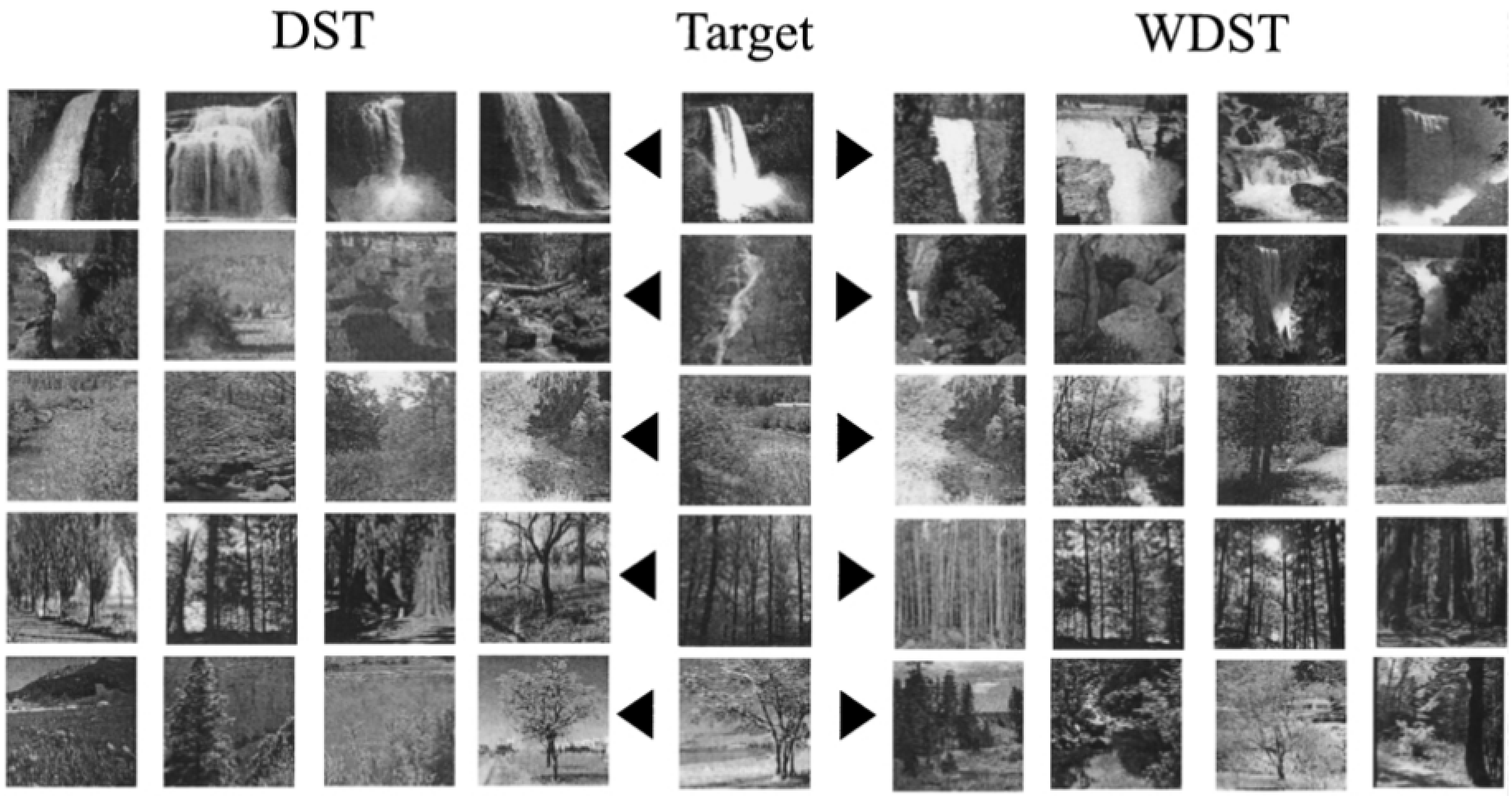
Organization of natural scenes in 2D (using WDST)

Source: Oliva & Torralba 2001



Organization of man-made scenes in 2D (using WDST)

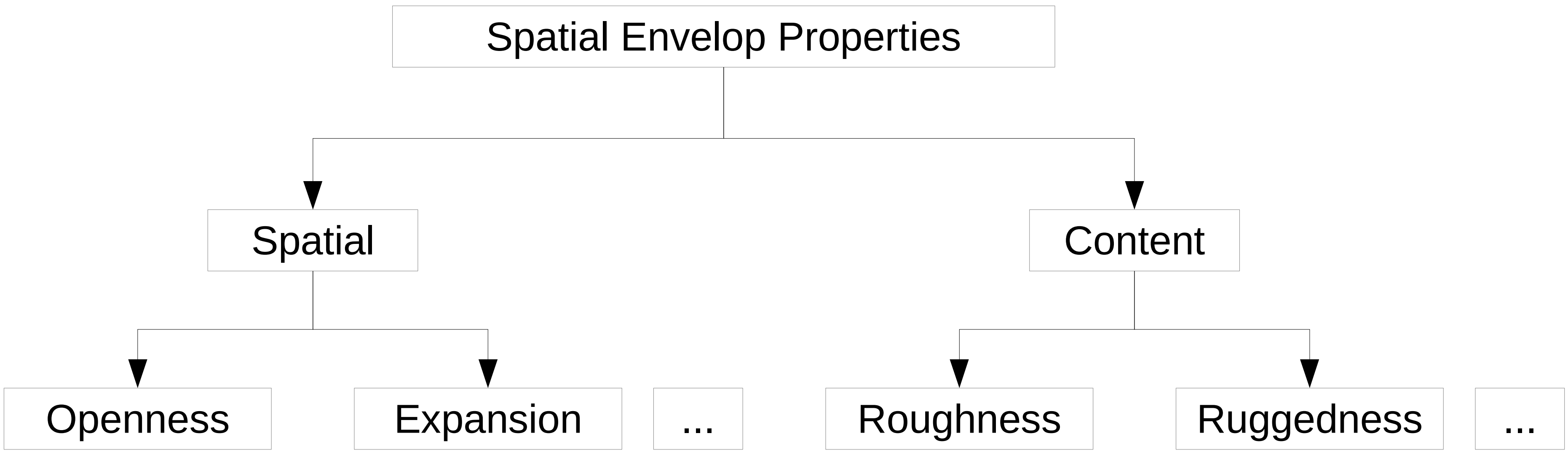
Source: Oliva & Torralba 2001



Examples of natural scenes with four neighbors sharing similar spatial envelope, estimated with the DST and the WDST procedures.

Source: Oliva & Torralba, 2001

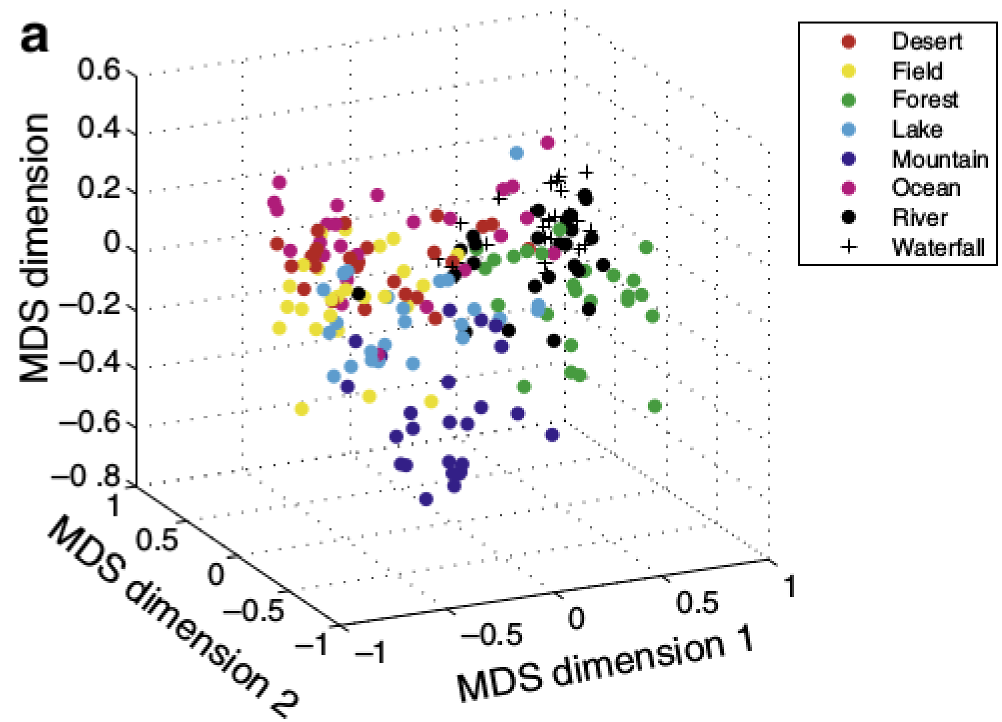
Summary: Spatial Envelop Representation



<u>In this framework</u>

Forest is a natural closed environment with a dense isotropic texture
NOT an environment with trees, bushes and leaves

Semantics of a scene emerges from a multi-dimensional projection on Spatial Envelop axes



Source: Green & Oliva, 2009

Target

Nearest neighbors

Source: Oliva, et al. 2010

- Space centered approach as opposed to object centered approach
  - Space is treated as a material object
- Provides interesting and complementary descriptors for space
  - Quantitative measures relate to perceptual properties
- Isovist
  - describes the visible volumes of a three-dimensional space
- Spatial envelope
  - captures layout and content features from a two-dimensional projected view

Comparison between Isovists and Spatial Envelope representations

- Space centered approach as opposed to object centered approach
  - Space is treated as a material object
- Provides interesting and complementary descriptors for space
  - Quantitative measures relate to perceptual properties

<u>Isovist</u>	**<u>Spatial envelope</u>**
Describes the visible volumes of a three-dimensional space	Captures layout and content features from a two-dimensional projected view

References

- Oliva, Park and Konkale. Representing, perceiving, and remembering the shape of visual space (Book Chapter), 2010 [overview]
- Benedict. To take hold of space ... 1979 [Isovist]
- Franz, et al. Exploring isovist-based correlates of spatial behavior and experience.
- Oliva and Torralba. Modeling the Shape of the Scene: ... Spatial Envelope, 2001 [Spatial Envelop Representation]