CS 4758/6758: Robot Learning

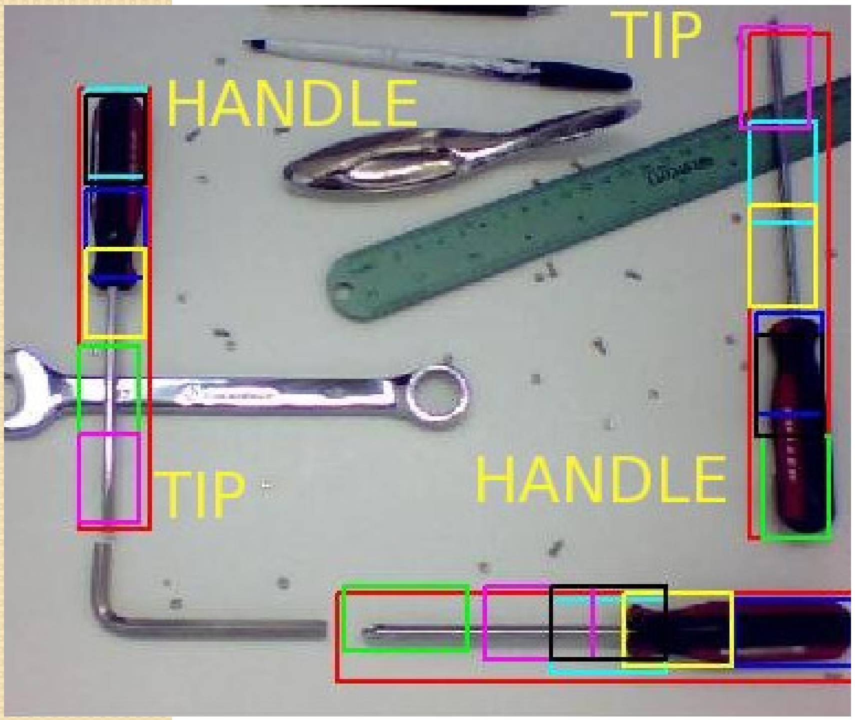
Spring 2010: Lecture 12

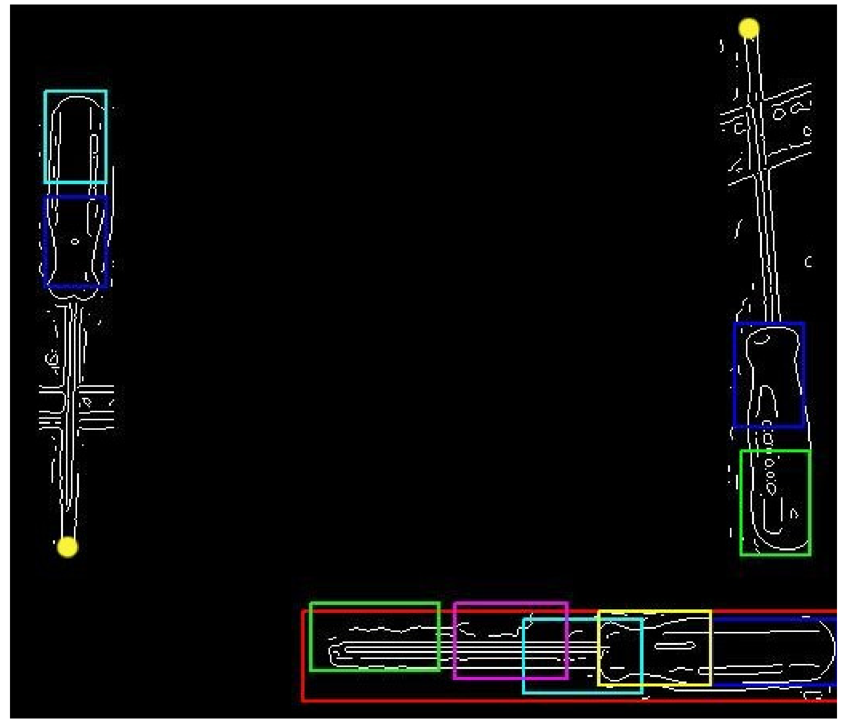
Ashutosh Saxena

Features

- Basic: Raw pixels, Color, Edges, Corners.
- Spatial Envelope Features.
- SIFT (Scale-Invariant Features)

Edge detection





- Convert a 2D image into a set of curves
  - Extracts salient features of the scene
  - More compact than pixels

Feature extraction: Corners and blobs

Features

- Basic: Raw pixels, Color, Edges, Corners.

Spatial Envelope Features

Oliva & Torralba (2006) propose a model that includes both local and global processing at different spatial scales
Local receptive fields are combined to form global receptive fields, without segmentation or grouping of low-level features

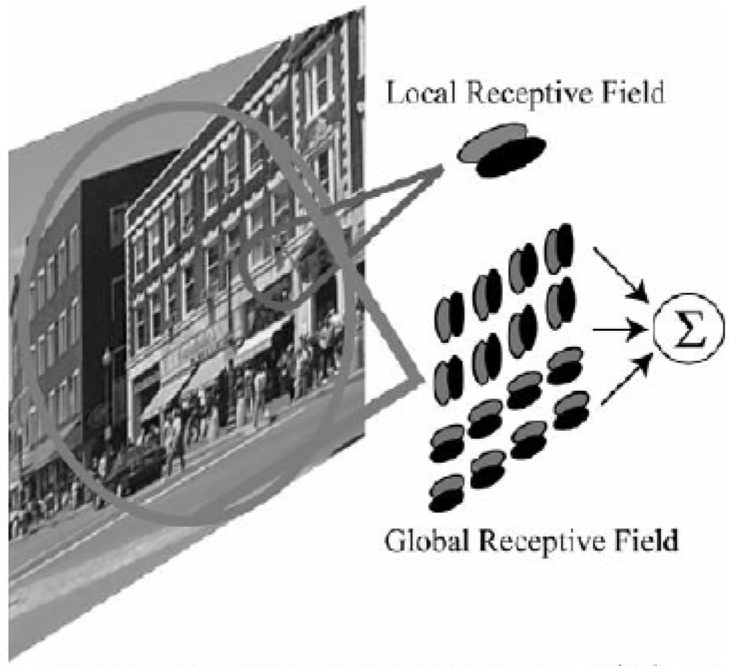


Figure 3: Illustration of a local receptive field and a global receptive field (RF). A local RF is tuned to a specific orientation and spatial scale, at a particular position in the image. A global RF is tuned to a spatial pattern of orientations and scales across the entire image. A global RF can be generated as a combination of local RFs and can, in theory, be implemented from a population of local RFs like the ones found in the early visual areas. Larger RFs, which can be selective to global scene properties, could be found in higher cortical areas (V4 or IT). The global feature illustrated in this figure is tuned to images with vertical structures at the top part and horizontal component at the bottom part, and will reply strongly to the scene street image.

Principal component analysis (for a particular feature such as orientation) is used to produce "global feature templates"

Different templates are created at different scales

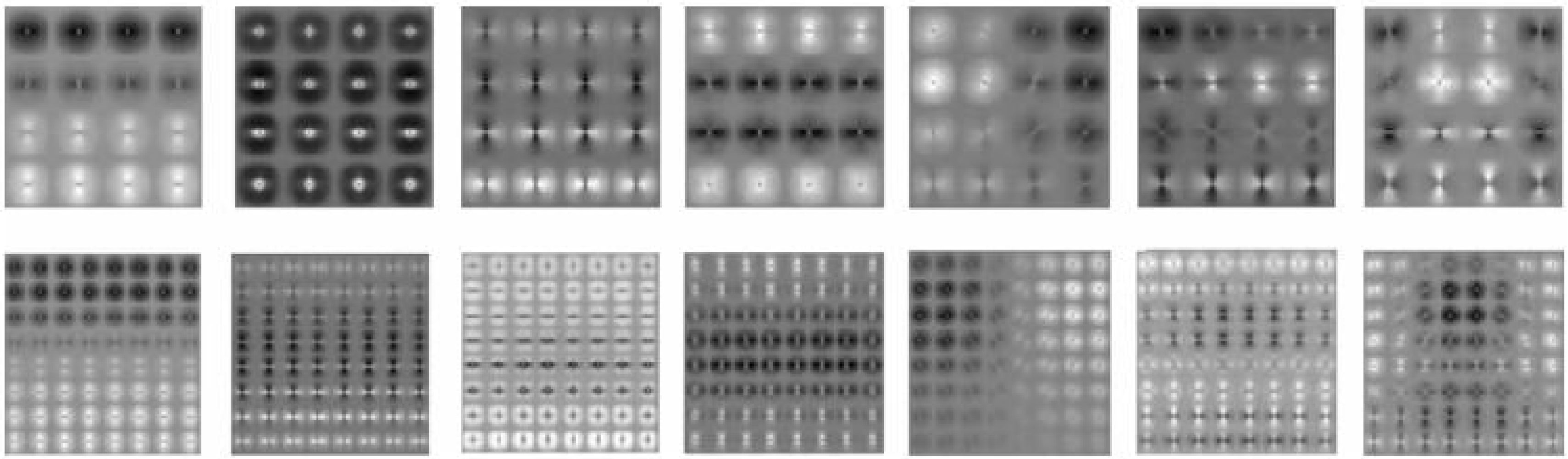


Figure 4. The Principal components of natural image statistics define the weights used to compute the global features. The set of weights are obtained by applying principal component analysis (PCA) to the responses of multiscale oriented filters to a large collection of natural images. The top row shows the 2nd to the 8th principal components for a spatial resolution of 2 cycles/image (4 x 4 regions). The first component behaves as a global average of the output of all orientations and scales and therefore it is not shown. The bottom row shows the PCs for a resolution of 4 cycles/image (8 x 8 regions). For each PC, each subimage shows, in a polar plot (low spatial frequencies are in the center of the plot), how the spatial scale and orientations are weighted at each spatial location. The white corresponds to positive value and the black to negative value. Here we refer to the PCs as global feature templates.

Oliva & Torralba (2001) propose that scenes can be defined in terms of properties such as:

openness, perspective, naturalness, expansion, depth, complexity, ruggedness, symmetry (to give the "spatial envelope" of a scene)

rather than by the type of object the scene contains

- Scenes with similar properties should have similar spatial envelopes
- Properties are determined by combining global feature templates

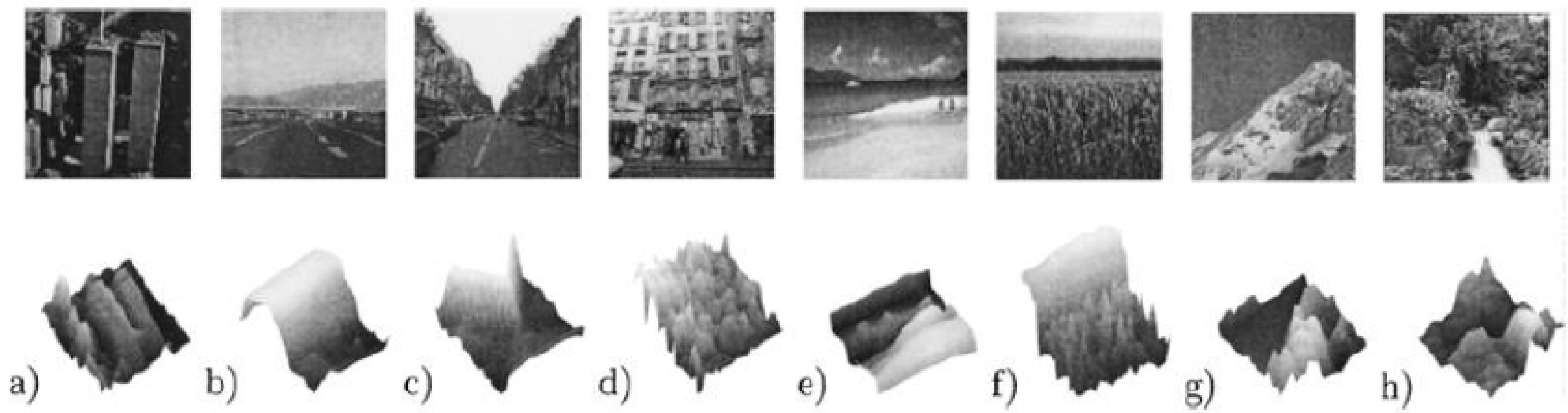


Figure 1. Scenes with different spatial envelopes and their surface representation, where the height level corresponds to the intensity at each pixel (images were low-passed): a) skyscrapers, b) an highway, c) a perspective street, d) view on a flat building, e) a beach, f) a field, g) a mountain and e) a forest. The surface shows the information really available after projection of the 3D scene onto the camera. Several aspects of the 3D scene have a direct transposition onto the 2D surface properties (e.g., roughness).

Robert Collins
CSE486, Penn State

SIFT Keys: General Idea

- Reliably extract same image points regardless of new magnification and rotation of the image.
- Normalize image patches, extract feature vector
- Match feature vectors using correlation

Robert Collins
CSE486, Penn State

SIFT Keys: General Idea

Want to detect/match same features regardless of

Translation : easy, almost every feature extraction and correlation matching algorithm in vision is translation invariant

Rotation : harder. Guess a canonical orientation for each patch from local gradients

Scaling : hardest of all. Create a multi-scale representation of the image and appeal to scale space theory to determine correct scale at each point.

Robert Collins
CSE486, Penn State

SIFT Vector

- Thresholded image gradients are sampled over 16x16 array of locations in scale space
- Create array of orientation histograms
- 8 orientations x 4x4 histogram array = 128 dimensions

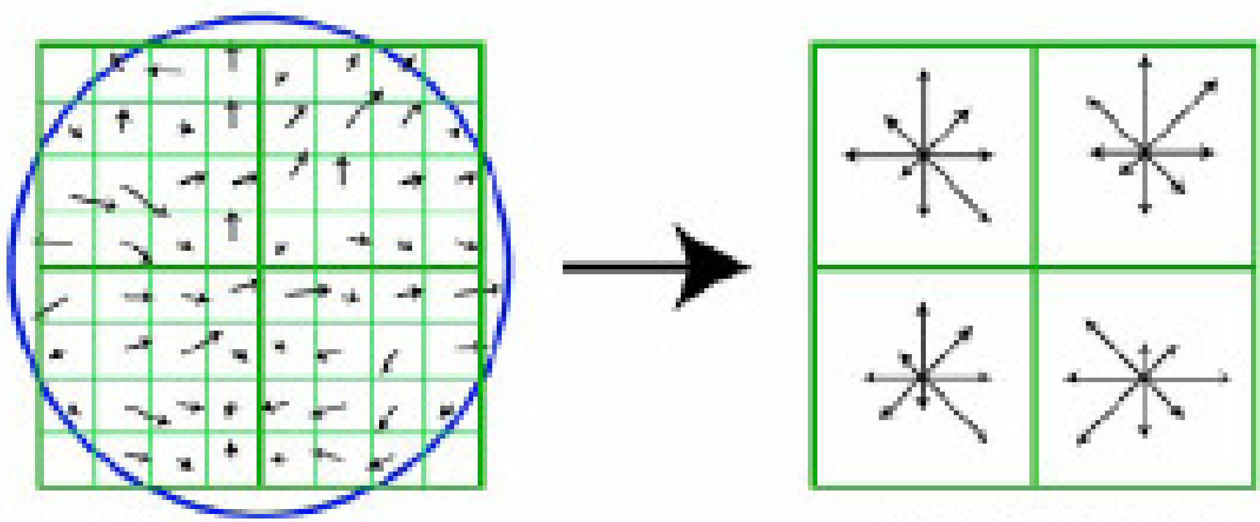


Image gradients Keypoint descriptor

Robert Collins
CSE486, Penn State

Example of KeyPoint Detection

Each keypoint has a center point (location), an orientation (rotation) and a radius (scale).

At this point, we could try to correlate patches (after first normalizing to a canonical orientation and scale).

Robert Collins
CSE486, Penn State

Application: Object Recognition

Note: since these are local, parts-based descriptors, they perform well even when some parts are missing (i.e. under occlusion).