Visual Recognition

Jianbo Shi

Computer and Information Science
University of Pennsylvania
Objects
- Face
- Pedestrian
- Car
- Cow
- Hand
- Chair

Scenes
- Mountain
- Beach
- Forest
- Highway
- Street
- Indoor

Objects in scenes
- Animal in natural scene
- Tree in urban scene
- Close-up person in urban scene
- Far pedestrian in urban scene
- Car in urban scene
- Lamp in indoor scene
Texture Patch Types

- Simple: clean step edge
- Textured: on either side, or step with noise
- Complex: wrong scale, or just a mess
- Invisible: boundary but no edge
Simple Patches
Textured Patches
Complex Patches
Filter Banks
Filter Bank
Dot filter response
Odd symmetric filter outputs
Even symmetric filter
Pooling
Input: an image patch

Output:
1) max over the region,
2) average over the region
Pooling using ave. filter bank response
Average filter bank response

squared responses

vertical

classification

horizontal

smoothed mean
Is mean of filter outputs sufficient?
Histogram of filter banks

- A histogram is a mapping from a set of d-dimensional integer vector $i$ to nonnegative real numbers.

$$f^r(i; I) = \left| \{ \bar{x} : t^r_{i-1} < I^r(\bar{x}) \leq t^r_i \} \right| .$$

The vector $i$ represents the bins in the relevant region of underlying feature space, defined by $I(x)$.
Adaptive binning: location of bins depends on the data itself,
Centers: are defined as prototypes \( \{ c_i \} \) and
Bins: are defined as the corresponding Voronoi tessellation.

\[
h_i = \left| \{ \mathbf{x} : i = \arg \min_j \|I(\mathbf{x}) - \mathbf{c}_j\| \} \right|.
\]
• For image contain a small amount of information, a finely quantized histogram is highly inefficient. But a too coarsely defined bin is also bad usually. Adaptive binning can achieve a good balance.
Texton: assign a label to each pixel
Pooling over texton

\[ f^r(i; I) = |\{x : t_{r-1}^r < f^r(x) \leq t_r^r\}|. \]

Adaptive binning for filter outputs
How to compare histograms?

\[ \chi^2(h_i, h_j) = \frac{1}{2} \sum_{m=1}^{K} \frac{[h_i(m) - h_j(m)]^2}{h_i(m) + h_j(m)} \]
2.2.1.1 Metric Space

A space $\mathcal{A}$ is called a metric space if for any of its two elements $x$ and $y$, there is a number $\rho(x, y)$, called the distance, that satisfies the following properties

- $\rho(x, y) \geq 0$  \hspace{1em} (non-negativity)
- $\rho(x, y) = 0$ if and only if $x = y$  \hspace{1em} (identity)
- $\rho(x, y) = \rho(y, x)$  \hspace{1em} (symmetry)
- $\rho(x, z) \leq \rho(x, y) + \rho(y, z)$  \hspace{1em} (triangle inequality)
Heuristic Histogram distances

(i) The *Minkowski-form distance* $\mathcal{L}_p$ is defined by:

$$D(I, J) = \left( \sum_i |f(i; I) - f(i; J)|^p \right)^{1/p}.$$
Image similarity with L1 distance

1) 0.00
2) 0.53
3) 0.61
4) 0.61

5) 0.63
6) 0.67
7) 0.70
8) 0.70
Non-parametric test statistics

The $\chi^2$-statistic is given by

$$D(I, J) = \sum_i \frac{(f(i; I) - \hat{f}(i))^2}{\hat{f}(i)}.$$
Image similarity w. chi-sqr statistics
Information-theoretic divergences

(i) The *Kullback–Leibler divergence* (KL) suggested in [10] as an image dissimilarity measure is defined by

$$ D(I, J) = \sum_i f(i; I) \log \frac{f(i; I)}{f(i; J)} . $$

(ii) The *Jeffrey–divergence* (JD) is defined by

$$ D(I, J) = \sum_i f(i; I) \log \frac{f(i; I)}{\hat{f}(i)} + f(i; J) \log \frac{f(i; J)}{\hat{f}(i)} . $$
Image similarity with Jeffrey divergence
Perceptual similarity

- Quadratic form
  
  \[ D(I, J) = \sqrt{(\vec{f}_I - \vec{f}_J)^T A (\vec{f}_I - \vec{f}_J)} , \]

- Earth Moving Distance
Image similarity w. quadratic-form
Problems with Binning
Problem with quadratic norm
Image similarity with Earth Moving Distance (EMD)

1) 0.00
29020.jpg
2) 8.16
29077.jpg
3) 12.23
29005.jpg
4) 12.64
29017.jpg

5) 13.82
20003.jpg
6) 14.52
53062.jpg
7) 14.70
29018.jpg
8) 14.78
29019.jpg
Earth Moving Distance

- Let $P$, $Q$ to be 2 histogram signature:
  - $P = \{(P_1, w_{p1}), \ldots, (P_m, w_{pm})\}$
  - $Q = \{(Q_1, w_{q1}), \ldots, (Q_n, w_{qn})\}$

- Find a optimal mapping from $P$ to $Q$

- Define a flow $F(i, j)$ so to minimize

\[
\text{WORK}(P, Q, F) = \sum_{i=1}^{m} \sum_{j=1}^{n} d_{ij} f_{i,j}.
\]
Earth Moving Distance

- Define a flow $F(i,j)$ so to minimize

$$\text{WORK}(P, Q; F) = \sum_{i=1}^{m} \sum_{j=1}^{n} d_{i,j} f_{i,j}.$$ 

- Such that,

$$f_{i,j} \geq 0 \quad 1 \leq i \leq m, \ 1 \leq j \leq n$$

$$\sum_{j=1}^{n} f_{i,j} \leq w_{p,i} \quad 1 \leq i \leq m$$

$$\sum_{i=1}^{m} f_{i,j} \leq w_{q,j} \quad 1 \leq j \leq n$$

$$\sum_{i=1}^{m} \sum_{j=1}^{n} f_{i,j} = \min(\sum_{i=1}^{m} w_{p,i}; \sum_{j=1}^{n} w_{q,j}).$$
\[
\begin{array}{cccc}
| f_{11} & \cdots & f_{1n} & w_{p_1} \\
| \vdots & \ddots & \vdots & \vdots \\
| f_{m1} & \cdots & f_{mn} & w_{p_m} \\
| w_{q_1} & \cdots & w_{q_n} & \cdot \\
\end{array}
\]
Earth Moving Distance

- Define a flow $F(i,j)$ so to minimize

$$\text{WORK}(P, Q, F) = \sum_{i=1}^{m} \sum_{j=1}^{n} d_{ij} f_{i,j}$$

- Earth moving distance is,

$$\text{EMD}(P, Q) = \frac{\sum_{i=1}^{m} \sum_{j=1}^{n} d_{ij} f_{i,j}}{\sum_{i=1}^{m} \sum_{j=1}^{n} f_{i,j}}$$
Theorem 3.1 If two signatures, $P$ and $Q$, have equal weights and the ground distance $d(p_i, q_j)$ is metric for all $p_i$ in $P$ and $q_j$ in $Q$, then $EMD(P, Q)$ is also metric.

Proof: To prove that a distance measure is metric, we must prove the following: positive definiteness ($EMD(P, Q) \geq 0$ and $EMD(P, Q) = 0$ iff $P \equiv Q$), symmetry ($EMD(P, Q) = EMD(Q, P)$), and the triangle inequality (for any signature $R$, $EMD(P, Q) \leq EMD(P, R) + EMD(R, Q)$).
How to compute EDM

- Max-flow
- Hungarian method:
  - http://mathlab.usc.edu/matlab/toolbox/fdident/pairs.html
- Linear programming
- Running time
comparision

Jeffrey divergence

EMD
X: # image retrieved,
Y: # relevant images