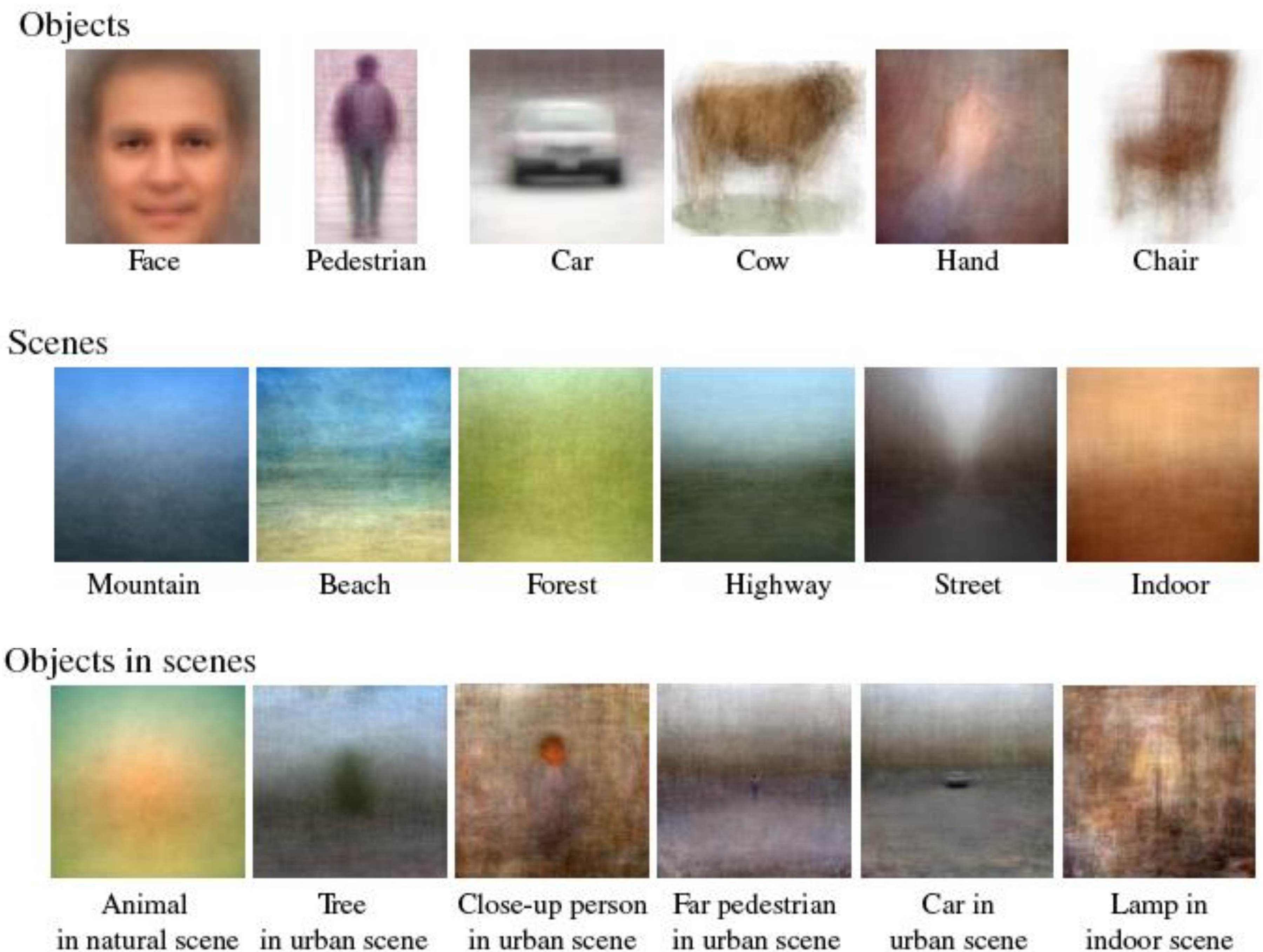
Visual Recognitio

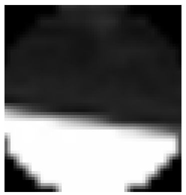
Jianbo Shi

Computer and Information Science
University of Pennsylvania

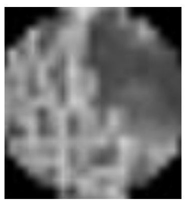


Objects
Face
Pedestrian
Car
Cow
Hand
Chair
Scenes
Mountain
Beach
Forest
Highway
Street
Indoor
Objects in scenes
Animal in natural scene
Tree in urban scene
Close-up person in urban scene
Far pedestrian in urban scene
Car in urban scene
Lamp in indoor scene

Texture Patch Types



- Simple: clean step edge

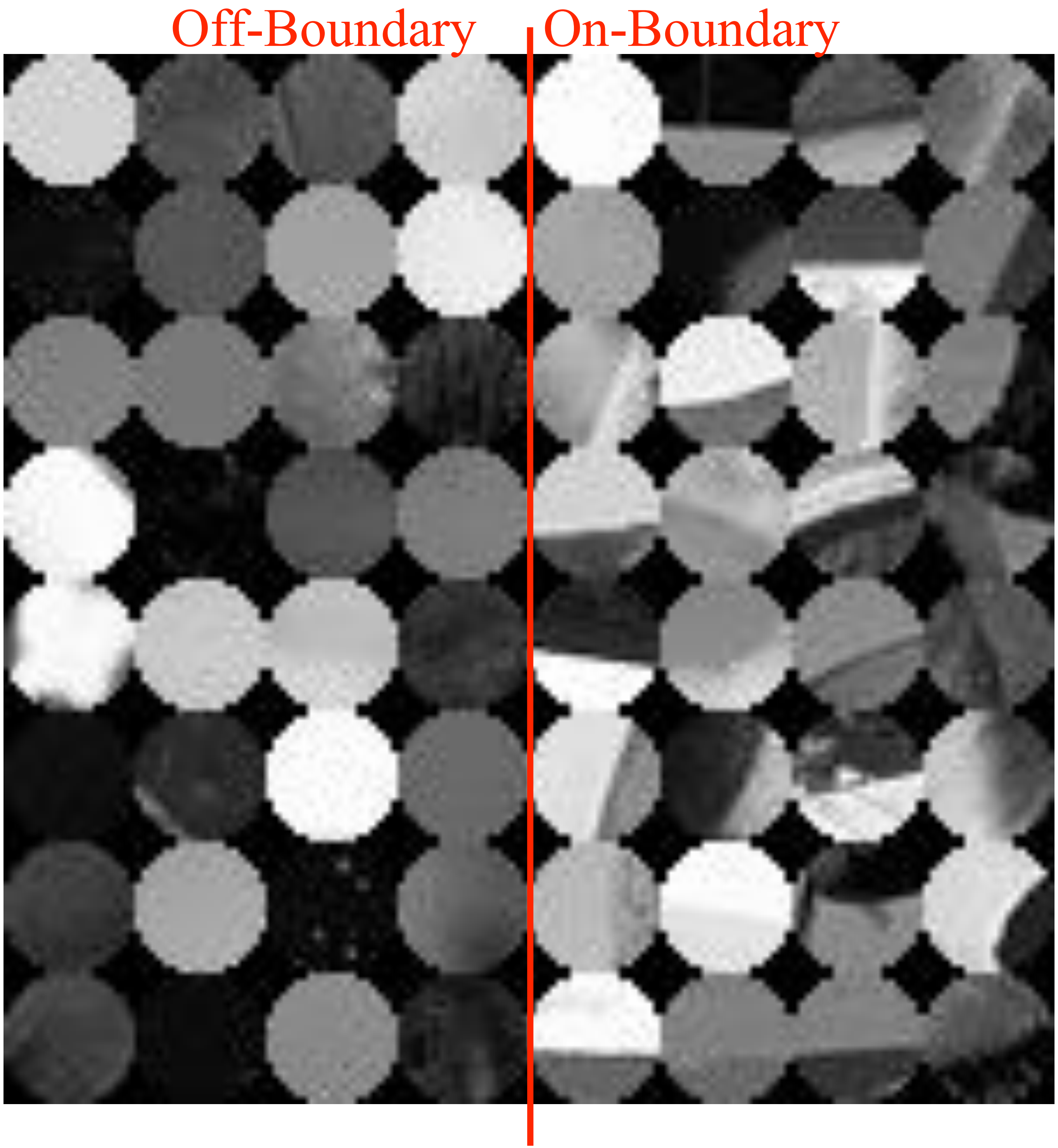


- Textured: on either side, or step with noise

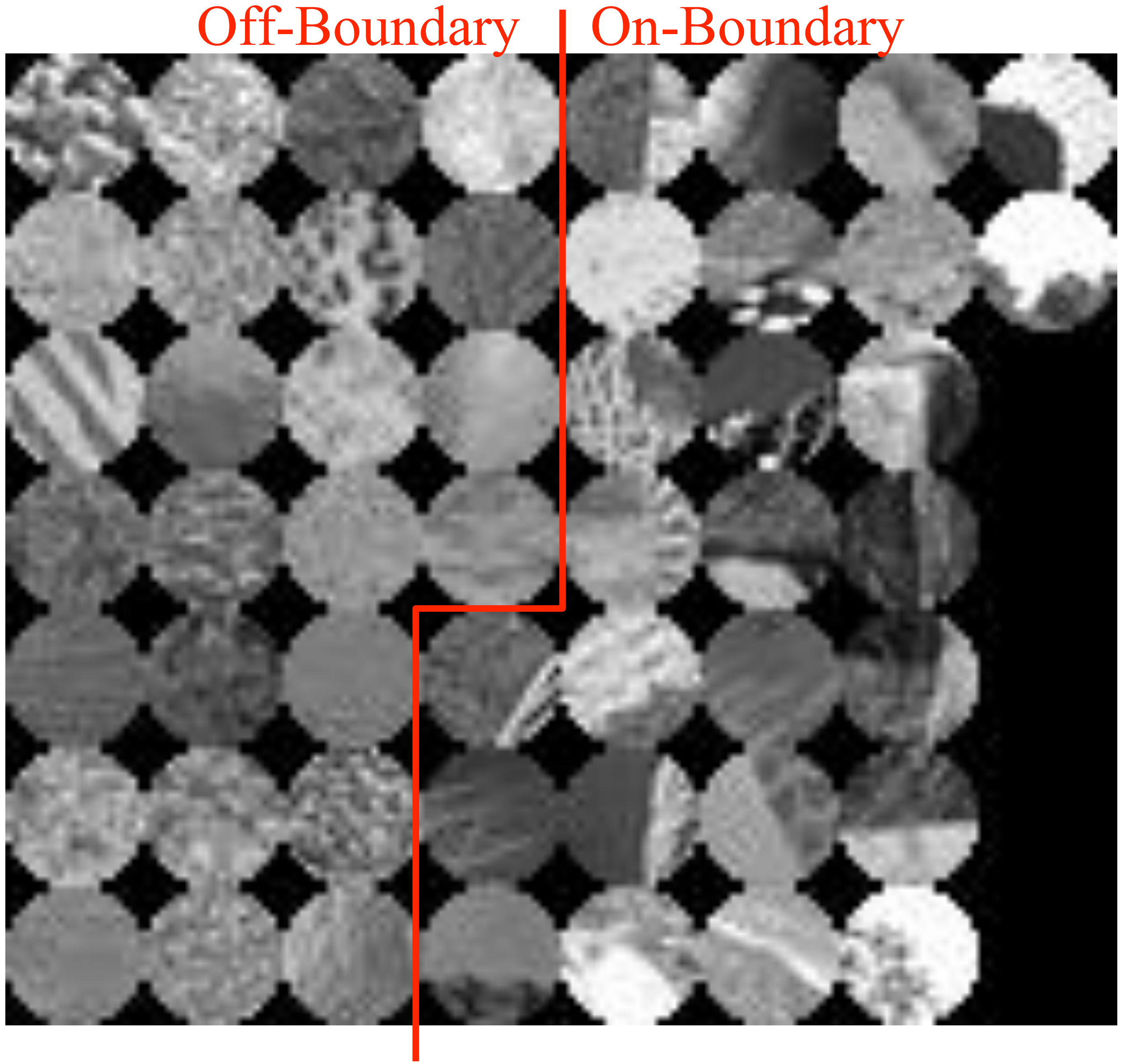
- Complex: wrong scale, or just a mess

- Invisible: boundary but no edge

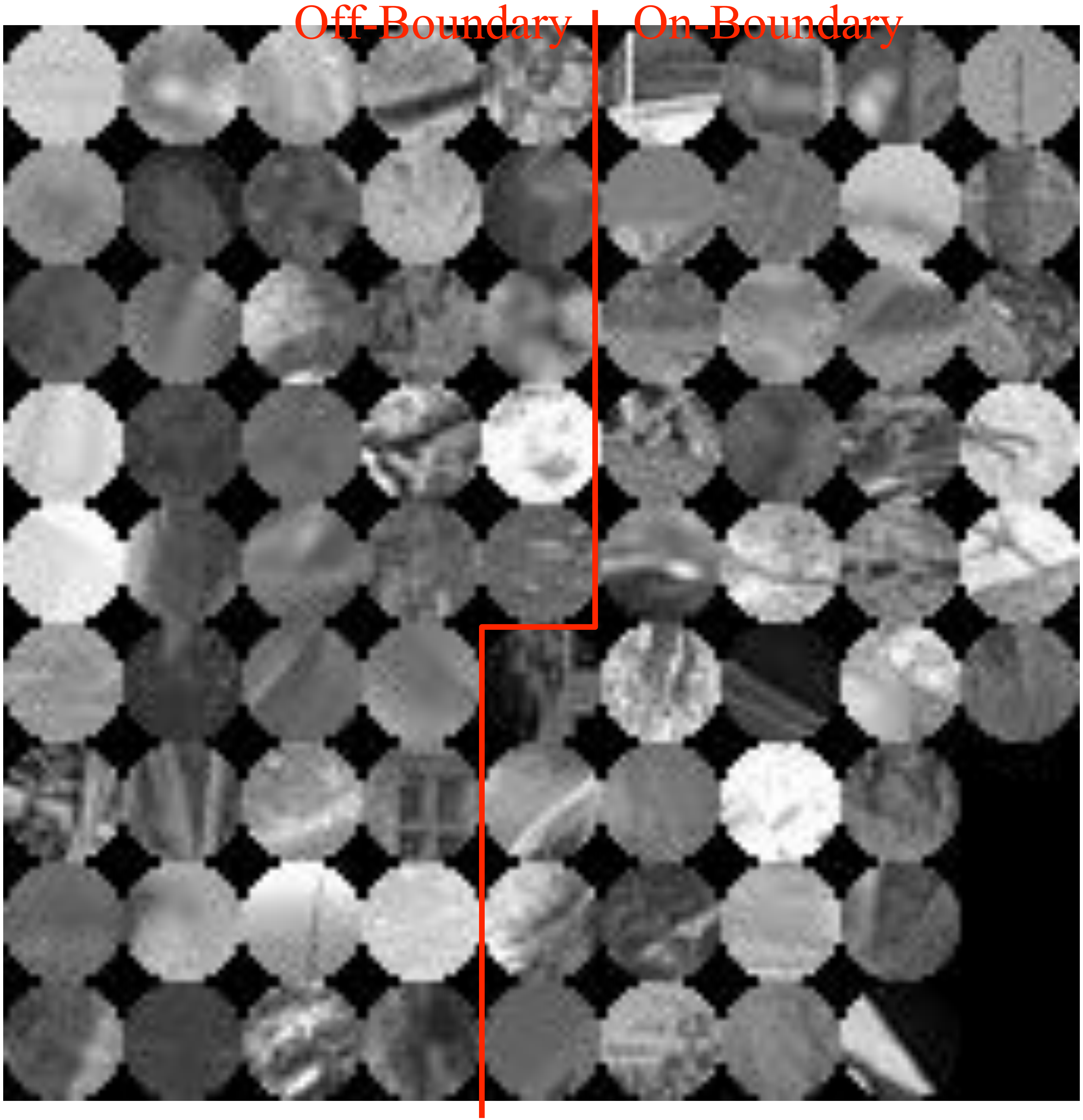
Simple Patches



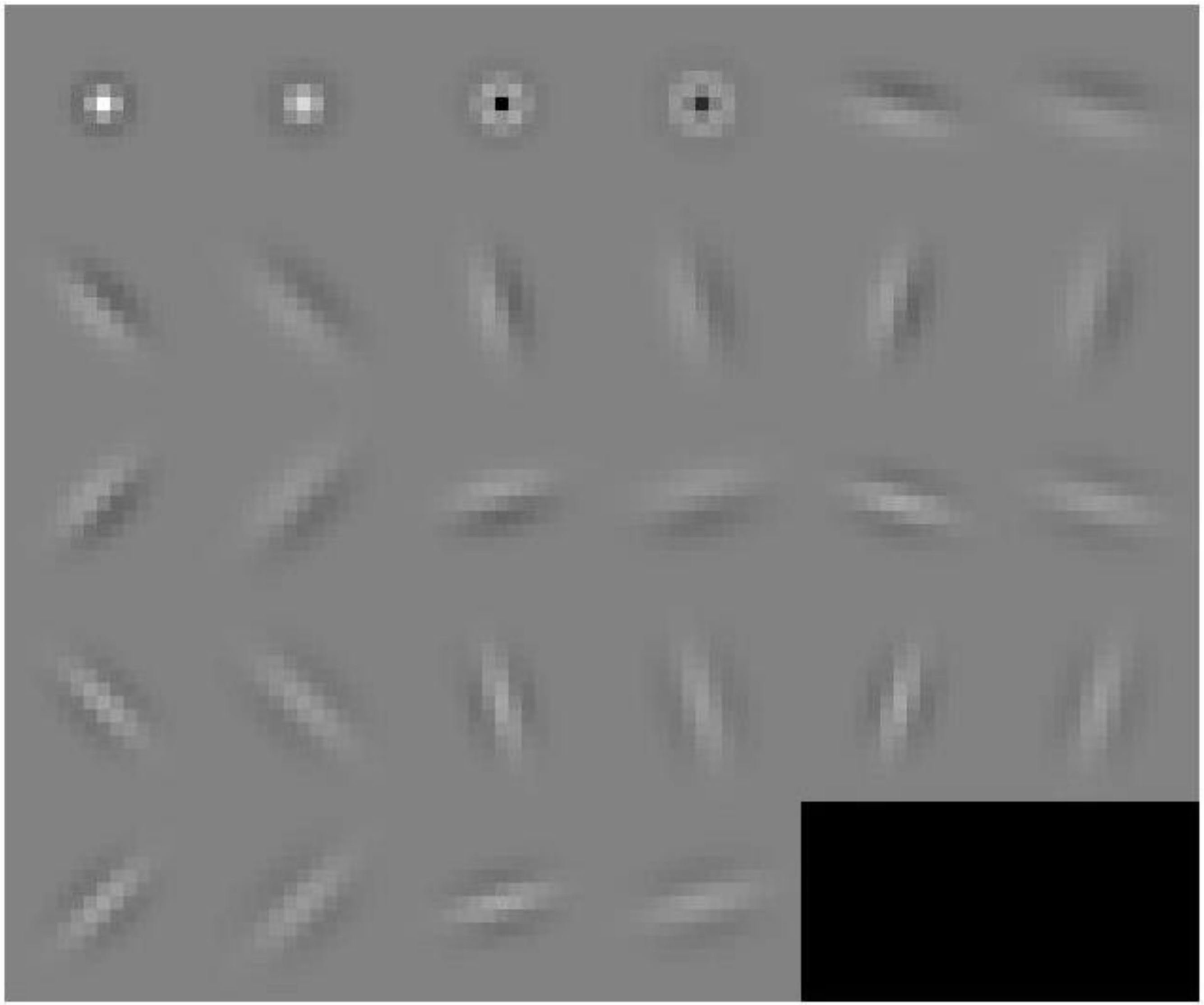
Textured Patches



Complex Patches

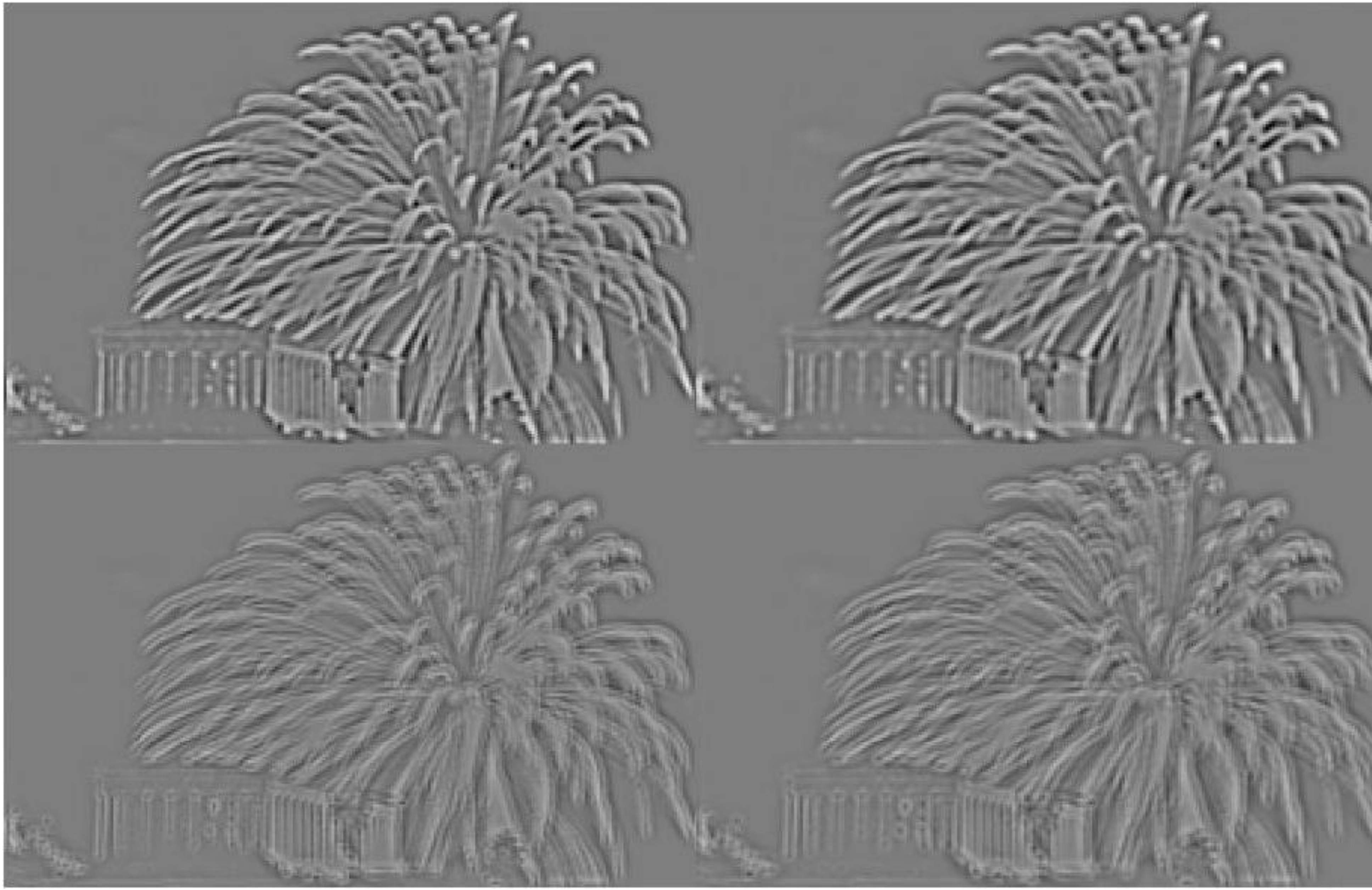


Filter Banks

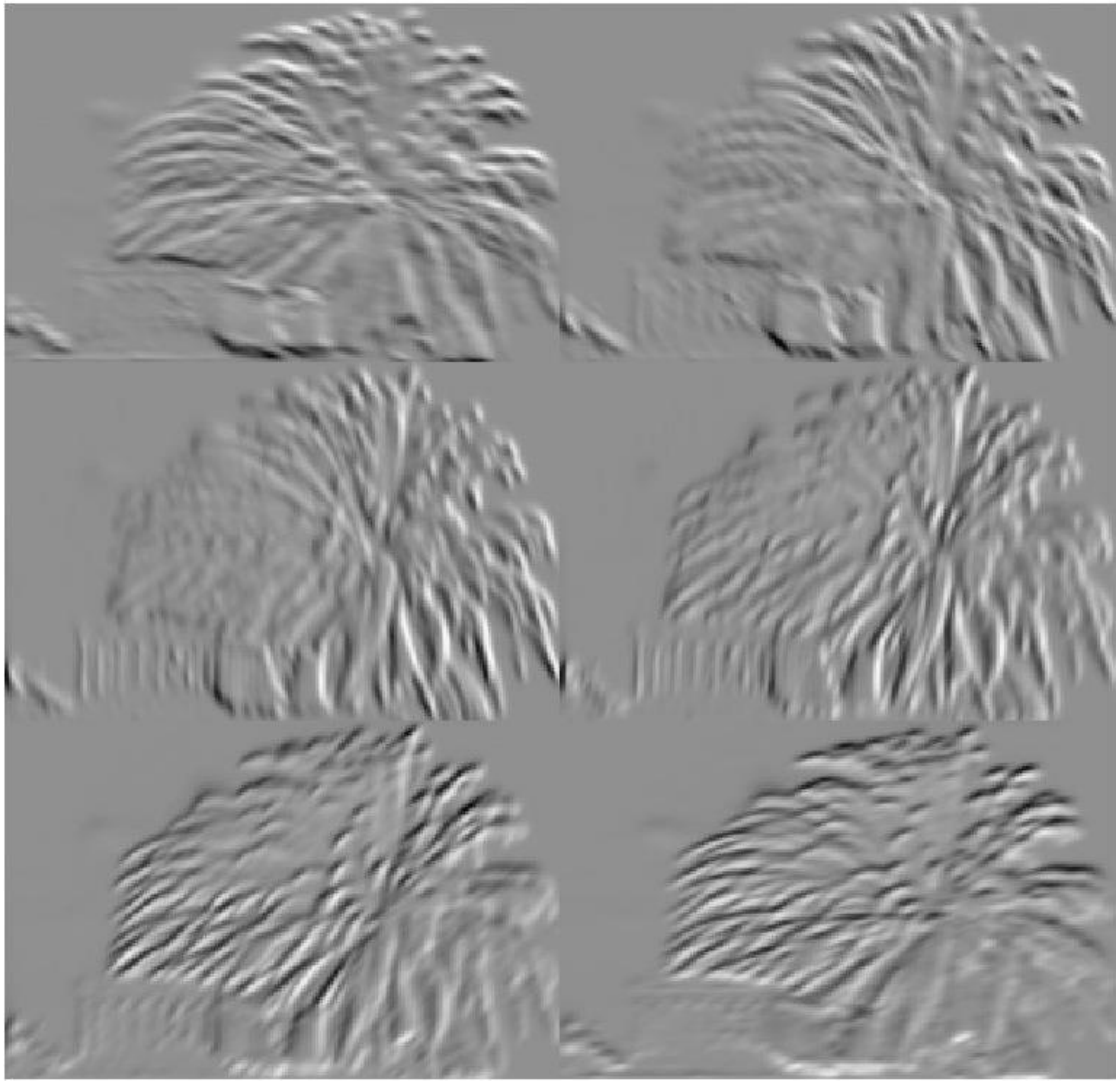


Filter Bank

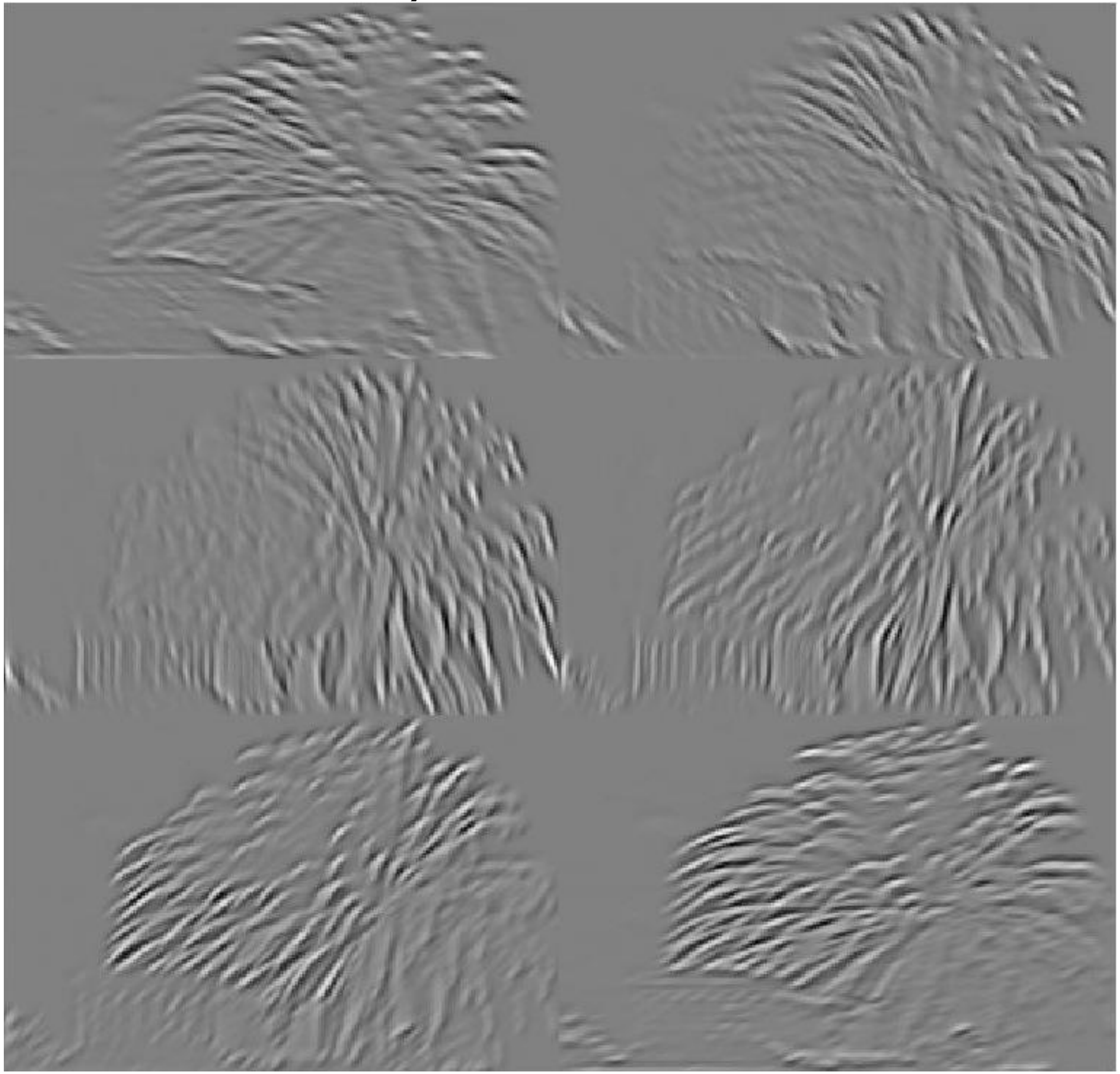
Dot filter response

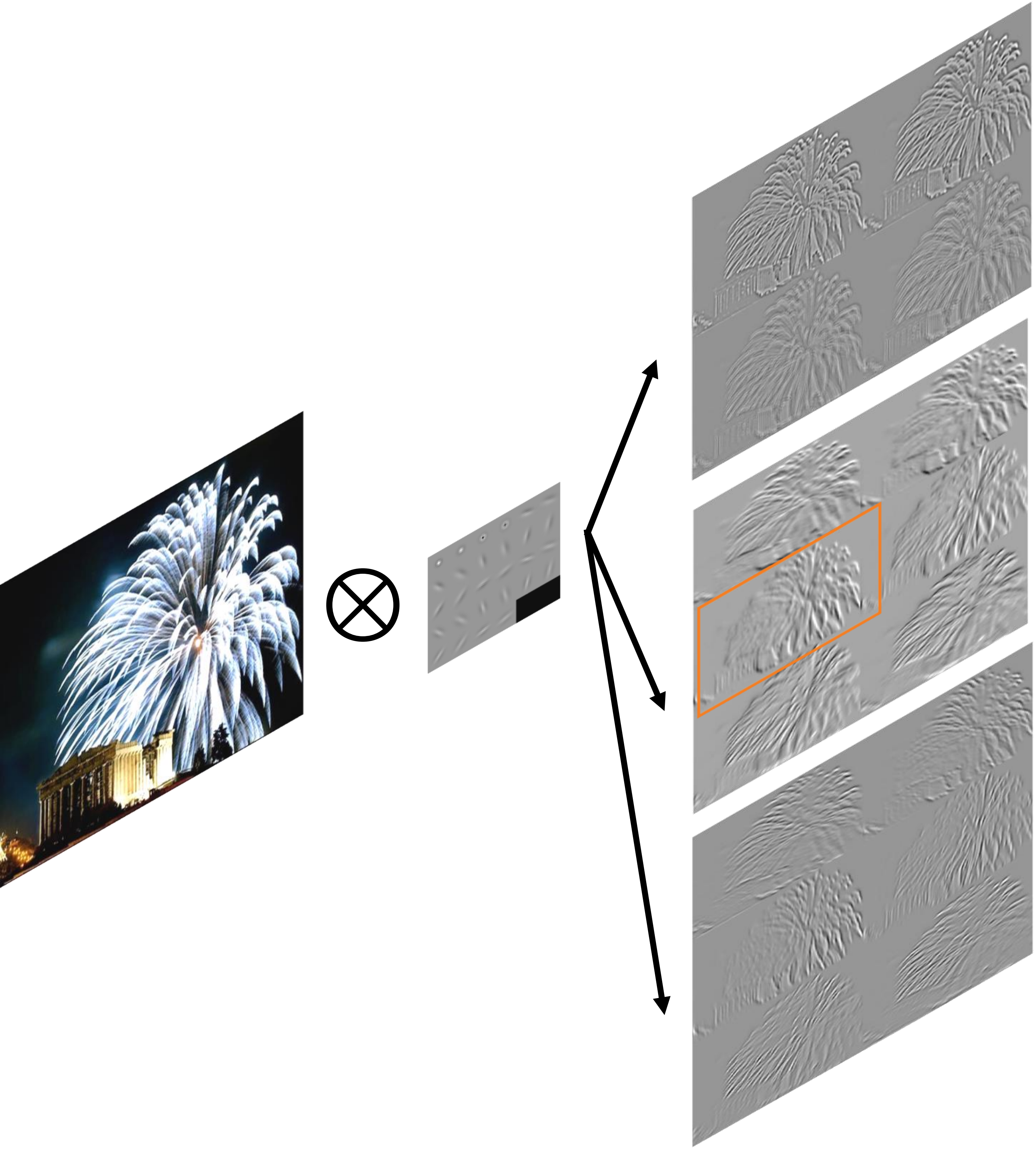


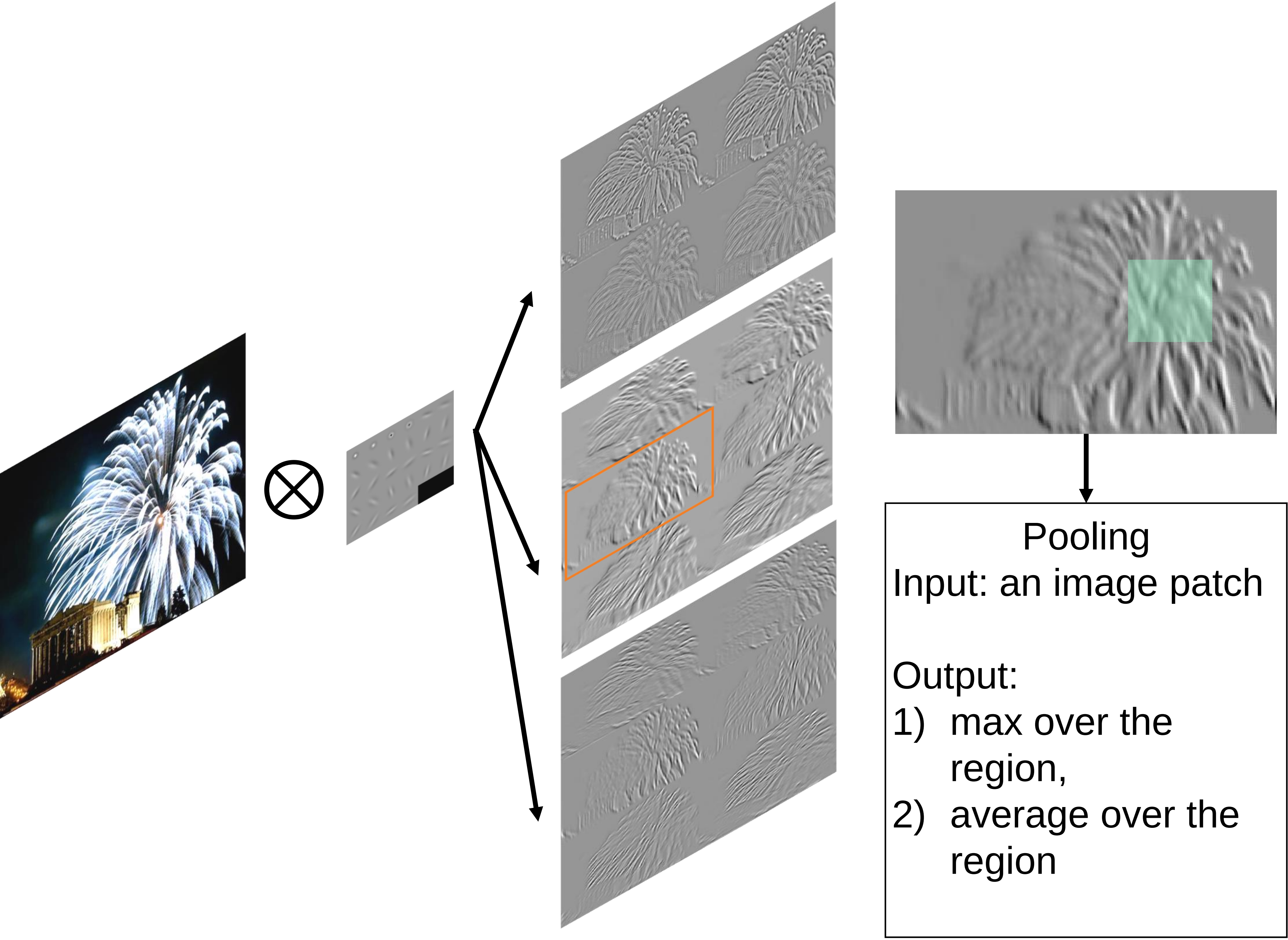
Odd symmetric fitler outputs

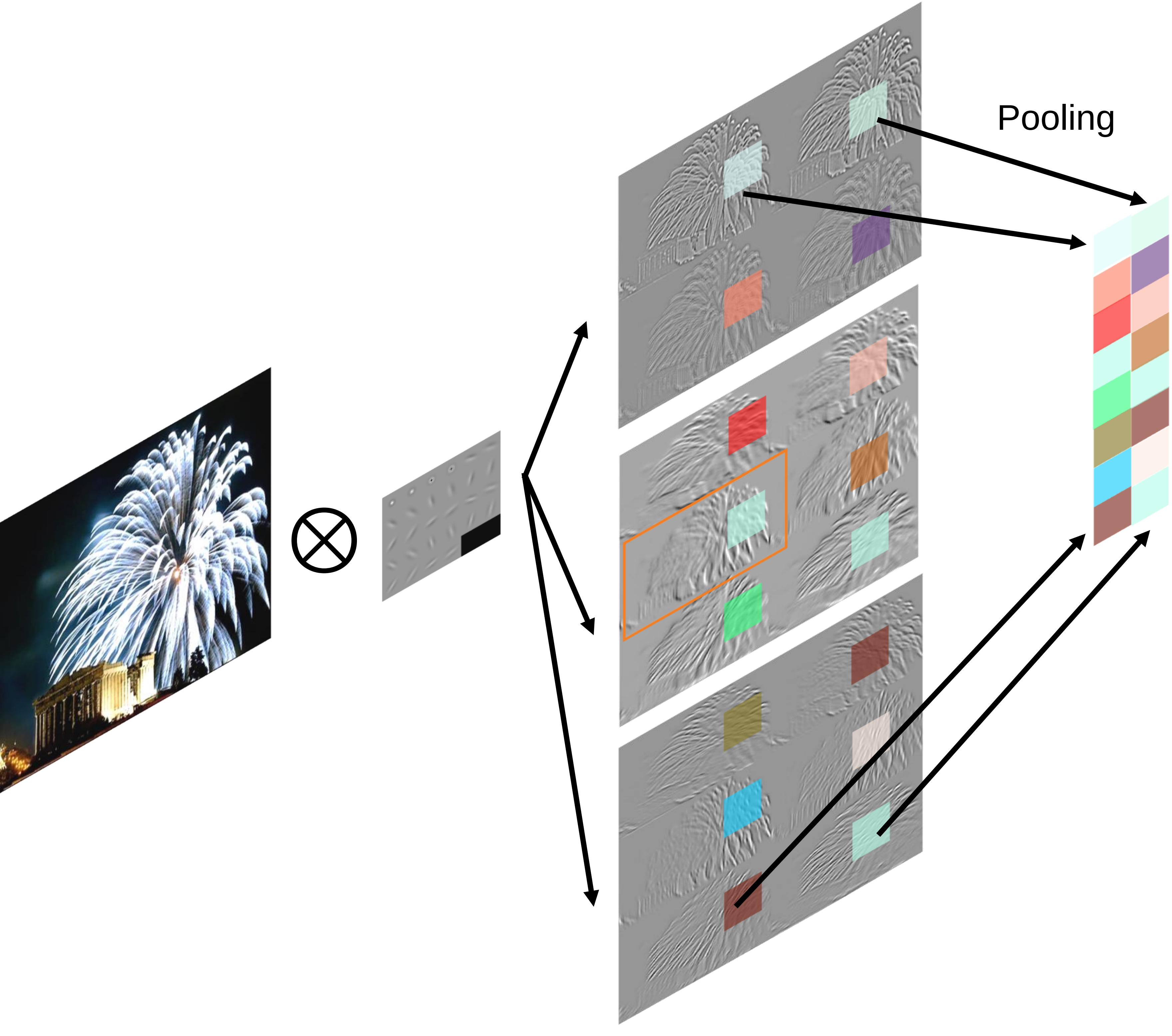


Even symmetric filter

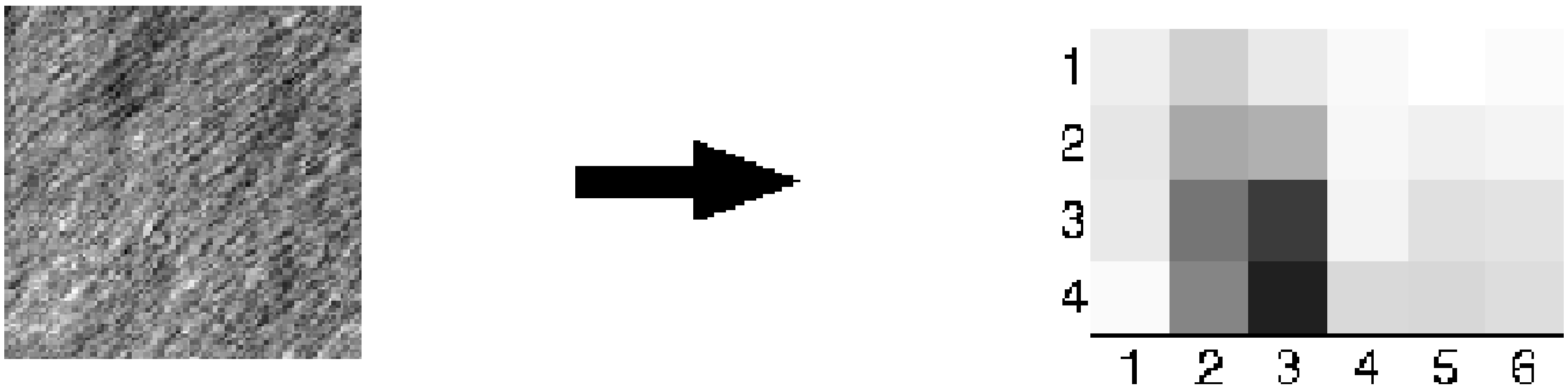


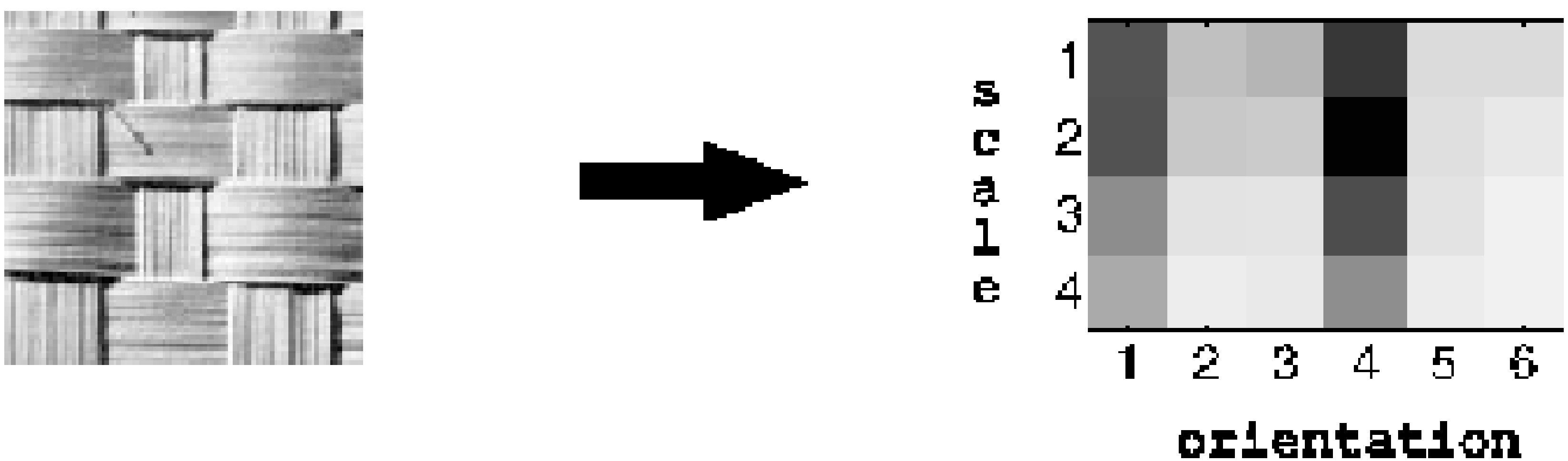


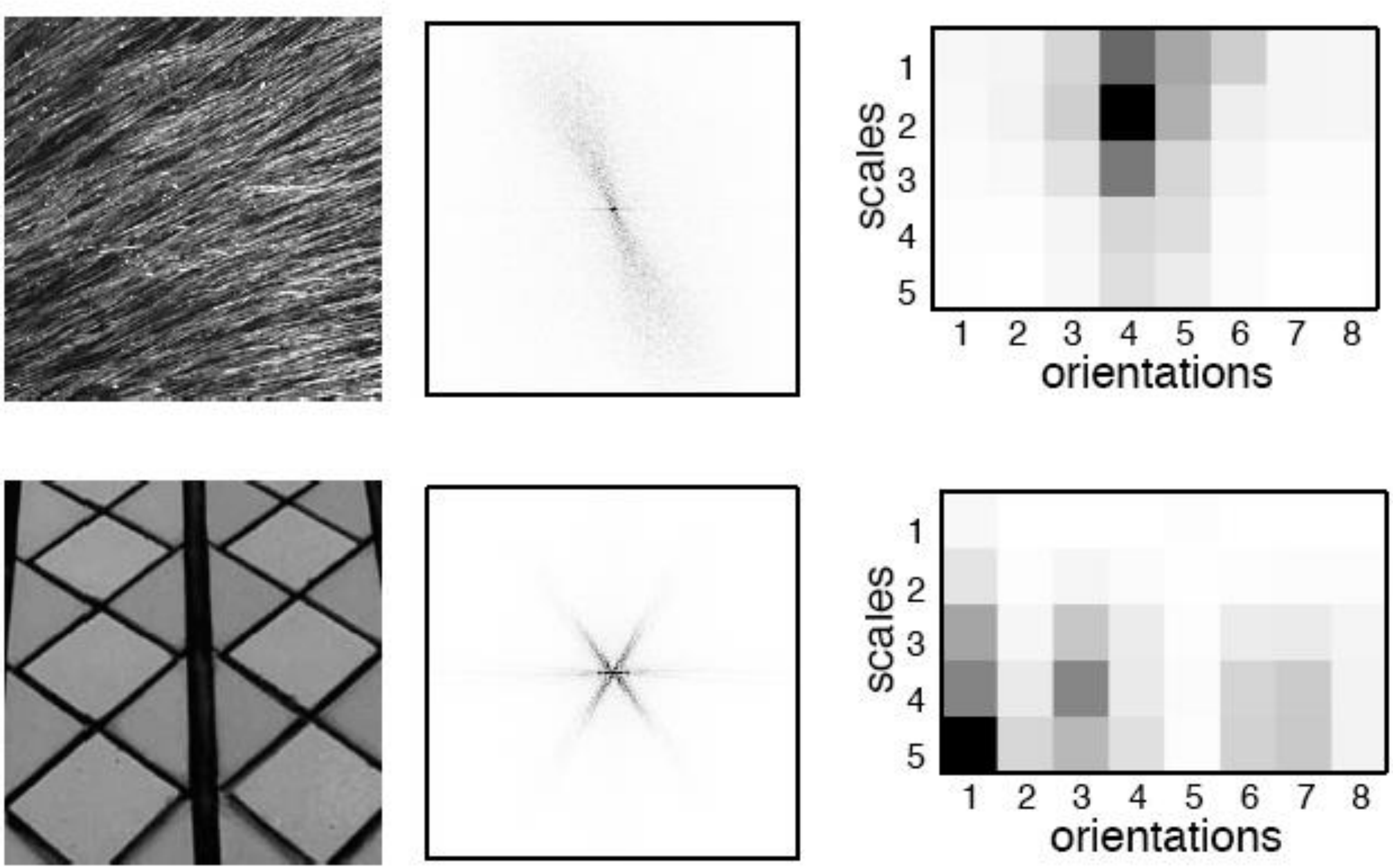

Pooling
Input: an image patch
Output:
1) max over the region,
2) average over the region


Pooling

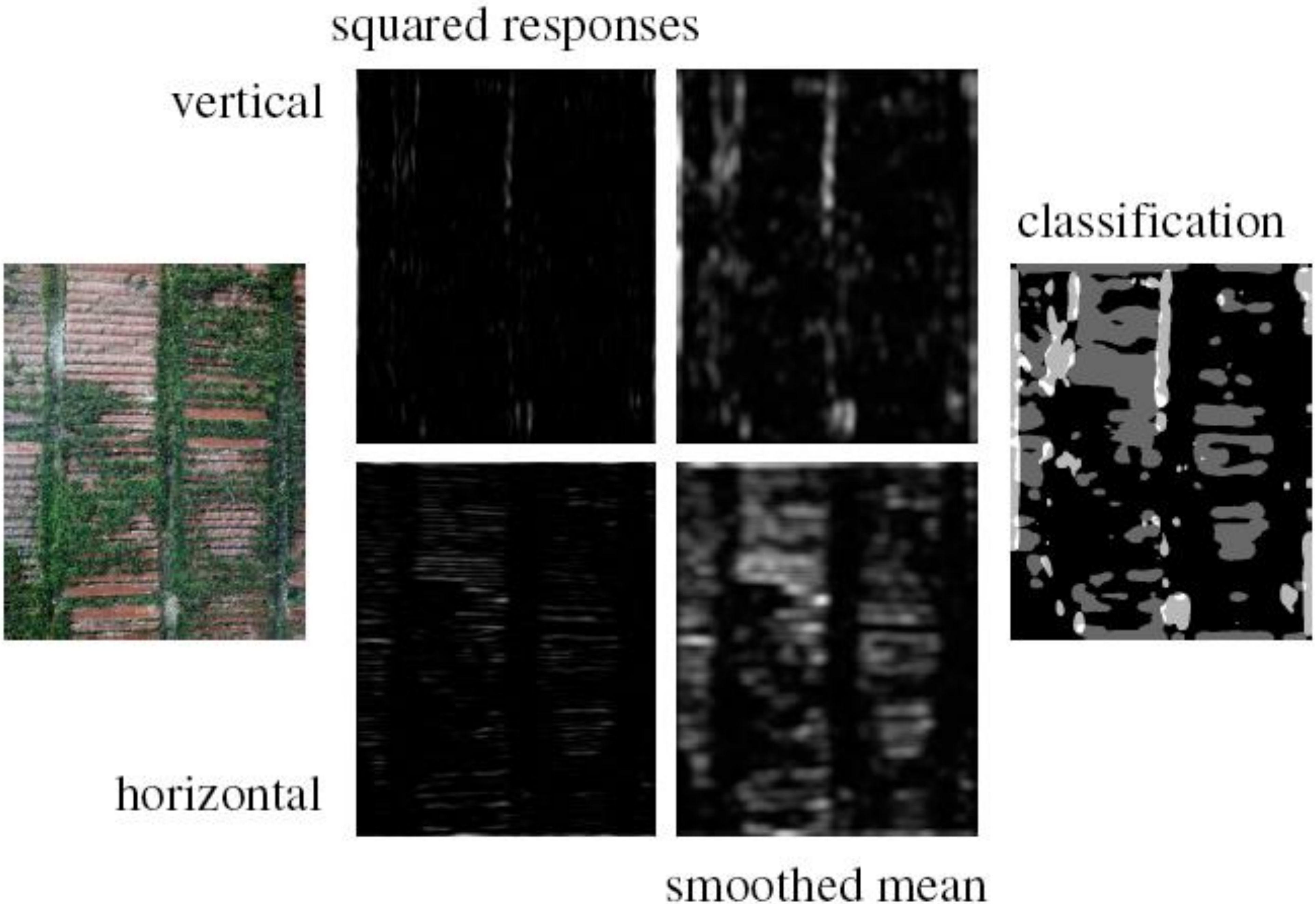
Pooling using ave. filter bank response

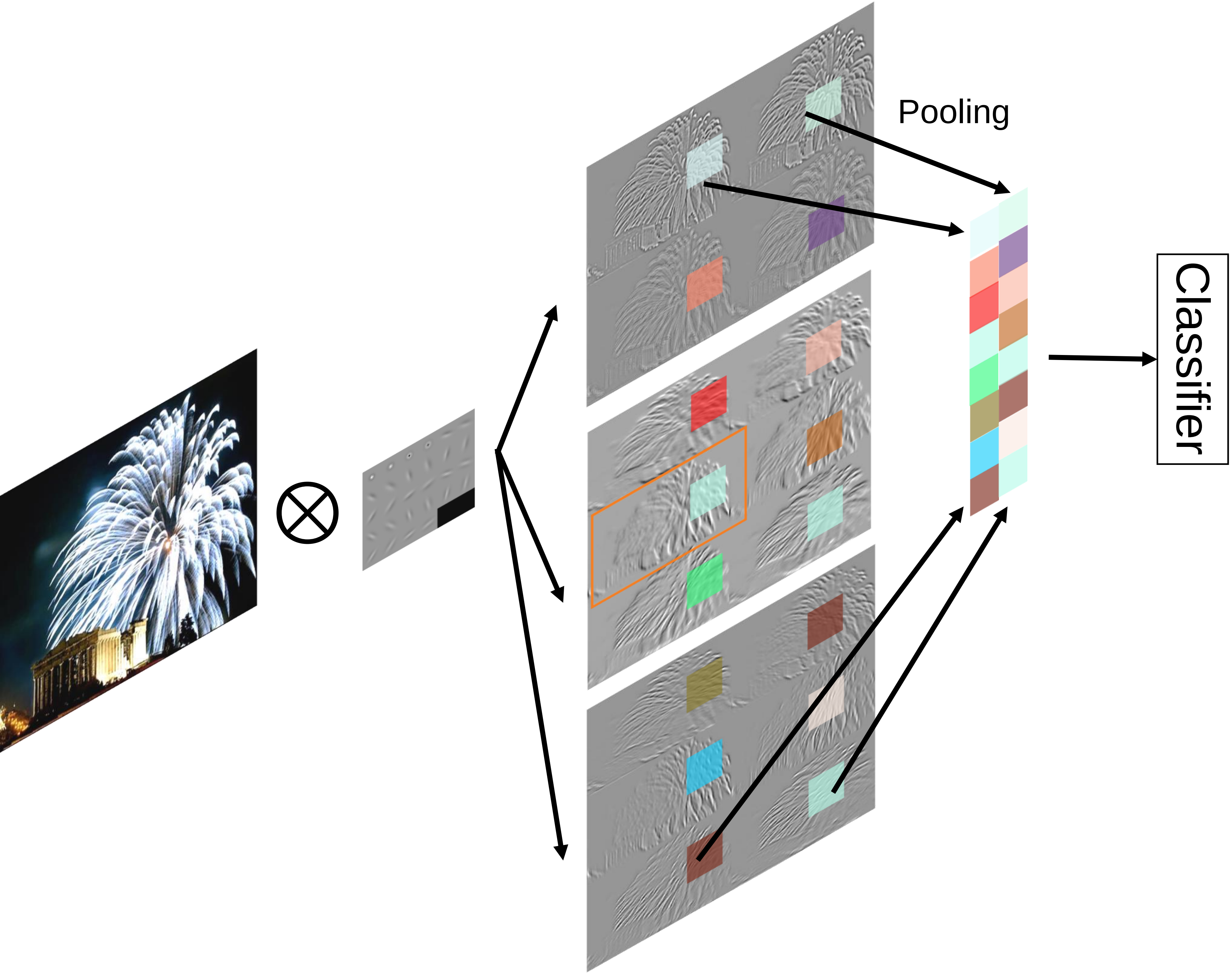




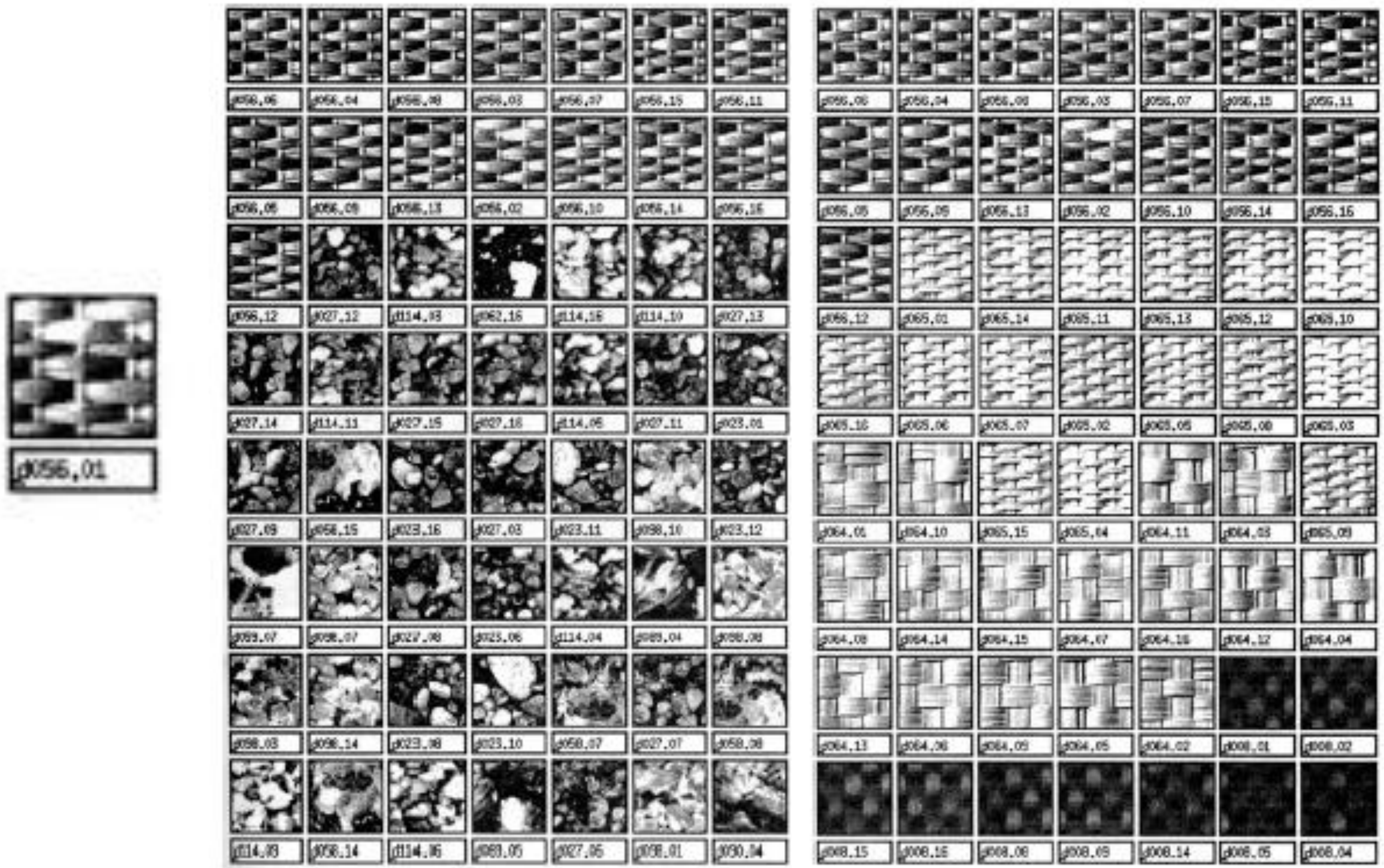

scales
1
2
3
4
5
1 2 3 4 5 6 7 8
orientations
scales
1
2
3
4
5
1 2 3 4 5 6 7 8
orientations

Average filter bank response




Pooling
Classifier

Is mean of filter outputs sufficent?

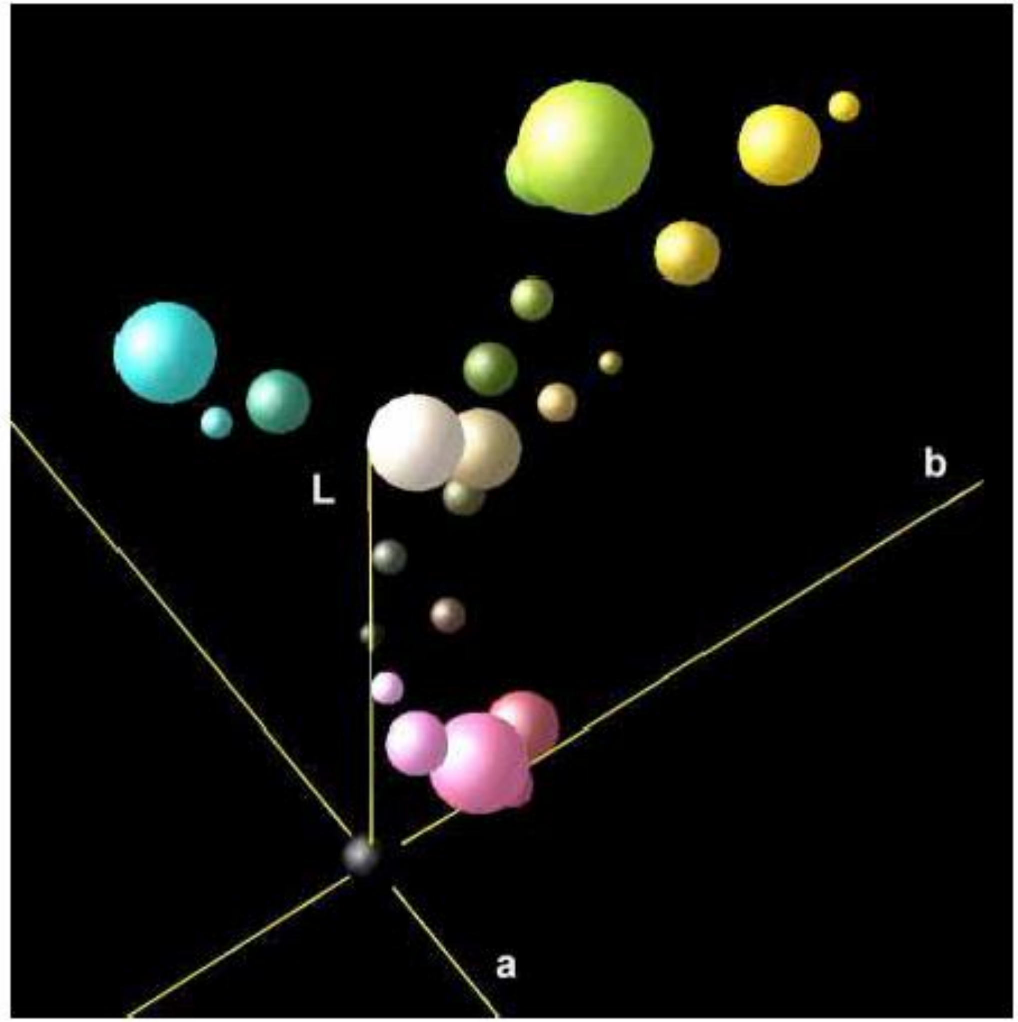


Histogram of filter banks

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

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

The vector **i** represents the bins in the relevant region of underlying feature space, defined by I(x)

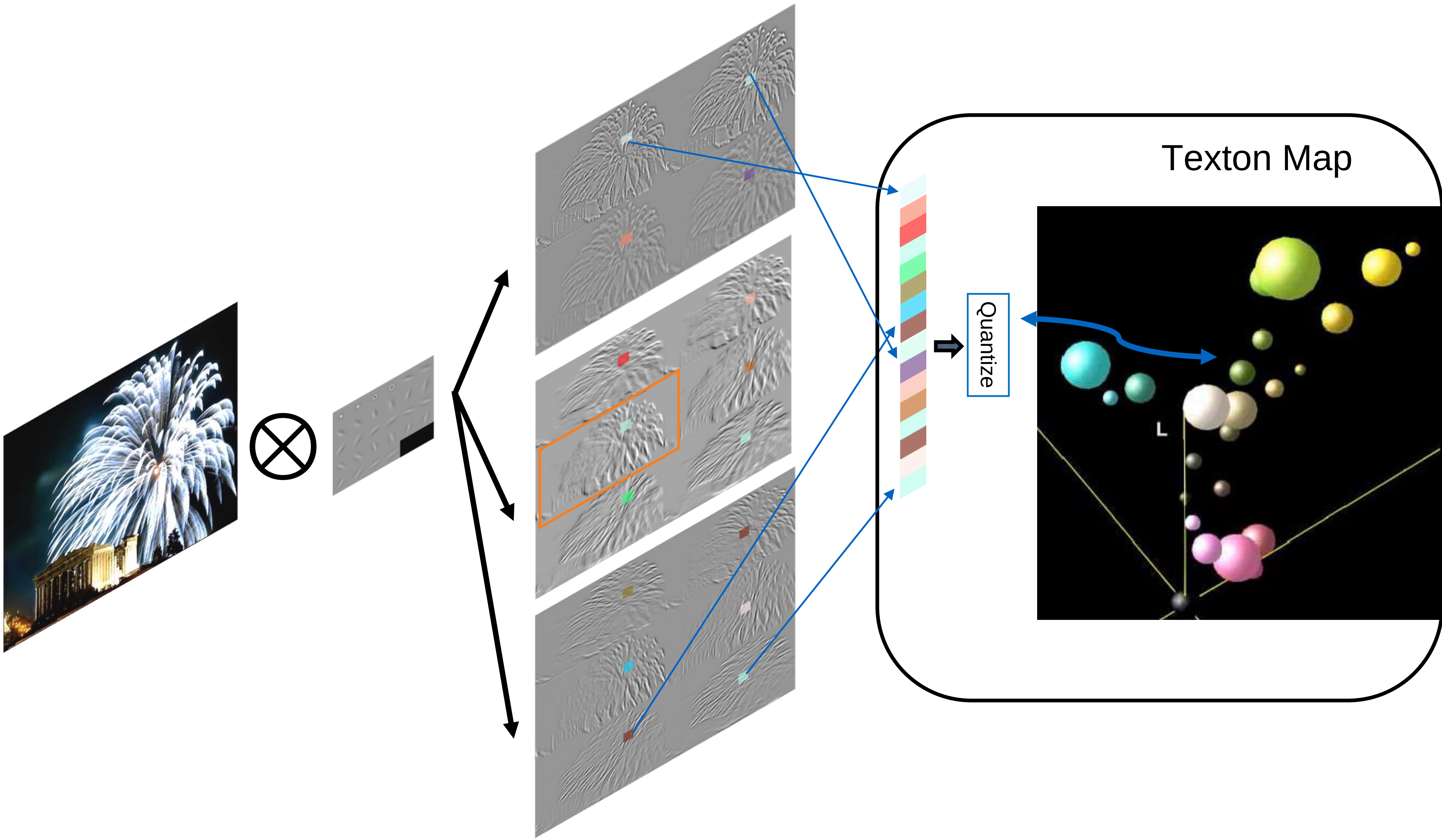


Adaptive binning: location of bins depends on the data itself,
Centers: are defined as prototypes {ci} and
Bins: are defined as the corresponding Voronoi tesslation.

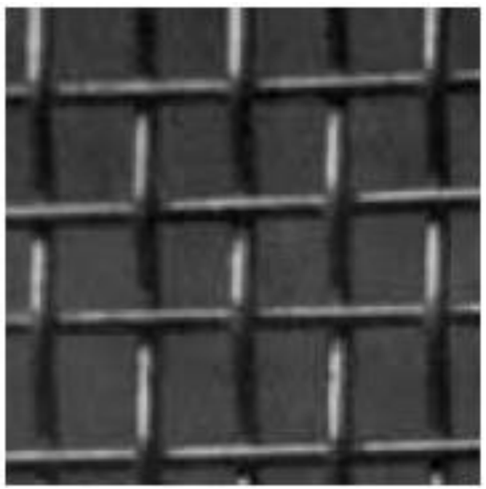
$$h_i = \left| \{ \mathbf{x} \; : \; i = \arg\min_j \| I(\mathbf{x}) - \mathbf{c}_j \| \} \right| .$$

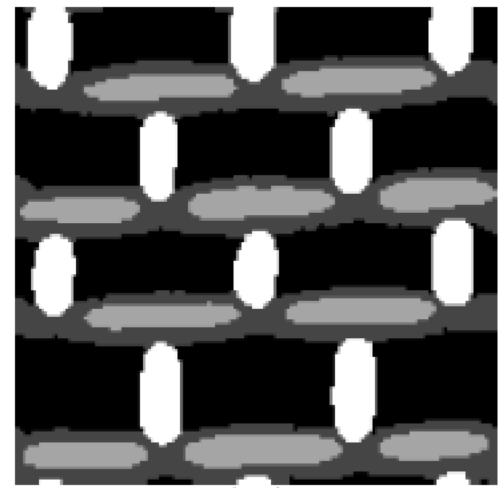
(a) (b) (c)

- 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 Map
Quantize
L

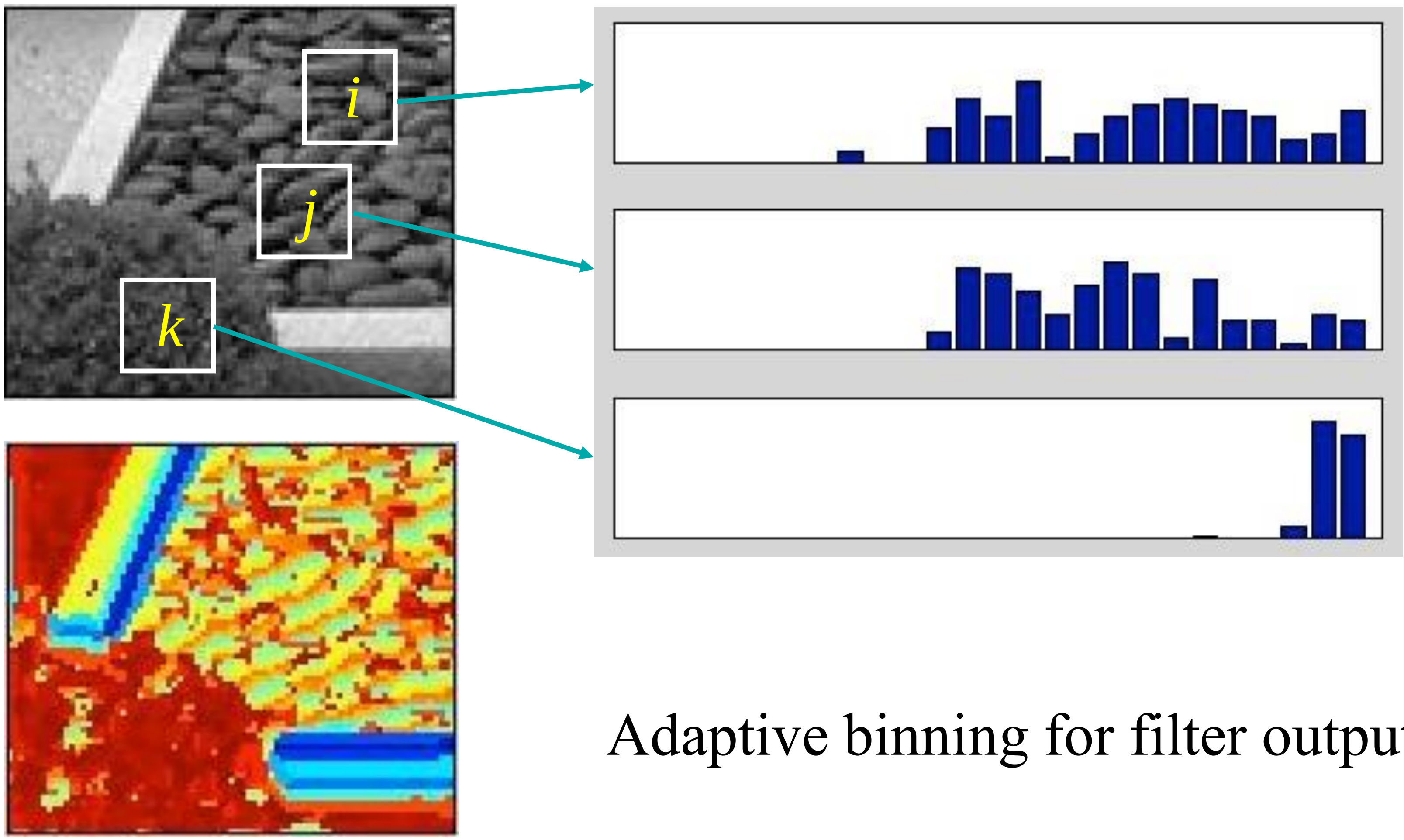
Texton: assign a label to each pixel





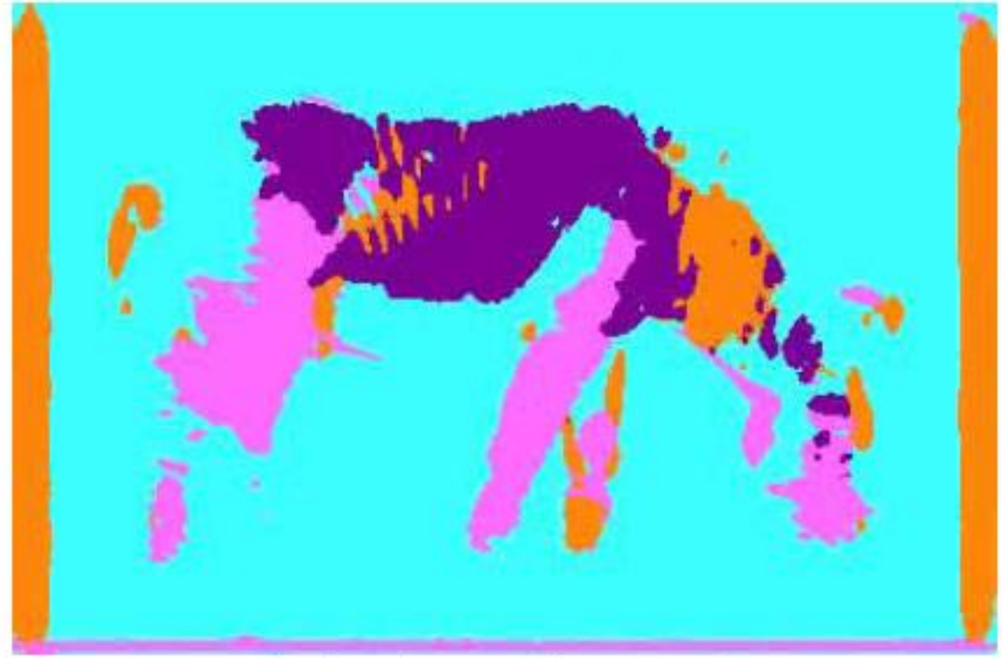
Pooling over texton

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

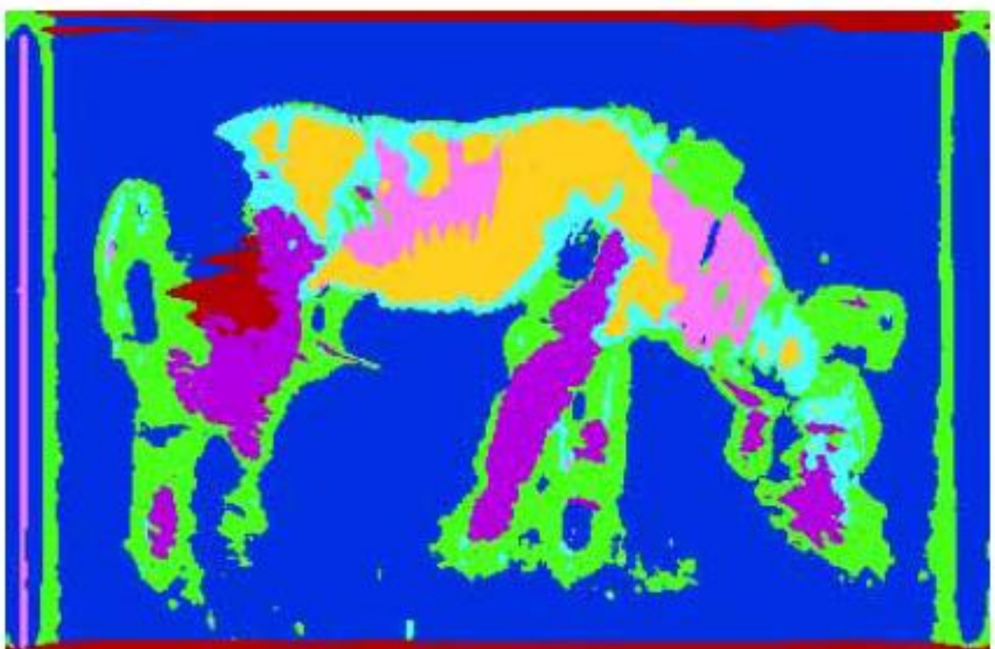


Adaptive binning for filter outputs

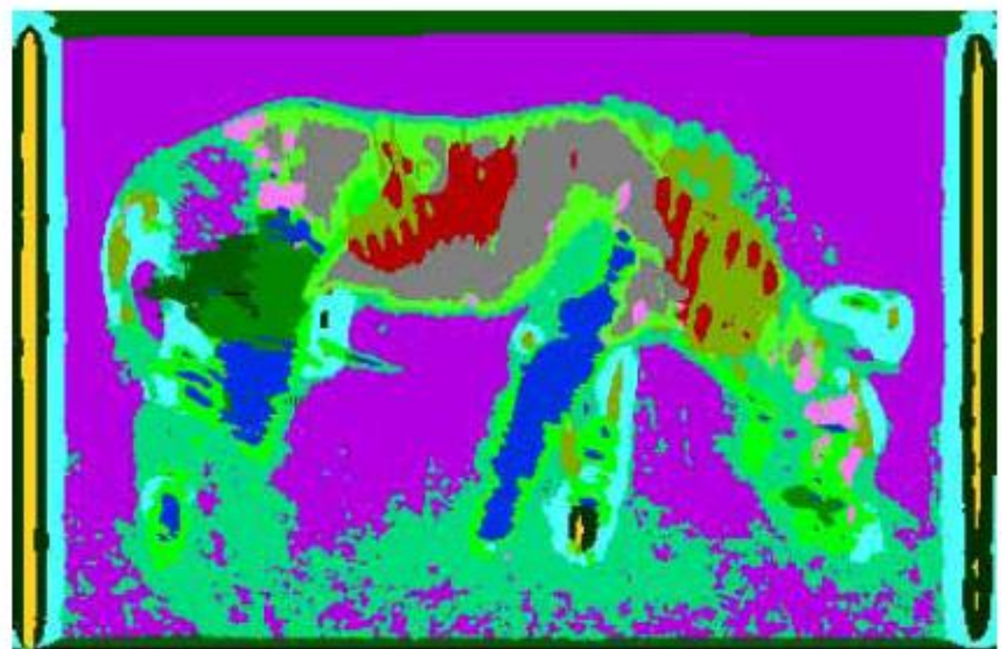
Zebra image



4-cluster assignment

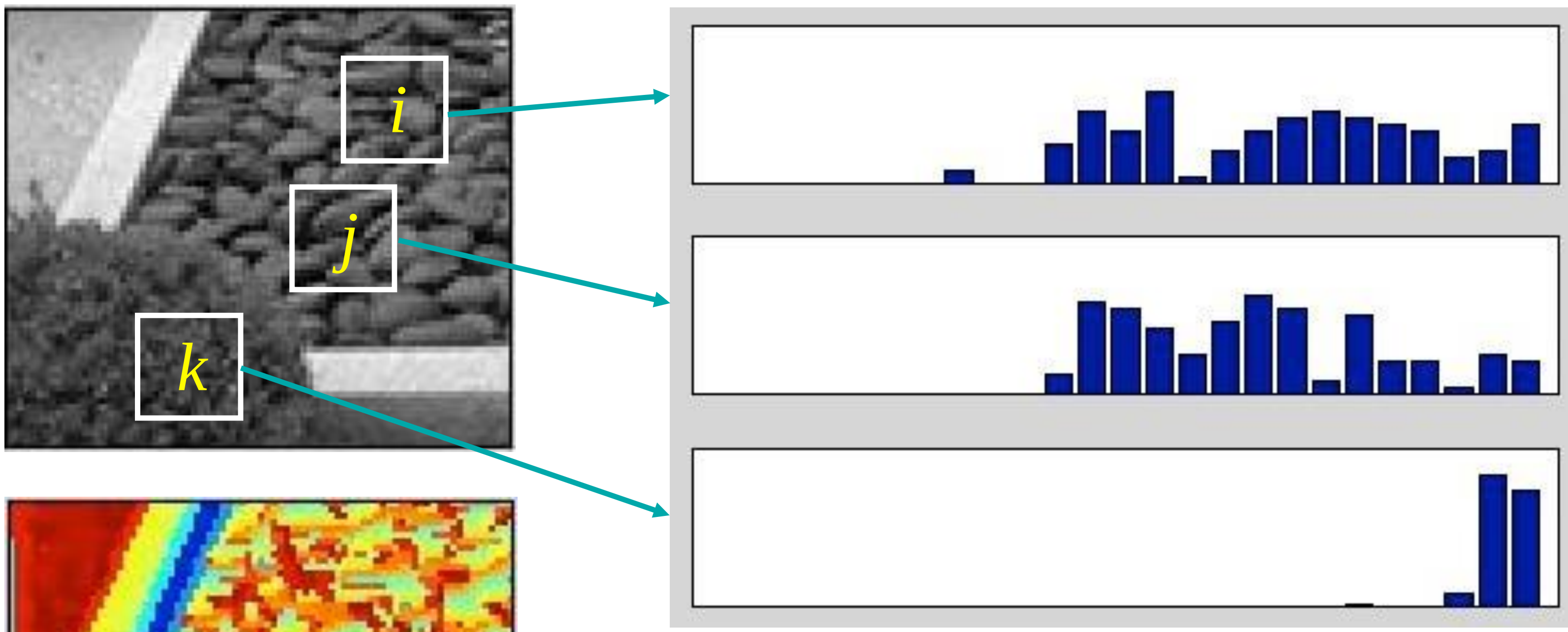


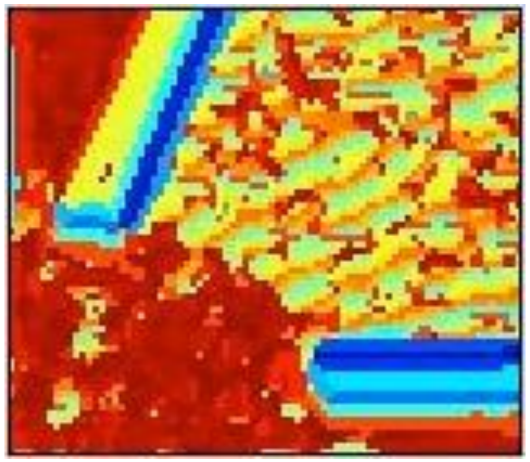
8-cluster assignment



16-cluster assignment

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

- $\rho(x, y) \geq 0$ (non-negativity)
- $\rho(x, y) = 0$ if and only if $x = y$ (identity)
- $\rho(x, y) = \rho(y, x)$ (symmetry)
- $\rho(x, z) \leq \rho(x, y) + \rho(y, z)$ (triangle inequality)

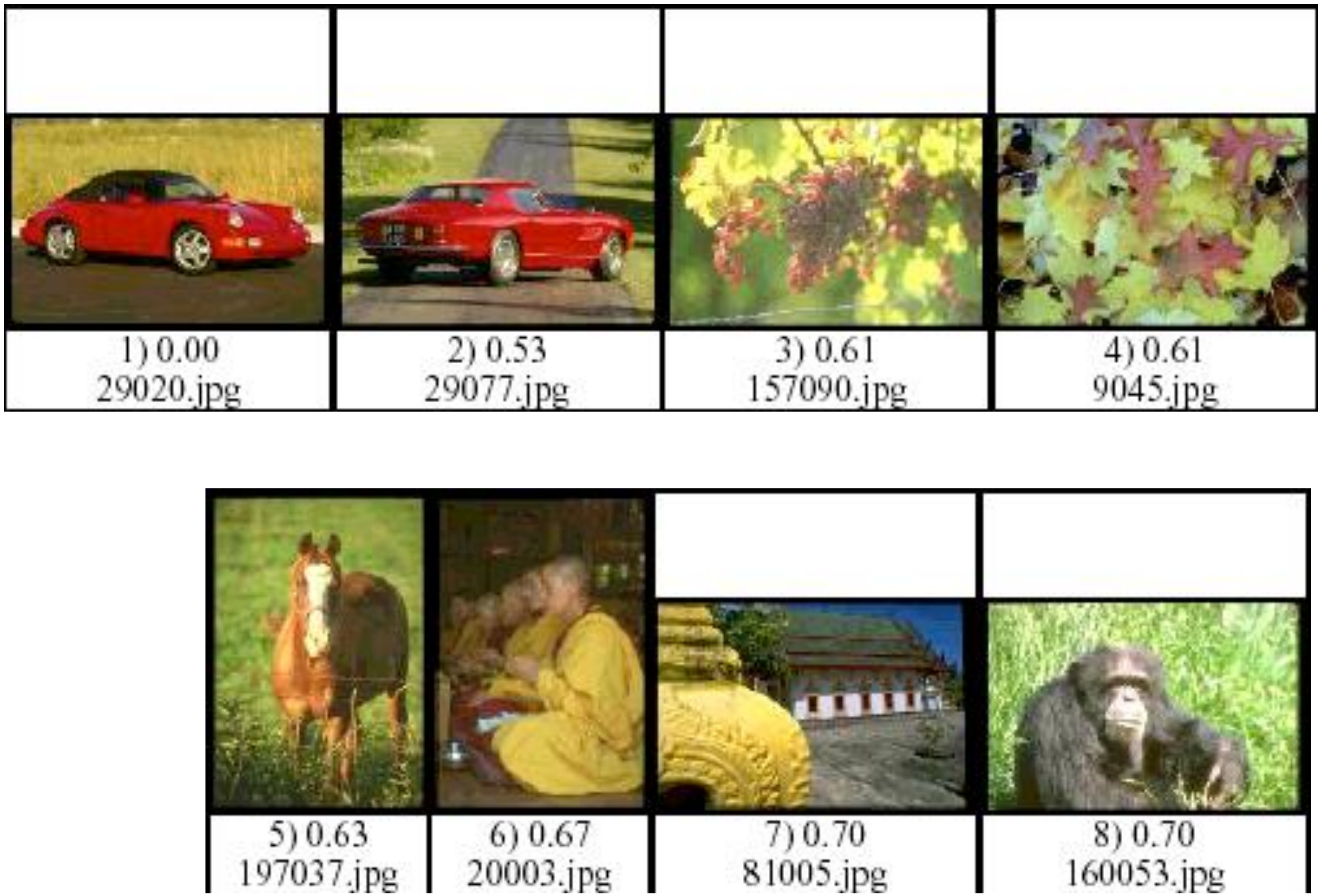
Heuristic Histogram distances

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

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

Bin-by-bin dissmilarity

Image similarity with L1 distance

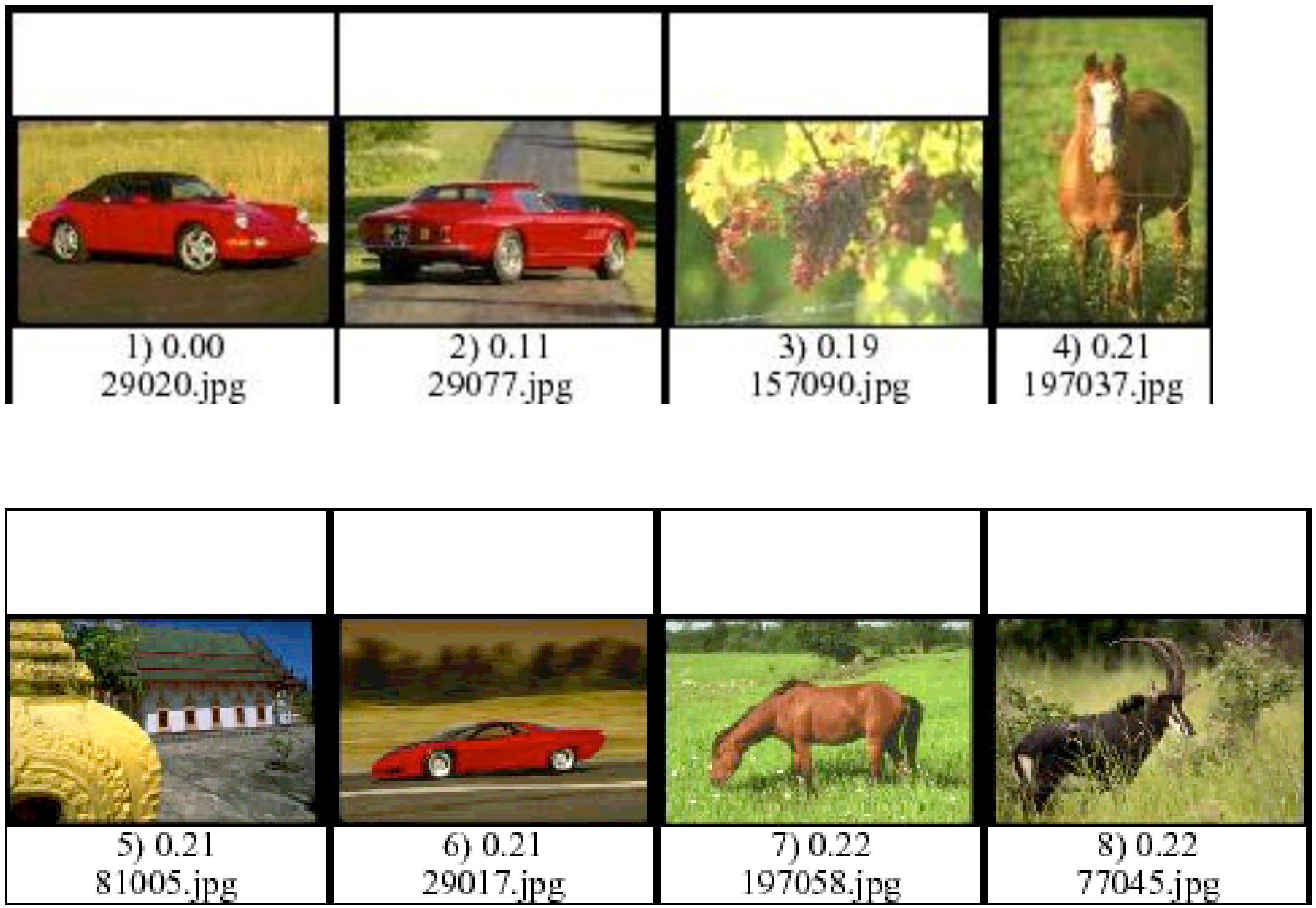


Non-parametric test statistics

The χ^2–*statistic* is given by

$$D(I,J) = \sum_i \frac{\left(f(i;I) - \hat{f}(i)\right)^2}{\hat{f}(i)},$$

Image similarity w. chi-sqr statistics



Information-theoretic divergences

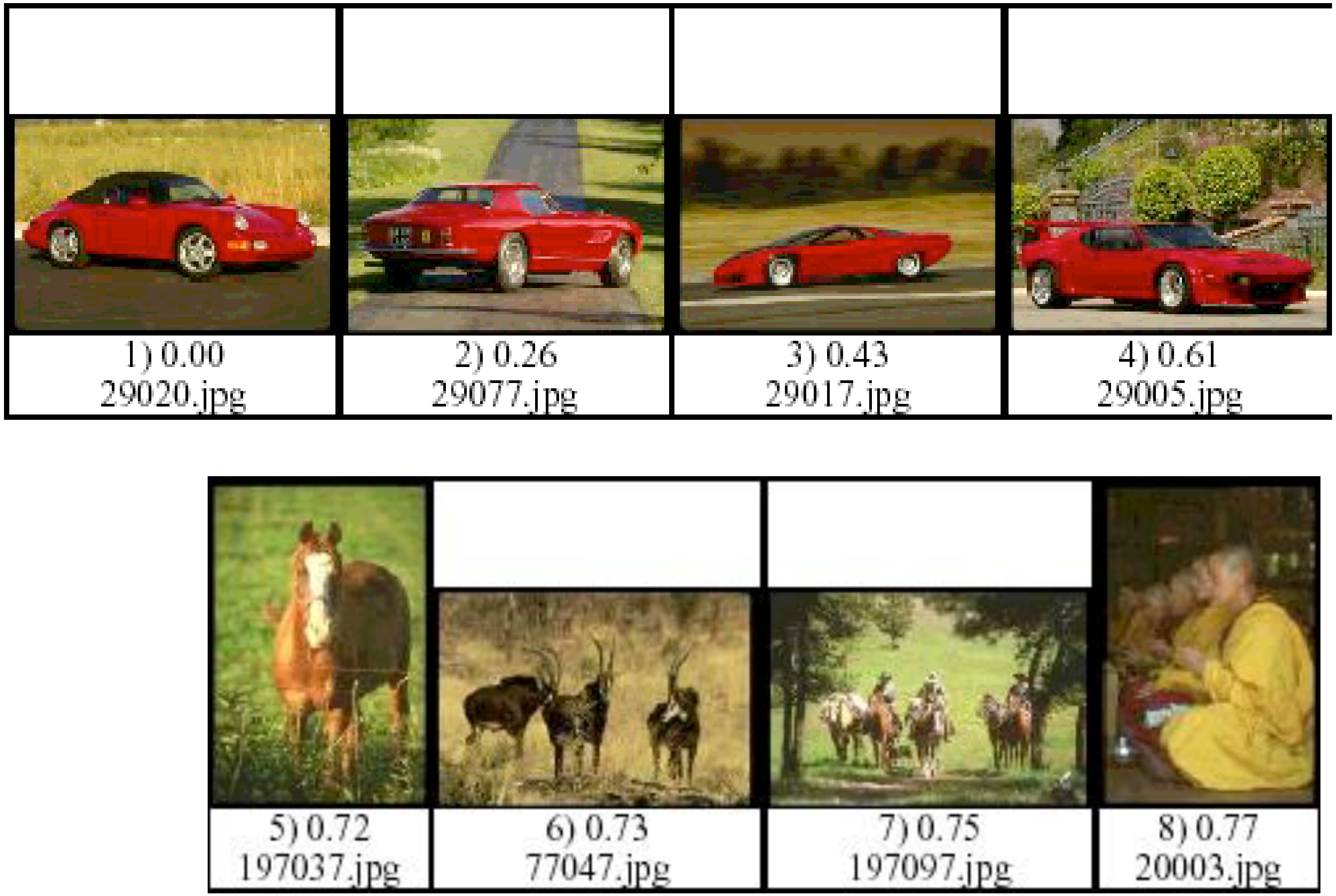
(i) The *Kullback–Leibler divergence* (KL) suggested in [10] as an image dissimilarity measure is defined by

$$D(I,J) = \sum_i f(i;I) \log \frac{f(i;I)}{f(i;J)} \quad . \qquad (9)$$

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

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

Image similarity with Jeffrey divergence



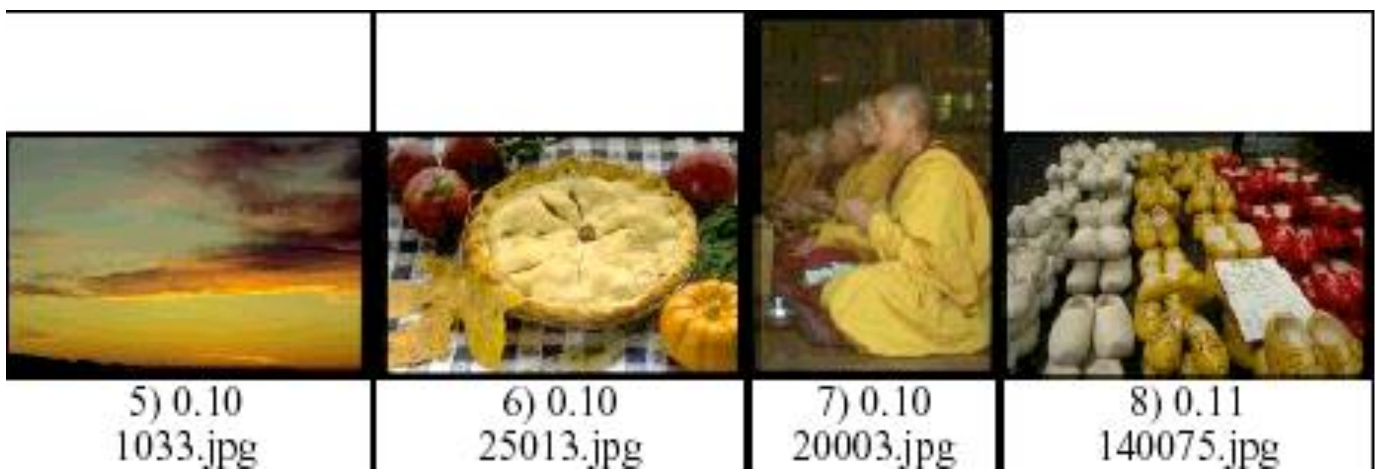
Perceptual similarity

- Quadratic form

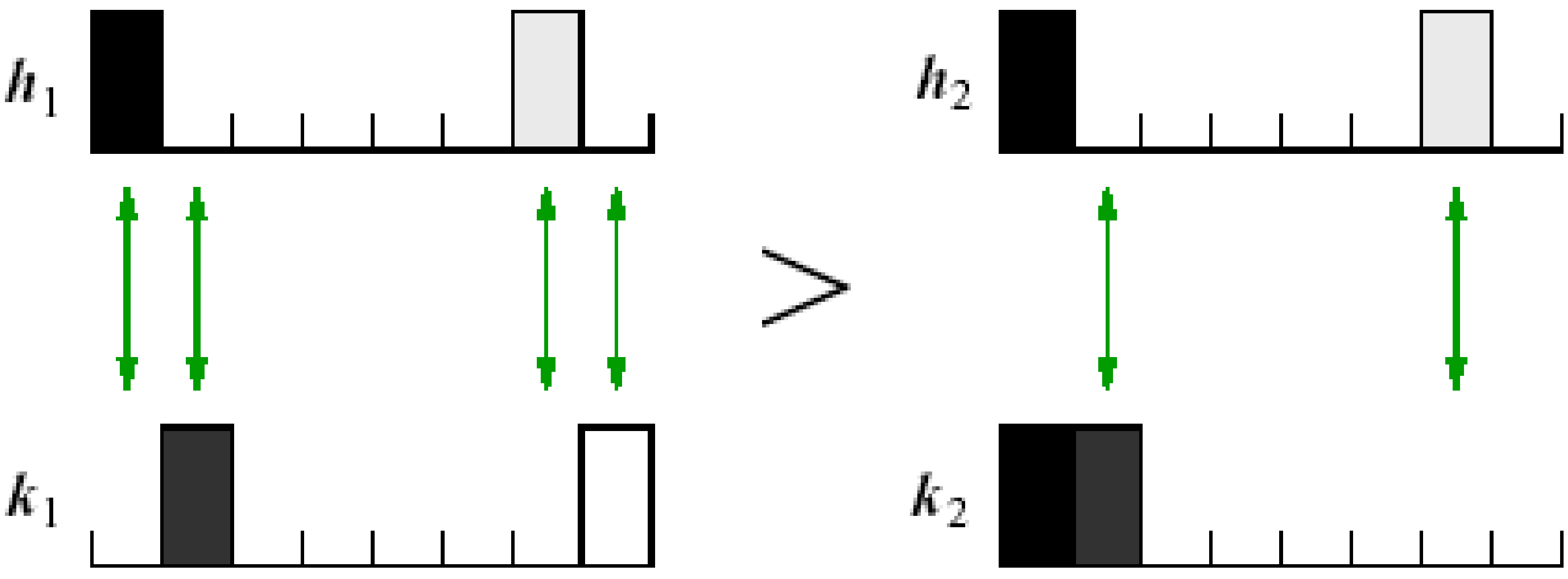
$$D(I, J) = \sqrt{(\vec{f}_I - \vec{f}_J)^T \mathbf{A} (\vec{f}_I - \vec{f}_J)} \; ,$$

- Earth Moving Distance

Image similarity w. quadratic-form

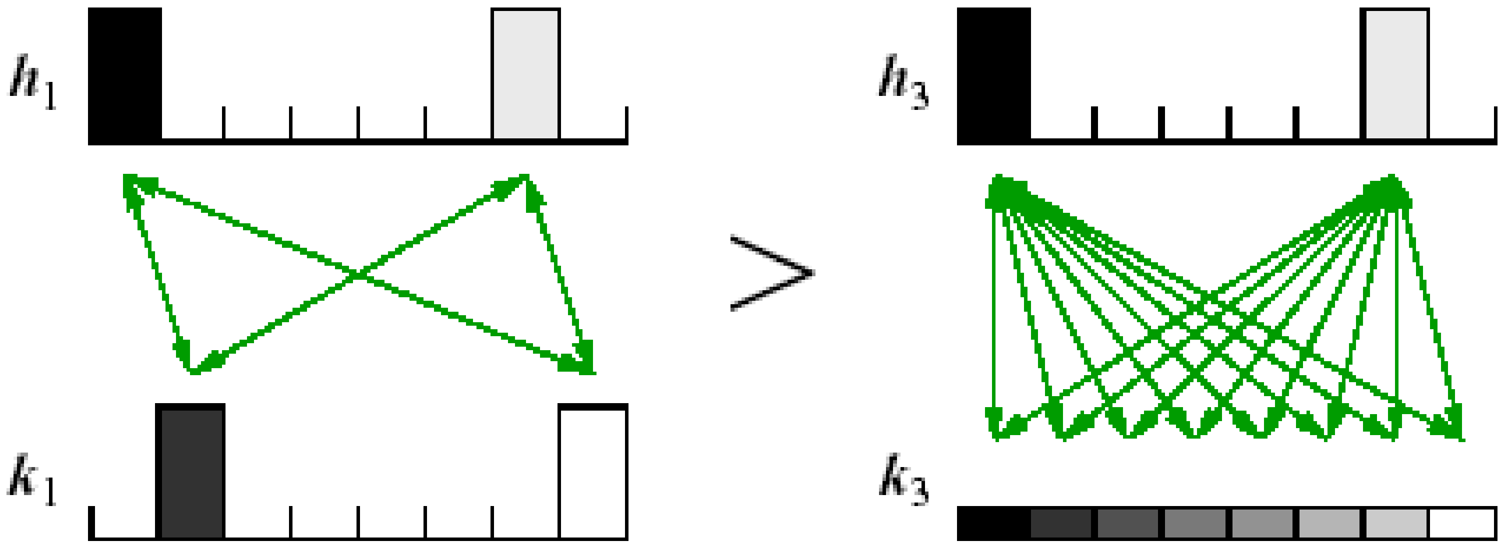


Problems with Binning



(a)

Problem with quadradic norm



(b)

Image similarity with Earth Moving Distance(EMD)

Earth Moving Distance

- Let P, Q to be 2 histogram signiture:
  - P={(P1,w_p1),…(Pm,w_pm)}
  - Q={(Q1:w_q1),…,(Qn,w_qn)}
- Find a optimal mapping from P to Q
- Define a flow F(i,j) so to minimize

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

Earth Moving Distance

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

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

- Such tnat,

$$\begin{aligned}
f_{ij} &\geq 0 & & 1 \leq i \leq m,\ 1 \leq j \leq n \\
\sum_{j=1}^{n} f_{ij} &\leq w_{p_i} & & 1 \leq i \leq m \\
\sum_{i=1}^{m} f_{ij} &\leq w_{q_j} & & 1 \leq j \leq n \\
\sum_{i=1}^{m}\sum_{j=1}^{n} f_{ij} &= \min\left(\sum_{i=1}^{m} w_{p_i}, \sum_{j=1}^{n} w_{q_j}\right),
\end{aligned}$$

$$\begin{array}{|ccc|c|} \hline f_{11} & \cdots & f_{1n} & w_{\mathbf{p}_1} \\ \vdots & & \vdots & \vdots \\ f_{m1} & \cdots & f_{mn} & w_{\mathbf{p}_m} \\ \hline w_{\mathbf{q}_1} & \cdots & w_{\mathbf{q}_n} & \\ \hline \end{array} .$$

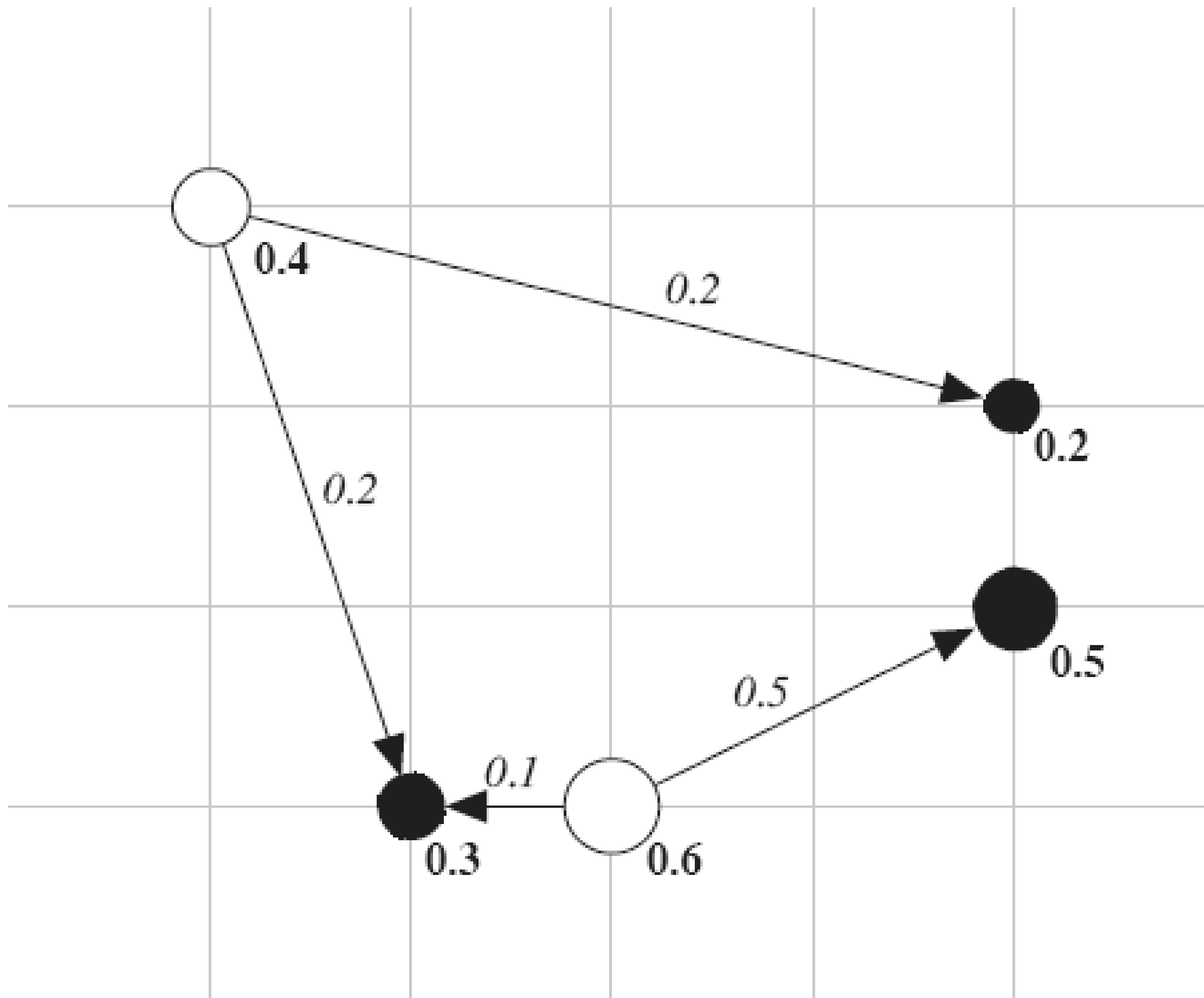
Earth Moving Distance

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

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

- Earth Moving Distance is,

$$\mathrm{EMD}(P, Q) = \frac{\sum_{i=1}^{m} \sum_{j=1}^{n} d_{ij} f_{ij}}{\sum_{i=1}^{m} \sum_{j=1}^{n} f_{ij}} ,$$

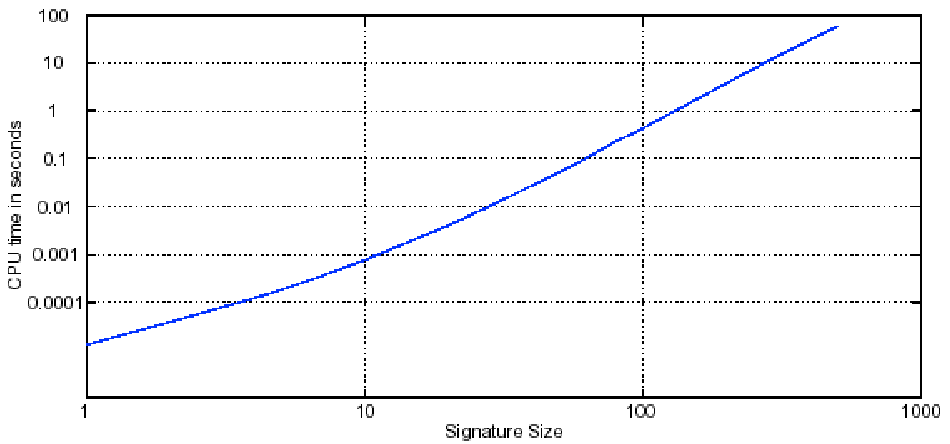

0.4
0.2
0.2
0.2
0.5
0.5
0.1
0.3
0.6

Theorem 3.1 *If two signatures, P and Q, have equal weights and the ground distance $d(\mathbf{p}_i, \mathbf{q}_j)$ is metric for all $\mathbf{p}_i$ in P and $\mathbf{q}_j$ 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 ($\mathrm{EMD}(P, Q) \geq 0$ and $\mathrm{EMD}(P, Q) = 0$ iff $P \equiv Q$), symmetry ($\mathrm{EMD}(P, Q) = \mathrm{EMD}(Q, P)$), and the triangle inequality (for any signature R, $\mathrm{EMD}(P, Q) \leq \mathrm{EMD}(P, R) + \mathrm{EMD}(R, Q)$).

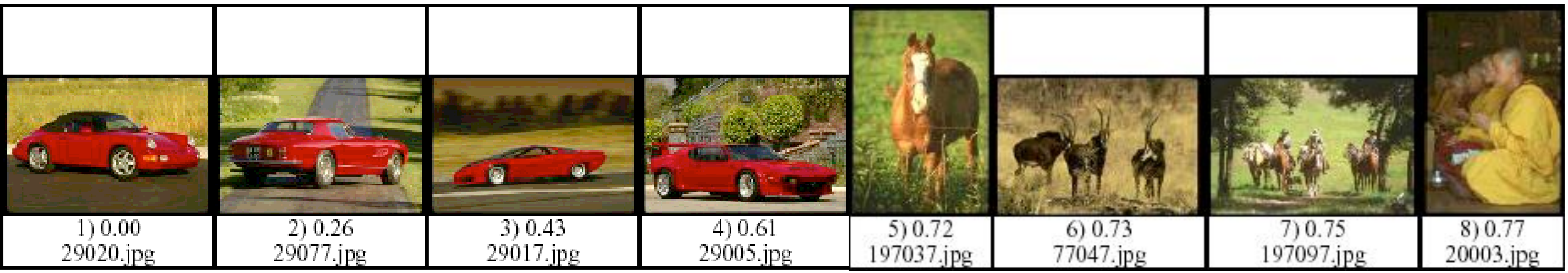
How to compute EDM

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

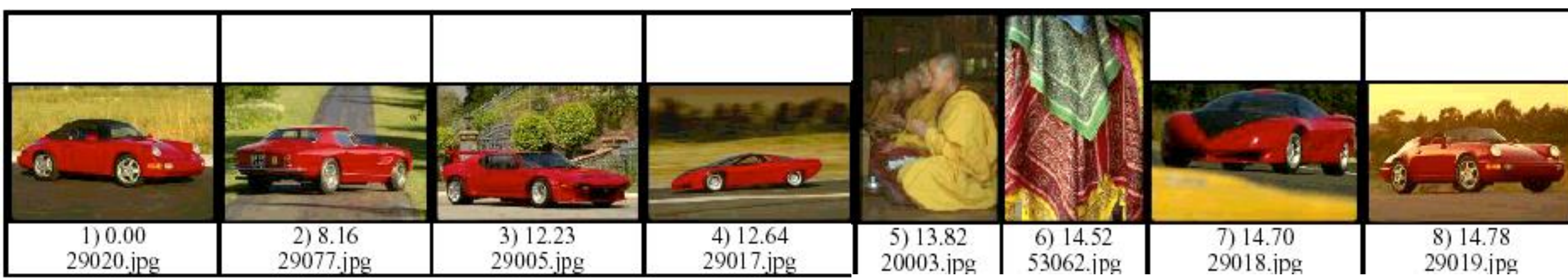


comparision

Jeffrey divergence



EMD



X: # image retrieved,
Y: #relavent images