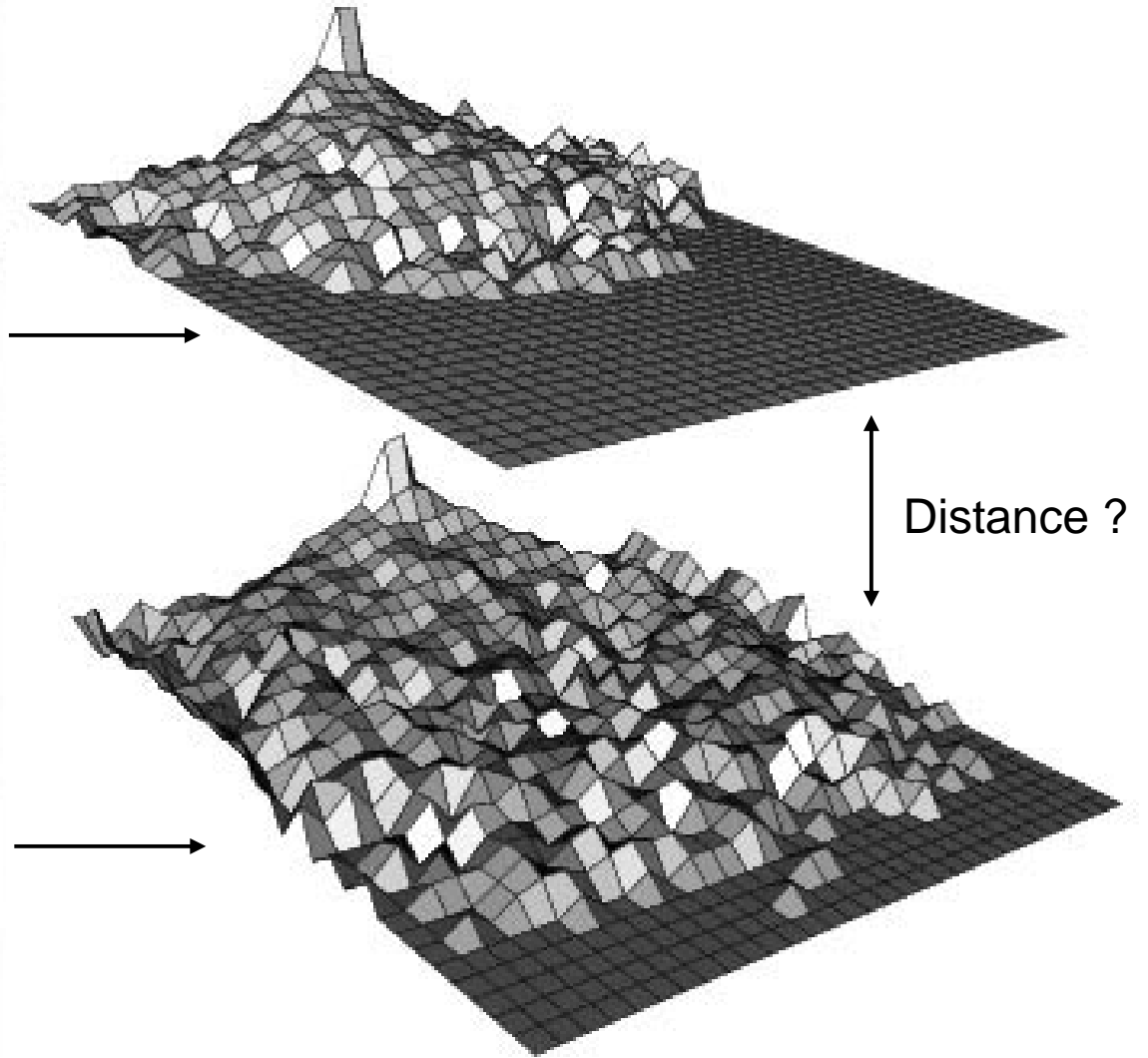
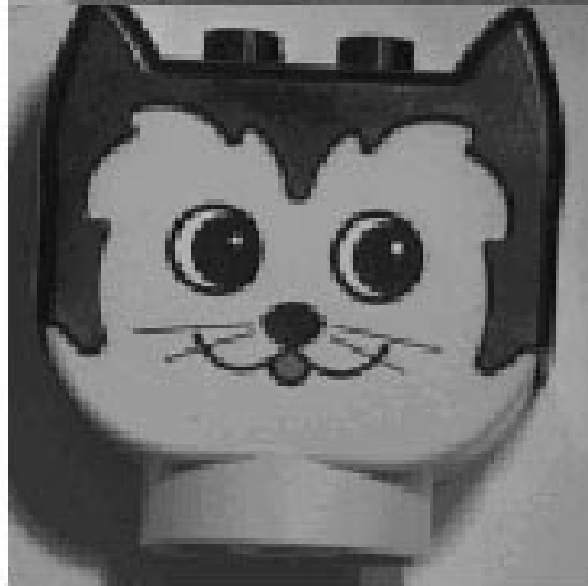
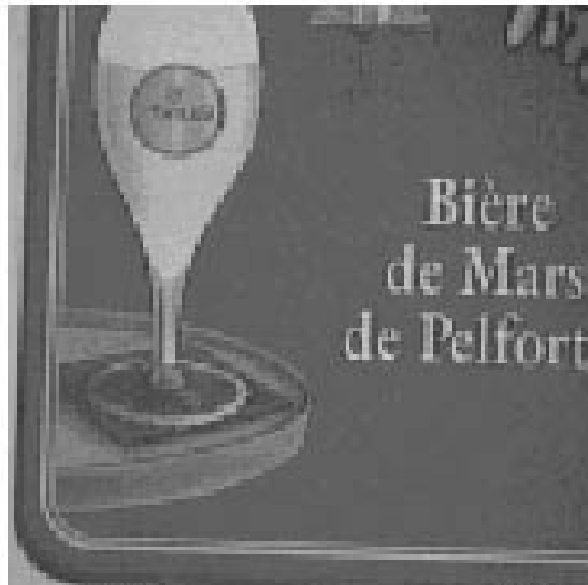


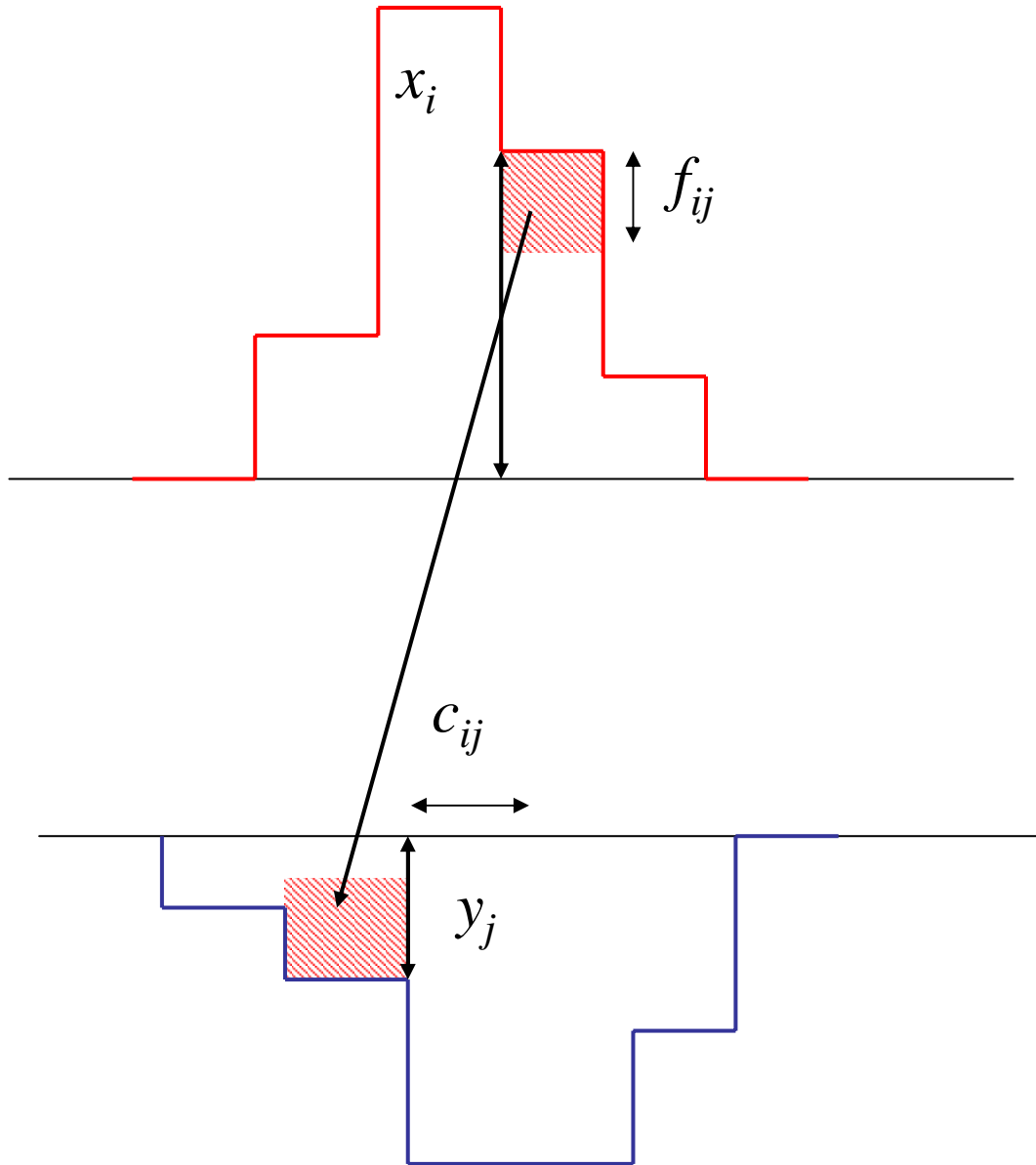
**Distances**



# Possible Distances

- L1, L2:  $\sum_i |x_i - y_i|$   $\sum_i |x_i - y_i|^2$
- Weighted L2 (multiple features):  $(X - Y)^T A(X - Y)$
- KLD:  $\sum_i x_i \log \frac{y_i}{x_i}$
- Histogram intersection:  $\frac{|X \cap Y|}{|X \cup Y|}$
- $\chi^2$ :  $\sum_i \frac{(x_i - y_i)^2}{x_i + y_i}$   $\chi^2 \sim \text{KLD} < \text{L1}$

# Earth Mover's Distance (EMD)

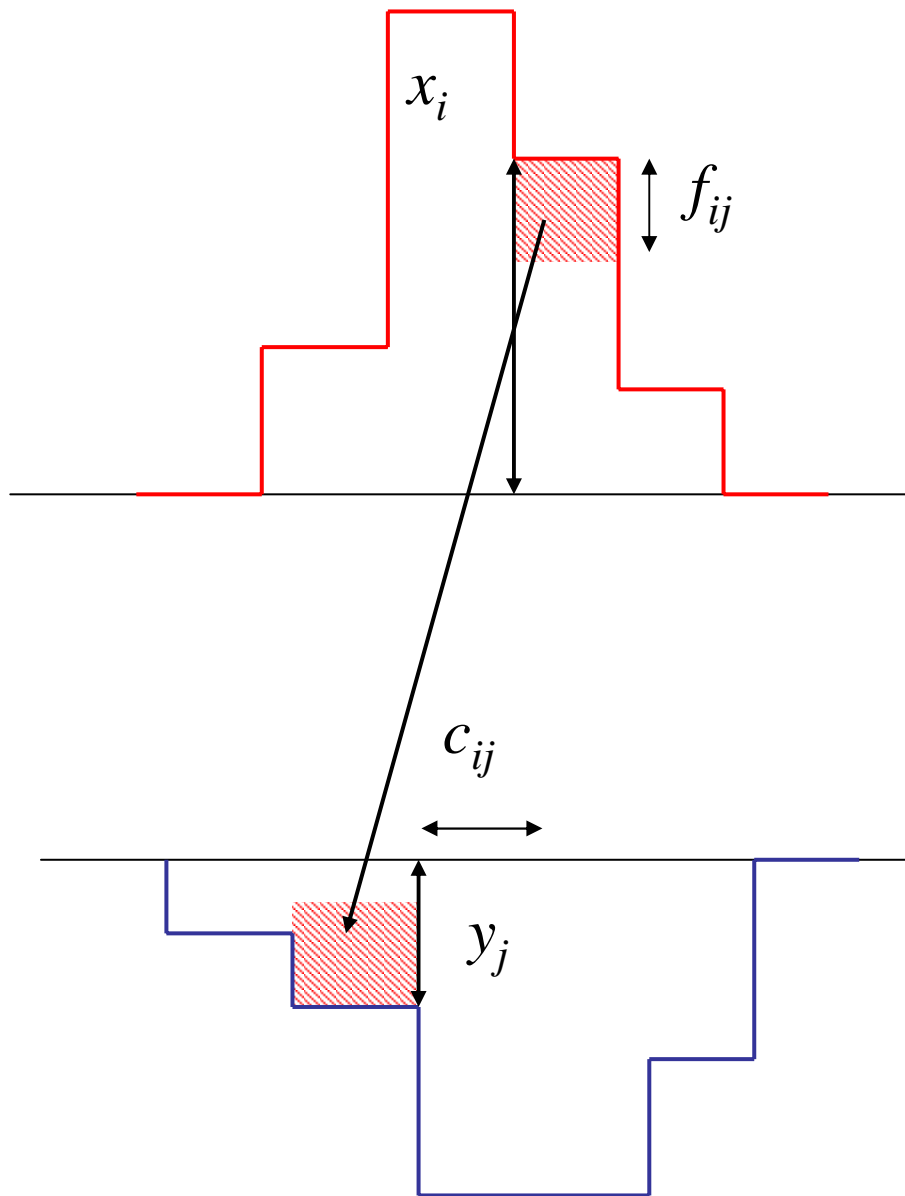


$$\text{Min} \sum_{i,j} c_{ij} f_{ij}$$

$$f_{ij} \geq 0$$

$$\sum_i f_{ij} = y_j$$

$$\sum_j f_{ij} \leq x_i$$



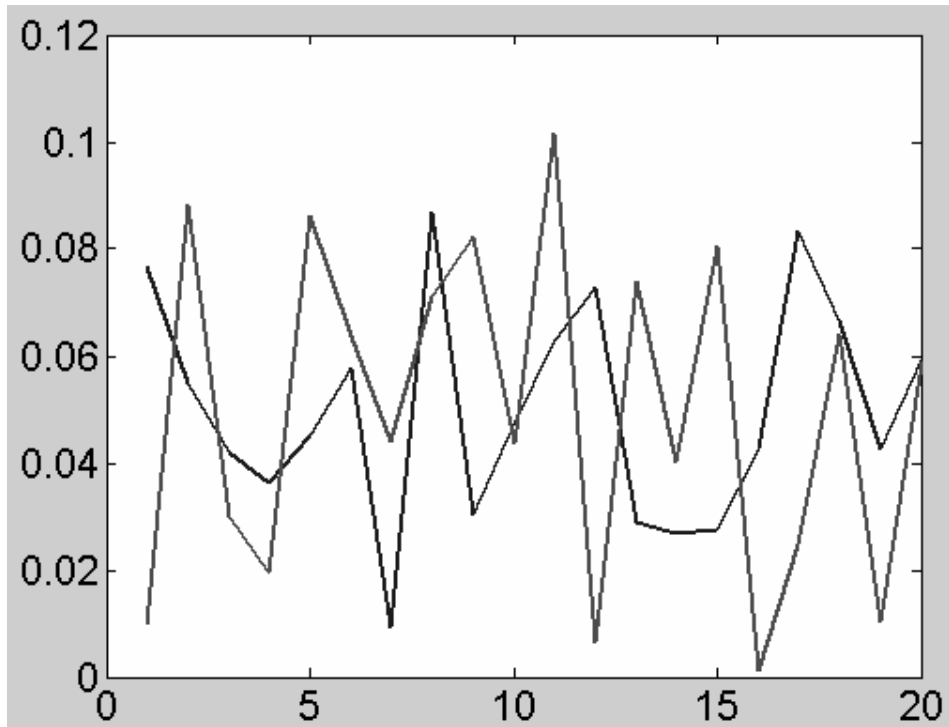
$$\text{Min} \sum_{i,j} c_{ij} f_{ij}$$

$$f_{ij} \geq 0$$

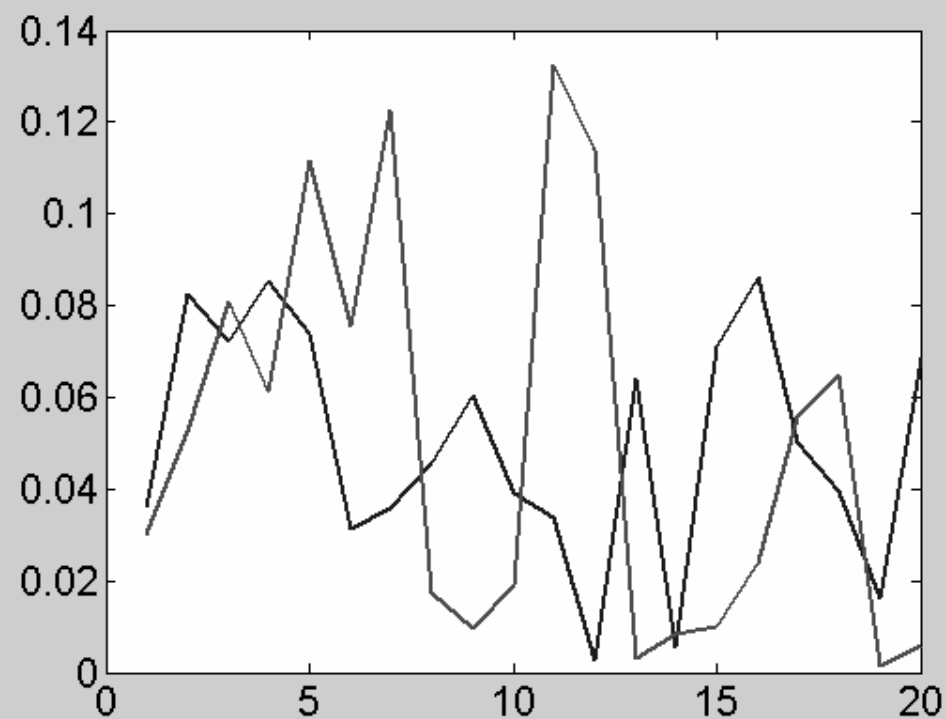
$$\sum_i f_{ij} \leq y_j$$

$$\sum_j f_{ij} \leq x_i$$

$$\sum_{i,j} f_{ij} = \text{Min}(X, Y)$$



EMD = 0.95  
L1 = 0.64



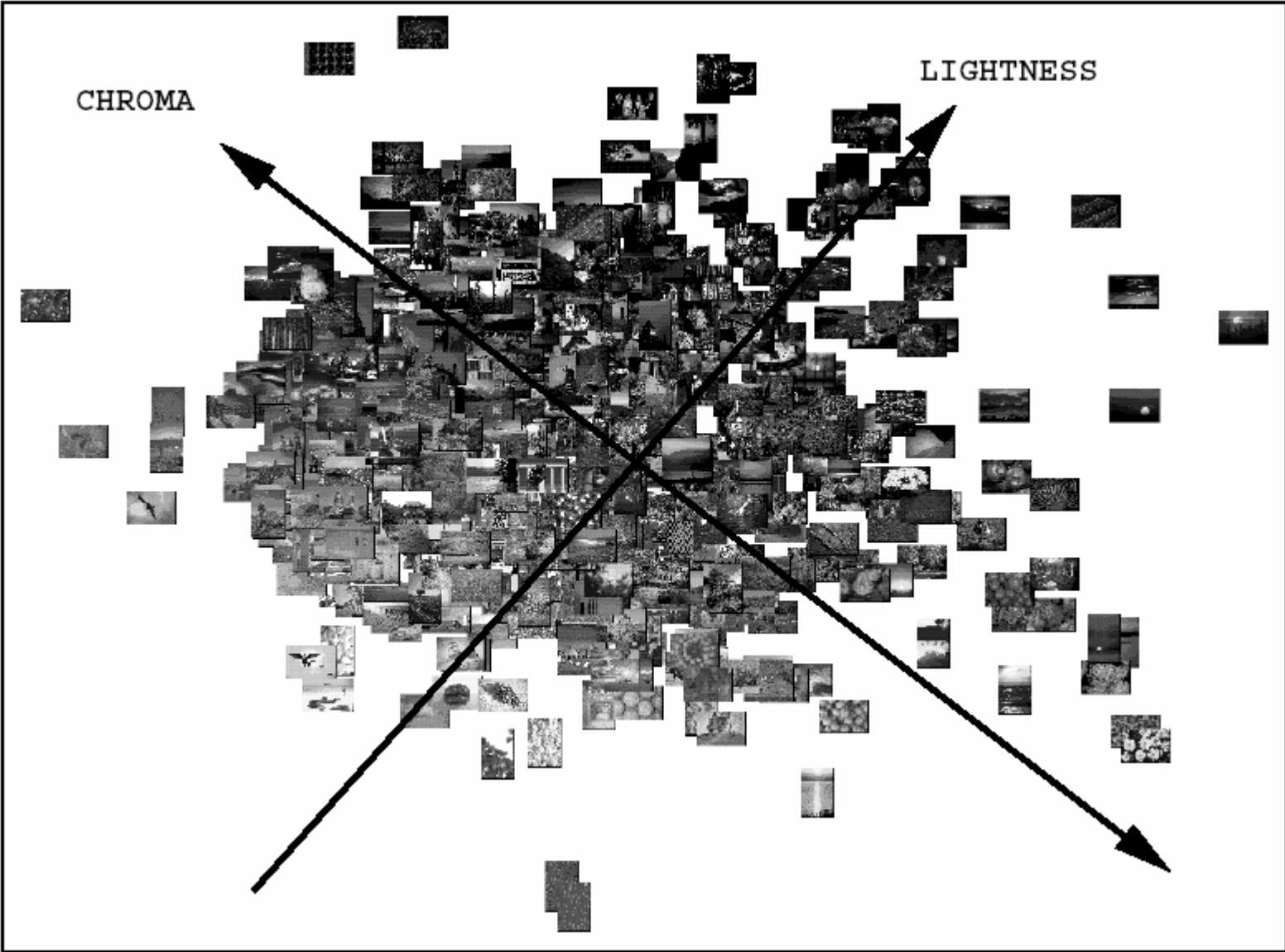
EMD = 2.25  
L1 = 0.84

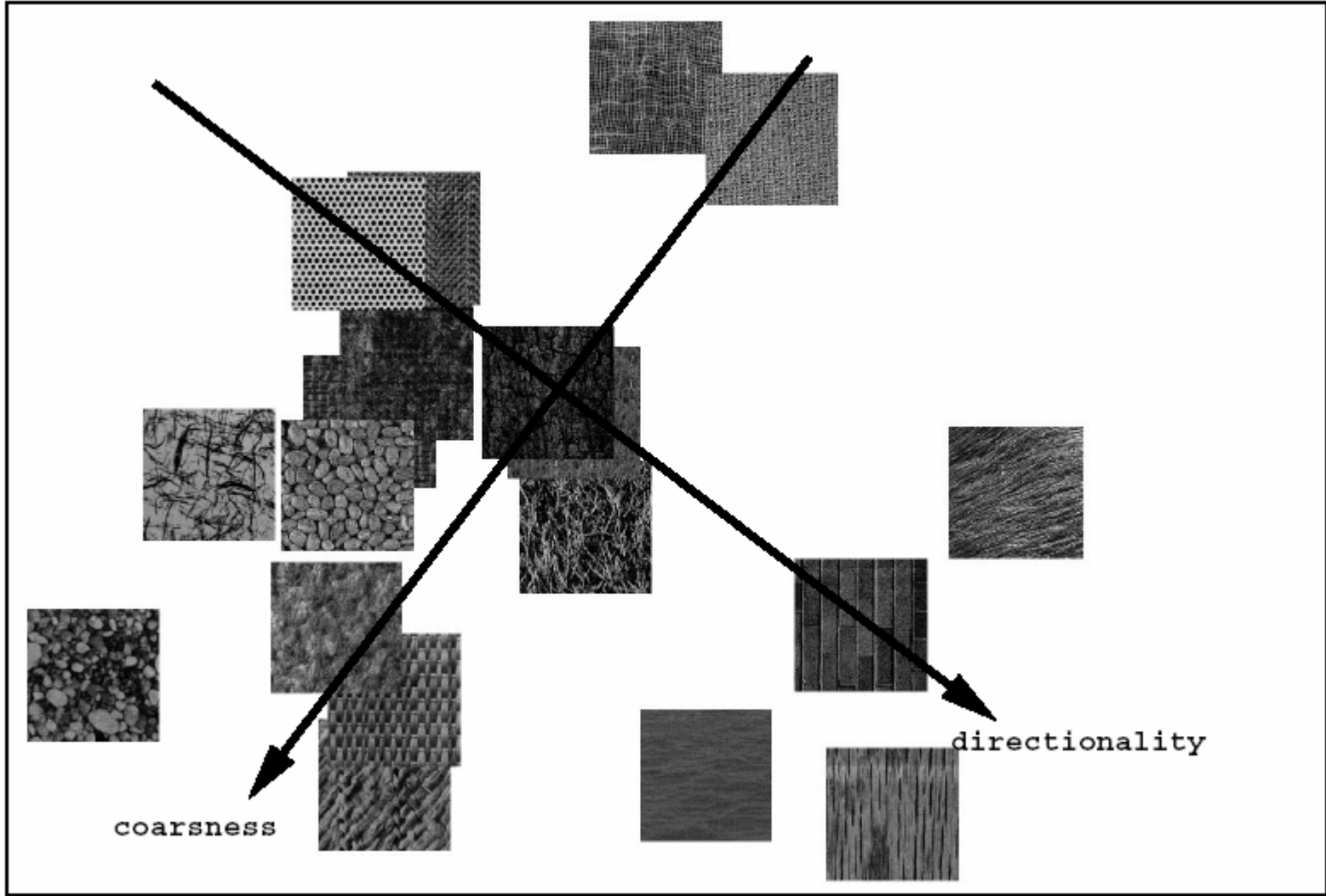
# Properties

- Application of “transportation problem. Can be solved by simplex method. Efficient approximate solution.
- Efficient lower bound. Useful for pruning pairs with large distances.
- For normalized histograms:

$$\sum_{i,j} c_{ij} f_{ij} \geq |\bar{X} - \bar{Y}|$$

- Code on line: <http://robotics.stanford.edu/~rubner/emd/default.htm>
- Y. Rubner, C. Tomasi, L.J. Guibas. A Metric for distributions with applications to image databases. ICCV, 1998.
- E.J. Russell. Extension of Dantzig’s algorithm to finding an initial near-optimal basis for the transportation problem. Operations Research, 17:187-191, 1969.





# Questions

- Dimensionality reduction (PCA, NMF, etc.)
- Choice of distances
- Combination of heterogeneous features
  - Weighted  $L^2$
  - Optimal classification distance
  - “feature” selection