Deep Learning for Computer Vision

Pyramid Matching

Vineeth N Balasubramanian

Department of Computer Science and Engineering
Indian Institute of Technology, Hyderabad

Acknowledgements

- Most of this lecture's slides are based on lectures of **Deep Learning for Vision** course taught by Prof Yannis Avrithis at Inria Rennes-Bretagne Atlantique

Recall: Descriptor Matching

- Given two images with descriptors $X, Y \subset \mathbb{R}^d$, $X_c = \{\mathbf{x} \in X : q(\mathbf{x}) = \mathbf{c}\}$ where q maps vector $\mathbf{x}$ to its nearest centroid, bag-of-words similarity on C is given by:

$$s_{BoW}(X, Y) \propto \sum_{\mathbf{c} \in C} w_c |X_c||Y_C| = \sum_{\mathbf{c} \in C} w_c \sum_{\mathbf{x} \in X_c} \sum_{\mathbf{y} \in Y_c} 1 \tag{1}$$

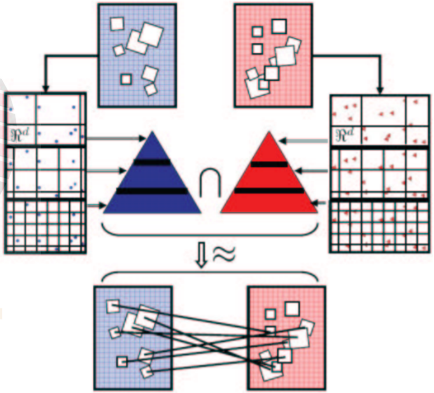
- More general form:

$$K(X, Y) := \gamma(X)\gamma(Y) \sum_{\mathbf{c} \in C} w_c M(X_c, Y_c)$$

where M is a within-cell matching function, and $\gamma(X)$ serves for normalization

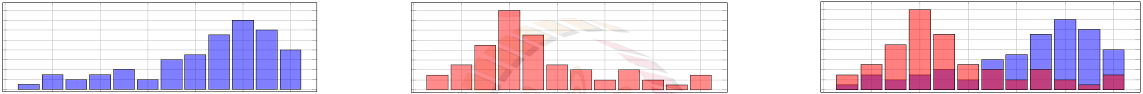
Going Beyond Single-level Matching: Pyramid Match Kernel (PMK)[1]

- **Pyramid matching:** an efficient method that maps unordered feature sets to multi-resolution histograms
- Computes a weighted histogram intersection to find implicit correspondences based on finest resolution histogram cell where a matched pair first appears
- Approximates similarity measured by optimal correspondences between feature sets of unequal cardinality



[1] Grauman and Darrell, The Pyramid Match Kernel: Discriminative Classification with Sets of Image Features, IEEE ICCV 2005, Vol 2, pp. 1458–1465

Histogram Intersection[2]



- Given two histograms $\mathbf{x}, \mathbf{y}$ of b bins each, their **histogram intersection** is:

$$\kappa_{HI}(\mathbf{x}, \mathbf{y}) = \sum_{i-1}^{b} \min(x_i, y_i)$$

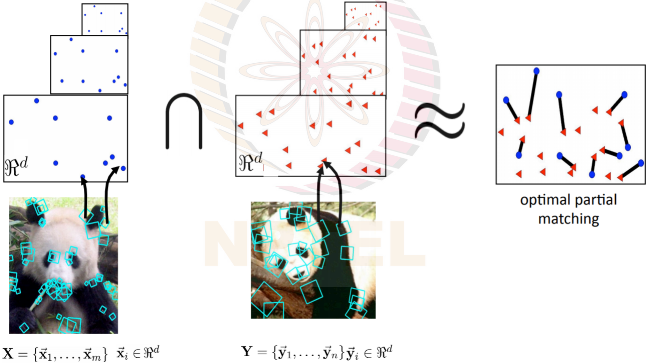
- This is related to L_1 distance as:

$$||\mathbf{x} - \mathbf{y}||_1 = ||\mathbf{x}||_1 + ||\mathbf{y}||_1 - 2\kappa_{HI}(\mathbf{x}, \mathbf{y})$$

[2]Swain and Ballard, Color Indexing, IJCV 1991, pp 11–32

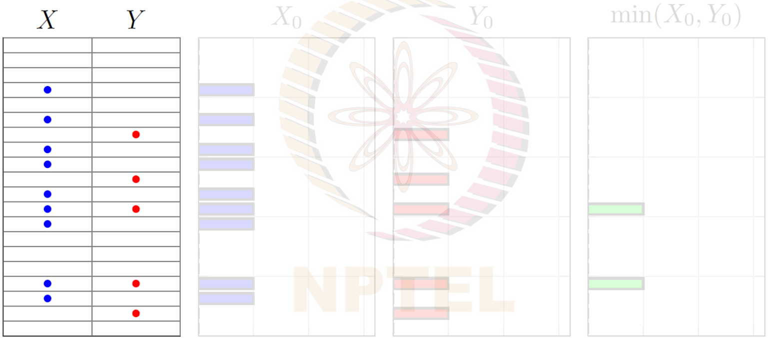
Pyramid Match Kernel (PMK)[3]

Weighted sum of histogram intersections at different levels of two images approximates their optimal pairwise matching



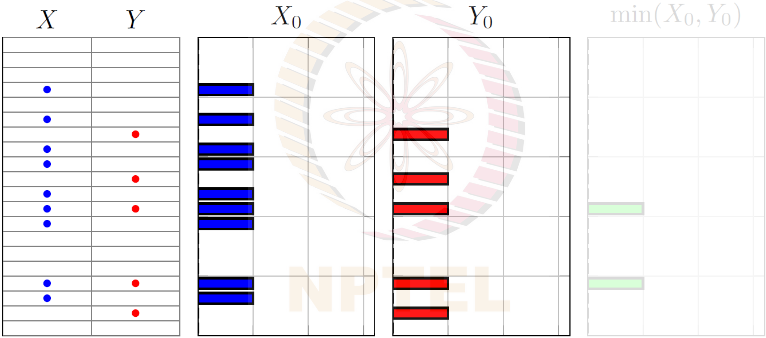
[3] Grauman and Darrell, The Pyramid Match Kernel: Discriminative Classification with Sets of Image Features, IEEE ICCV 2005, Vol 2, pp. 1458–1465

Pyramid Match Kernel: Method



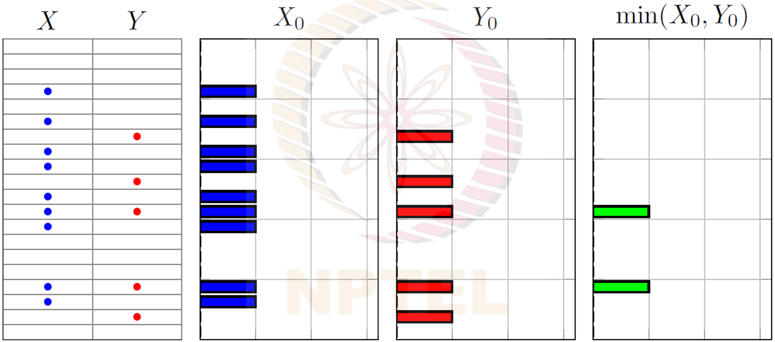
- 1-D point sets X, Y on grid of size 1

Pyramid Match Kernel: Method



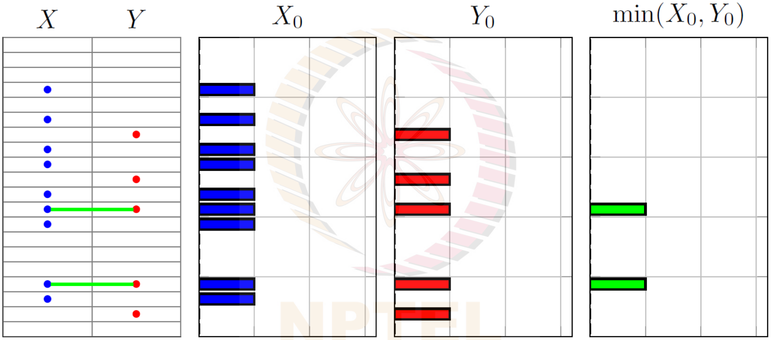
- 1-D point sets X, Y on grid of size 1 - level 0 histograms

Pyramid Match Kernel: Method



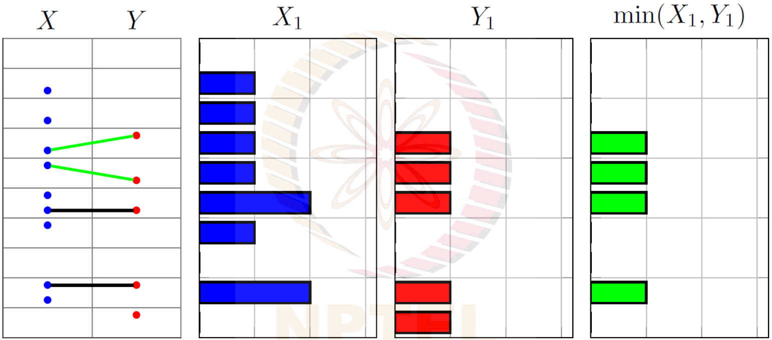
- 1-D point sets X, Y on grid of size 1 - level 0 histograms - intersection

Pyramid Match Kernel: Method



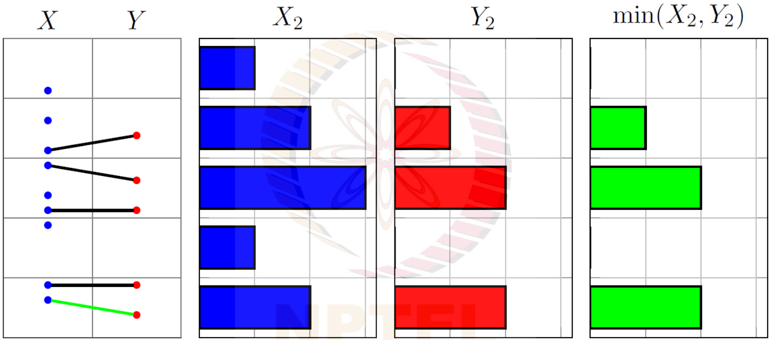
- 1-D point sets X, Y on grid of size 1 - level 0 histograms - intersection
- 2 matches weighted by 1
- Total similarity score: $2 \times 1 = 2$

Pyramid Match Kernel: Method



- 1-D point sets X, Y on grid of size 2 - level 1 histograms - intersection
- (2 matches weighted by 1) + (2 weighted by $\frac{1}{2}$)
- Total similarity score: $2 \times 1 + 2 \times \frac{1}{2} = 3$

Pyramid Match Kernel: Method



- 1-D point sets X, Y on grid of size 4 - level 2 histograms - intersection
- (2 matches weighted by 1) + (2 weighted by $\frac{1}{2}$) + (1 weighted by $\frac{1}{4}$)
- Total similarity score: $2 \times 1 + 2 \times \frac{1}{2} + 1 \times \frac{1}{4} = \mathbf{3.25}$

Pyramid Match Kernel

- Given a set $X = \{\mathbf{x}_1, \cdots, \mathbf{x}_n\} \subset \mathbb{R}^d$, where distances of elements range in $[1, D]$

Pyramid Match Kernel

- Given a set $X = \{\mathbf{x}_1, \cdots, \mathbf{x}_n\} \subset \mathbb{R}^d$, where distances of elements range in $[1, D]$
- Let X_i be a histogram of X in $\mathbb{R}^d$ on a regular grid of side length 2^i
  - i ranges from -1 (a base case where there is no intersection, 0 (where each bin has atmost one element), and so on to $L = \lceil \log_2 D \rceil$, where all of X is contained in a single bin

Pyramid Match Kernel

- Given a set $X = \{\mathbf{x}_1, \cdots, \mathbf{x}_n\} \subset \mathbb{R}^d$, where distances of elements range in $[1, D]$
- Let X_i be a histogram of X in $\mathbb{R}^d$ on a regular grid of side length 2^i
  - i ranges from -1 (a base case where there is no intersection, 0 (where each bin has atmost one element), and so on to $L = \lceil \log_2 D \rceil$, where all of X is contained in a single bin
- Given two images with descriptors $X, Y \subset \mathbb{R}^d$, their **pyramid match** is:

$$\begin{aligned} K_\Delta(X, Y) &= \gamma(X)\gamma(Y) \sum_{i=0}^{L} \frac{1}{2^i} \Big(\underbrace{\kappa_{HI}(X_i, Y_i)}_{\text{Matches at this level}} - \underbrace{\kappa_{HI}(X_{i-1}, Y_{i-1})}_{\text{Matches at previous level}} \Big) \\ &= \gamma(X)\gamma(Y) \left(\frac{1}{2^L} \kappa_{HI}(X_L, Y_L) + \sum_{i=0}^{L-1} \frac{1}{2^{i+1}} \kappa_{HI}(X_i, Y_i) \right) \end{aligned} \tag{2}$$

where $\gamma(X)$ serves for normalization

Pyramid Match Kernel

- Given a set $X = \{\mathbf{x}_1, \cdots, \mathbf{x}_n\} \subset \mathbb{R}^d$, where distances of elements range in $[1, D]$
- Let X_i be a histogram of X in $\mathbb{R}^d$ on a regular grid of side length 2^i
  - i ranges from -1 (a base case where there is no intersection, 0 (where each bin has atmost one element), and so on to $L = \lceil \log_2 D \rceil$, where all of X is contained in a single bin
- Given two images with descriptors $X, Y \subset \mathbb{R}^d$, their **pyramid match** is:

$$\begin{aligned} K_\Delta(X, Y) &= \gamma(X)\gamma(Y) \sum_{i=0}^{L} \frac{1}{2^i} \Bigg(\underbrace{\kappa_{HI}(X_i, Y_i)}_{\text{Matches at this level}} - \underbrace{\kappa_{HI}(X_{i-1}, Y_{i-1})}_{\text{Matches at previous level}} \Bigg) \\ &= \gamma(X)\gamma(Y) \left(\frac{1}{2^L} \kappa_{HI}(X_L, Y_L) + \sum_{i=0}^{L-1} \frac{1}{2^{i+1}} \kappa_{HI}(X_i, Y_i) \right) \end{aligned} \tag{2}$$

where $\gamma(X)$ serves for normalization

Counts number of new pairs matched

PMK is a Positive Definite Kernel

- K_Δ can be written as a weighted sum of κ_{HI} terms, with non-negative coefficients
- κ_{HI} can be written as a sum of min terms
- min can be written as a dot product:

x	$\phi(x)$							
3	1	1	1	0	0	0	0	0
5	1	1	1	1	1	0	0	0
$\min(x, y) = 3$	1	1	1	0	0	0	0	0

- Therefore, so can K_Δ

PMK as an Embedding[4]

- There is an explicit embedding for κ_{HI}, therefore also for K_{Δ}. What could it be?

[4] Indyk and Thaper, Fast Image Retrieval via Embeddings, WSCTV. 2003

PMK as an Embedding[4]

- There is an explicit embedding for κ_{HI}, therefore also for K_Δ. What could it be?
- If $|X| \leq |Y|$ and $\pi : X \to Y$ is one-to-one, then $K_\Delta(X, Y)$ approximates the optimal pairwise matching:

$$\max_{\pi} \sum_{\mathbf{x} \in X} ||\mathbf{x} - \pi(\mathbf{x})||_1^{-1}$$

[4] Indyk and Thaper, Fast Image Retrieval via Embeddings, WSCTV. 2003

PMK as an Embedding[4]

- There is an explicit embedding for κ_{HI}, therefore also for K_Δ. What could it be?
- If $|X| \leq |Y|$ and $\pi : X \to Y$ is one-to-one, then $K_\Delta(X, Y)$ approximates the optimal pairwise matching:

$$\max_{\pi} \sum_{\mathbf{x} \in X} ||\mathbf{x} - \pi(\mathbf{x})||_1^{-1}$$

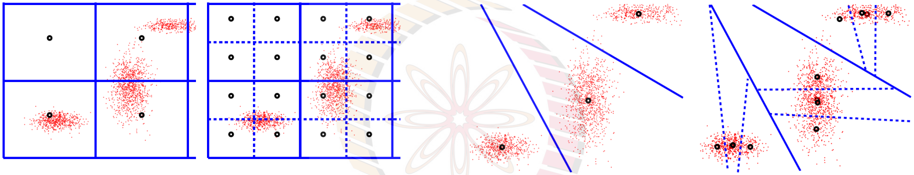
- This is similar to the **Earth mover's distance**:

$$\min_{\pi} \sum_{\mathbf{x} \in X} ||\mathbf{x} - \pi(\mathbf{x})||_1$$

- But PMK is a *similarity* measure; it allows partial matching and does not penalize clutter, except for the normalization

[4] Indyk and Thaper, Fast Image Retrieval via Embeddings, WSCTV. 2003

PMK and Vocabulary Tree[5]



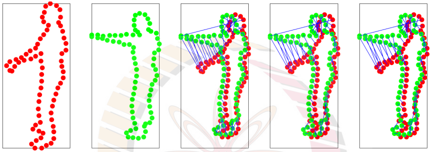
Uniform bins

Vocabulary-guided bins

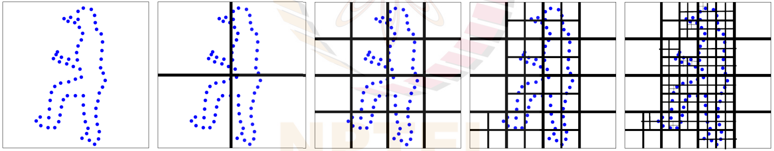
- Replace regular grid with hierarchical vocabulary cells
- Compared to vocabulary tree, there is a principle in assigning cell weights
- Still, its approximation quality can suffer at high dimensions

[5] Grauman and Darrell, Approximate Correspondences in High Dimensions, NeurIPS 2007

PMK and Spatial Matching[6]



Optimal Matching

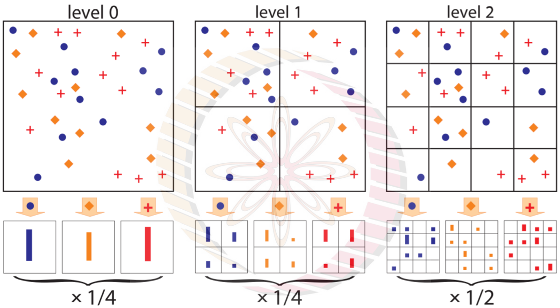


Representation

- Same idea, applied to image 2-D coordinate space for spatial matching
- Matching cost is only based on point coordinates; No appearance

[6]Grauman and Darrell, Fast Contour Matching Using Approximate Earth Mover's Distance, CVPR 2004

Spatial Pyramid Matching (SPM)[7]

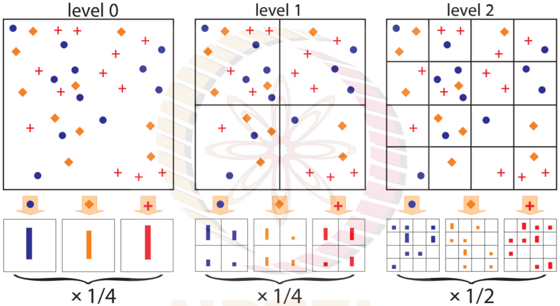


- If $X^{(j)}$, $Y^{(j)}$ are the feature coordinates of images X, Y with descriptors assigned to visual word j,

$$K_{SP}(X, Y) = \sum_{j=1}^{k} K_{\Delta}(X^{(j)}, Y^{(j)})$$

[7] Lazebnik et al, Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories, CVPR 2006

Spatial Pyramid Matching (SPM)[7]



- Coupled with BoW, it is a set of joint appearance-geometry histograms
- Robust to deformation but not invariant to transformations; Applied for global scene classification

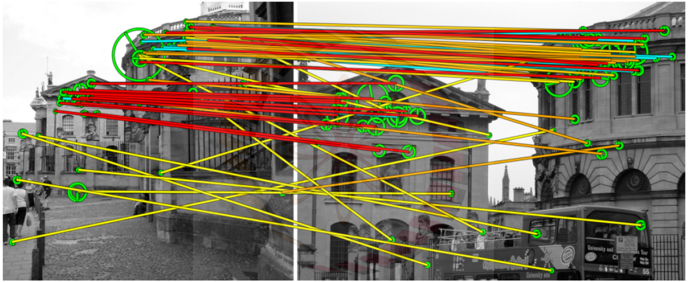
[7] Lazebnik et al, Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories, CVPR 2006.

Hough Pyramid Matching (HPM)[8]

Fast Pyramid Matching

[8] Tolias and Avrithis, Speeded-up, relaxed spatial matching, ICCV 2011

Hough Pyramid Matching (HPM)[8]



Hough Pyramid Matching

- Work with a single set of correspondences instead of two sets of features
- Determine a transformation hypothesis by a pair of features and then use histograms to collect votes in the transformation space

[8]Tolias and Avrithis, Speeded-up, relaxed spatial matching, ICCV 2011

Hough Pyramid Matching

- A **local feature** p in image P has position $\mathbf{t}(p)$, scale $s(p)$ and orientation $\theta(p)$ given by matrix $R(p) \in \mathbb{R}^{2\times2}$:

$$F(p) = \begin{pmatrix} s(p)R(p) & \mathbf{t}(p) \\ \mathbf{0}^\top & 1 \end{pmatrix}$$

Hough Pyramid Matching

- A **local feature** p in image P has position $\mathbf{t}(p)$, scale $s(p)$ and orientation $\theta(p)$ given by matrix $R(p) \in \mathbb{R}^{2\times 2}$:

$$F(p) = \begin{pmatrix} s(p)R(p) & \mathbf{t}(p) \\ \mathbf{0}^\top & 1 \end{pmatrix}$$

- A **correspondence** $c = (p, q)$ is a pair of features $p \in P$, $q \in Q$ of two images P, Q and determines relative similarity transformation from p to q:

$$F(c) = F(q)F(p)^{-1} = \begin{pmatrix} s(c)R(c) & \mathbf{t}(c) \\ \mathbf{0}^\top & 1 \end{pmatrix}$$

with translation $\mathbf{t}(c) = \mathbf{t}(q) - s(c)R(c)\mathbf{t}(p)$, relative scale $s(c) = s(q)/s(p)$ and rotation $R(c) = R(q)R(p)^{-1}$ or $\theta(c) = \theta(q) - \theta(p)$

Hough Pyramid Matching

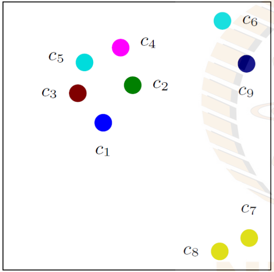
- The 4-DoF relative transformation represented by 4-D vector:

$$f(c) = (\mathbf{t}(c), s(c), \theta(c))$$

- To enforce one-to-one mapping, two correspondences $c = (p, q)$ and $c' = (p', q')$ are said to be **conflicting** if they refer to the same feature on either image, i.e., $p = p'$ or $q = q'$

Hough Pyramid Matching

correspondences

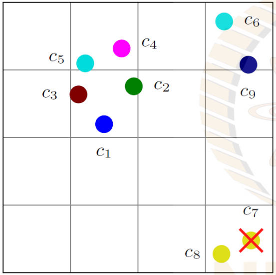


	p	q	similarity score
c_1			$(2 + \frac{1}{2}2 + \frac{1}{4}2)w(c_1)$
c_2			$(2 + \frac{1}{2}2 + \frac{1}{4}2)w(c_2)$
c_3			$(2 + \frac{1}{2}2 + \frac{1}{4}2)w(c_3)$
c_4			$(1 + \frac{1}{2}3 + \frac{1}{4}2)w(c_4)$
c_5			$(1 + \frac{1}{2}3 + \frac{1}{4}2)w(c_5)$
c_6			0
c_7			0
c_8			$(\frac{1}{4}6)w(c_8)$
c_9			$(\frac{1}{4}6)w(c_9)$

- Correspondence c weighted by $w(c)$, based e.g. on visual word

Hough Pyramid Matching

pyramid level 0

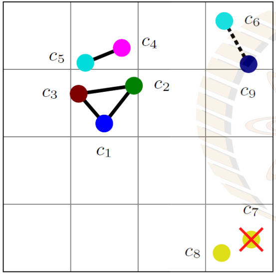


	p	q	similarity score
c_1			$(2 + \frac{1}{2}2 + \frac{1}{4}2)w(c_1)$
c_2			$(2 + \frac{1}{2}2 + \frac{1}{4}2)w(c_2)$
c_3			$(2 + \frac{1}{2}2 + \frac{1}{4}2)w(c_3)$
c_4			$(1 + \frac{1}{2}3 + \frac{1}{4}2)w(c_4)$
c_5			$(1 + \frac{1}{2}3 + \frac{1}{4}2)w(c_5)$
c_6			0
c_7			0
c_8			$(\frac{1}{4}6)w(c_8)$
c_9			$(\frac{1}{4}6)w(c_9)$

- Correspondence c weighted by $w(c)$, based e.g. on visual word
- Conflicting correspondences in same bin are **erased**

Hough Pyramid Matching

pyramid level 0

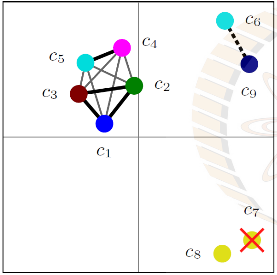


	p	q	similarity score
c_1			$(2 + \frac{1}{2}2 + \frac{1}{4}2)w(c_1)$
c_2			$(2 + \frac{1}{2}2 + \frac{1}{4}2)w(c_2)$
c_3			$(2 + \frac{1}{2}2 + \frac{1}{4}2)w(c_3)$
c_4			$(1 + \frac{1}{2}3 + \frac{1}{4}2)w(c_4)$
c_5			$(1 + \frac{1}{2}3 + \frac{1}{4}2)w(c_5)$
c_6			0
c_7			0
c_8			$(\frac{1}{4}6)w(c_8)$
c_9			$(\frac{1}{4}6)w(c_9)$

- Correspondence c weighted by $w(c)$, based e.g. on visual word
- Conflicting correspondences in same bin are **erased**
- In bin b with n_b correspondences, each correspondence groups with $[n_b - 1]_+$ others
- Level 0 Weight 1

Hough Pyramid Matching

pyramid level 1

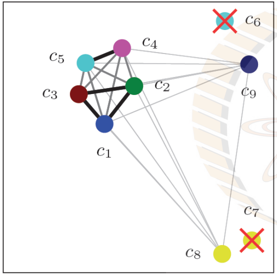


	p	q	similarity score
c_1			$(2 + \frac{1}{2}2 + \frac{1}{4}2)w(c_1)$
c_2			$(2 + \frac{1}{2}2 + \frac{1}{4}2)w(c_2)$
c_3			$(2 + \frac{1}{2}2 + \frac{1}{4}2)w(c_3)$
c_4			$(1 + \frac{1}{2}3 + \frac{1}{4}2)w(c_4)$
c_5			$(1 + \frac{1}{2}3 + \frac{1}{4}2)w(c_5)$
c_6			0
c_7			0
c_8			$(\frac{1}{4}6)w(c_8)$
c_9			$(\frac{1}{4}6)w(c_9)$

- Correspondence c weighted by $w(c)$, based e.g. on visual word
- Conflicting correspondences in same bin are **erased**
- In bin b with n_b correspondences, each correspondence groups with $[n_b - 1]_+$ others
- Level 1 Weight $\frac{1}{2}$

Hough Pyramid Matching

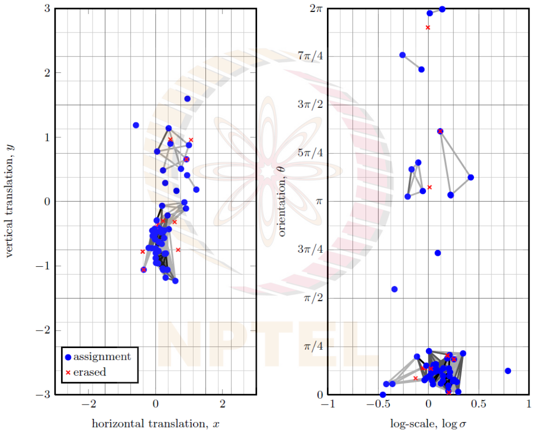
pyramid level 2



	p	q	similarity score
c_1			$(2+\frac{1}{2}2+\frac{1}{4}2)w(c_1)$
c_2			$(2+\frac{1}{2}2+\frac{1}{4}2)w(c_2)$
c_3			$(2+\frac{1}{2}2+\frac{1}{4}2)w(c_3)$
c_4			$(1+\frac{1}{2}3+\frac{1}{4}2)w(c_4)$
c_5			$(1+\frac{1}{2}3+\frac{1}{4}2)w(c_5)$
c_6			0
c_7			0
c_8			$(\frac{1}{4}6)w(c_8)$
c_9			$(\frac{1}{4}6)w(c_9)$

- Correspondence c weighted by $w(c)$, based e.g. on visual word
- Conflicting correspondences in same bin are **erased**
- In bin b with n_b correspondences, each correspondence groups with $[n_b - 1]_+$ others
- Level 2 Weight $\frac{1}{4}$

Hough Pyramid Matching



- *Mode Seeking:* We are looking for regions where density is maximized in transformation space

Hough Pyramid Matching

- Linear in number of correspondences; no need to count inliers
- Robust to deformations and multiple matching surfaces, invariant to transformations
- Only applies to same instance matching

Homework

Readings

- Chapter 16.1.4, Forsyth and Ponce, *Computer Vision: A Modern Approach (2nd ed.)*

Other Resources

- Pyramid Match Kernel project page

References

- Michael J. Swain and Dana H. Ballard. "Color Indexing". In: *Int. J. Comput. Vision* 7.1 (Nov. 1991), 11–32.
- Piotr Indyk and Nitin Thaper. "Fast Image Retrieval via Embeddings". In: *WSCTV*. 2003.
- Kristen Grauman and Trevor Darrell. "Fast Contour Matching Using Approximate Earth Mover's Distance". In: *IEEE CVPR*. Oct. 2004, pp. I–220.
- K. Grauman and T. Darrell. "The Pyramid Match Kernel: Discriminative Classification with Sets of Image Features". In: *IEEE ICCV*. Nov. 2005, Vol 2, pp. 1458–1465.
- Svetlana Lazebnik, Cordelia Schmid, and J. Ponce. "Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories". In: *IEEE CVPR*. Vol. 2. Feb. 2006, pp. 2169 –2178.
- Kristen Grauman and Trevor Darrell. "Approximate Correspondences in High Dimensions". In: *Neural Information Processing Systems*. 2007.
- Giorgos Tolias and Yannis Avrithis. "Speeded-up, relaxed spatial matching". In: *IEEE ICCV*. 2011, pp. 1653–1660.
- David Forsyth and Jean Ponce. *Computer Vision: A Modern Approach*. 2 edition. Boston: Pearson Education India, 2015.