Filters and other potions

P. Perona - Caltech
MIT - 21 November 2013
what
where
Architectures
Architecture 1

The vision black box

Image(s)

Feature extraction:
- texture
- stereo disparity
- color contrast
- motion flow
- edgels
...

Grouping:
- image regions

Perceptual organization:
- 2.5D sketch:
  - boundaries, junctions,
  - foreground, background

Surface shape, scene depth, spatial relationships, 3D motion

Recognition, surface properties

Objects, verbs, categories...

Motor cognition

[Marr '82]
features?

Le Corbusier, Villa Savoye

http://flickr.com/photos/ikura/1398271367/
edges

Le Corbusier, Villa Savoye
http://www.iit.edu/~stawraf/perspx.jpg
Matched filtering

\[ \max_{\|k\| = 1} \langle I^*, k \rangle \]
Matched filtering

\[ \langle f, g \rangle \triangleq \int_{[a,b]} f(x) g(x) \, dx \quad \triangleq \sum_i f(x_i) g(x_i) \]

\[ \langle f, g(-x) \rangle = f * g \triangleq \int_{[a,b]} f(t) g(t-x) \, dt \quad \triangleq \sum_i f(x_i) g(x_{i-j}) \]
Matched filtering 2D

\[ J = \max_{x,y,\theta} \langle I, k_\theta \rangle \]
Architecture 2

pattern recognition
forward (afferent)

Stimulus

backward (efferent)
selective attention

[Fukushima '80]
[DeValois '85]
Column
Hypercolumn
Dense sampling
Fig. 2. Architecture of LeNet-5, a Convolutional Neural Network, here for digits recognition. Each plane is a feature map, i.e. a set of units whose weights are constrained to be identical.

[LeCun et al. 1998]
Figure 1: For each octave of scale space, the initial image is repeatedly convolved with Gaussians to produce the set of scale space images shown on the left. Adjacent Gaussian images are subtracted to produce the difference-of-Gaussian images on the right. After each octave, the Gaussian image is down-sampled by a factor of 2, and the process repeated.

[Lowe 2004]
translation, rotation, scale invariance

[Hinton et al. '12]
Texture gradient \((x,y)\)

\[
\begin{align*}
&\text{max} \\
&\text{PIR}_{1}^{(x,y)} \quad \text{PIR}_{2}^{(x,y)} \\
&\text{PIR}_{n}^{(x,y)} \\
&\text{R}_{1}^{(x,y)} \quad \text{R}_{2}^{(x,y)} \\
&\text{R}_{n}^{(x,y)} \\
&\ast F_{1} \quad \ast F_{j} \\
&\ast F_{m} \\
&I^{(x,y)}
\end{align*}
\]

96 filters
6 orientations
2 center-surround
14 scale samples over 2.2 binary octaves
Detection Performance

Caltech pedestrians: 1M frames, 250K hand-annotated
Detection Performance

(e) Caltech-Train

(a) INRIA [7]
Detection Performance

Viola & Jones ‘01
Dalal-Triggs ‘05
Dollar et al. ‘08
Dollar et al. ‘10
Walk et al. ‘10

log-average miss rate vs. frames per second

- 0.004 MultiFtr+Motion
- 0.005 Pls
- 0.005 MultiFtr+CSS
- 0.010 Shapelet
- 0.014 HogLbp
- 0.017 MultiFtr
- 0.020 FtrMine
- 0.036 HikSvm
- 0.054 HOG
- 0.089 VJ
- 0.098 LatSvm-V1
- 0.101 PoseInv
- 0.164 LatSvm-V2
- 0.278 ChnFtrs
- 2.670 FPDW
filter technology
Scale, orientation, elongation....
lots of CPU cycles
how do we make computations efficient?
Separability

Cost = \text{m} \times \text{n}

R(i, j) = \sum_{h=1:M, k=1:N} k(h, k) I(i - h, j - k)

R(i, j) = \sum_{h=1:M} \sum_{k=1:N} k(h)k'(k) I(i - h, j - k)

Cost = \text{m} + \text{n}

[Adelson & Bergen, ’85]
Separability and decomposition

[Adelson & Bergen, '85]
Steerability

[Freeman & Adelson, '91]
General decomposition

\[ k(x, \theta) = \sum_{i=1}^{D} b_i(\theta) f_i(x) \]

\[ k(x, y) = \sum_{i=1}^{D} f_i(x) g_i(y) \]

\[ k(x, y; \theta) = \sum_{i=1}^{D} b_i(\theta) f_i(x) g_i(y) \]
Design?
\[ A = U S V^T \]
Approximation

\[ K(x, y; \theta) = \sum_{i=1}^{D} b_i(\theta) f_i(x, y) \]

\[ K(x, y; \theta) \approx \sum_{i=1}^{R} b_i(\theta) f_i(x, y) \quad R \ll D \]
Theorem 5. Let $L : A \rightarrow B$ be a linear compact operator between two Hilbert spaces. Let $a_i, b_i, \sigma_i$ be the singular value decomposition of $L$, where the $\sigma_i$ are in decreasing order of magnitude. Then

1) An optimal $n$-dimensional approximation to $L$ is

$$L_n = \sum_{i=1}^{n} \sigma_i a_i b_i$$

2) The approximation error is

$$\delta_n(L) = \sigma_{n+1}, \Delta_n^2(L) = \sum_{i=n+1}^{N} \sigma_i^2.$$
Theorem 2. Let $X$ and $T$ be locally compact Hausdorff spaces and $F \in L_2(X \times T)$. Then $L$ is well defined and is a compact operator.

Such a kernel is commonly called a Hilbert-Schmidt kernel.

A second result tells us that if a linear operator is compact, then it has a discrete spectrum (see [54]):

Theorem 3. Let $L$ be a compact operator on (complex) normed spaces, then the spectrum $S$ of $L$ is at most denumerable.

A third result says that if $L$ is continuous and operates on Hilbert spaces then the compactness property transfers to the adjoint of $L$ (see [54]):

Theorem 4. Let $L$ be a compact operator on Hilbert spaces, then the adjoint $L^*$ is compact.
reconstruction error v.s. components number

\[ \log(\text{error}) \]

\[ \text{n. components} \]

[4 components]

[9 components]

[15 components]
reconstruction error vs components number

log(error)

n. components

(gaus-1)

(gaus-2)

(gaus-3)

[Perona '95]
Tensor Factorization

\[ k(x, y; \theta) = \sum_{i=1}^{D} b_i(\theta) f_i(x) g_i(y) \]

- Not a convex problem
- Gradient descent

[Shy, Perona '96]
Including scale by resampling
Fig. 13. Pyramidal scheme for the 2-D multirate separable scalable/steerable decomposition.

[Manduchi et al. '98]
[cfr. Simoncelli et al]
Exploiting Image Statistics
sampling the gradient

(a) upsampling gradients (2x)

upsampling 2x

original

upsampled
(b) downsampling gradients (2x)

M → Σ_{ij} → h

≈ 0.34x

downsample 2x

M → Σ_{ij} → h

≈ 0.27x

downsample 2x

(c) downsampling normalized gradients (2x)

\tilde{M} → Σ_{ij} → h

≈ 0.34x

downsample 2x

\tilde{M} → Σ_{ij} → h

≈ 0.27x

downsample 2x

\[\text{probabilities} \quad \mu=0.33, \sigma=0.039\]

\[\text{probabilities} \quad \mu=0.26, \sigma=0.020\]

\[\text{probabilities} \quad \mu=0.34, \sigma=0.059\]

\[\text{probabilities} \quad \mu=0.27, \sigma=0.040\]

[Dollar et al. 2013]
Gradient histograms

[Dollar et al. 2013]
Power law feature scaling

(a) histograms of gradients

(b) histograms of normalized gradients

(d) grayscale images

(e) local standard deviation
Power law feature scaling

(b) histograms of normalized gradients

(c) difference of gaussians (DoG)

(e) local standard deviation

(f) HOG [21]
Individual images

(a) histograms of gradients

(b) histograms of normalized gradients

(d) grayscale images

(e) local standard deviation

[Dollar et al. 2013]
Fast computations

Standard Pipeline

1 → 1/2 → 1/4

Proposed Pipeline

1 → 1/2 → 1/4
Fast computations

(a) dense image pyramid
(b) classifier pyramid
(c) hybrid approach

[Dollar et al. 2013]
Performance

[Dollar et al. 2013]
Conclusions

• Filtering front-end
• Need fine sampling of scale, orientation, ...
• Scalable, separable and steerable approximations
• Exploiting image statistics to extrapolate
• Fast and accurate detection