Speeded-up, relaxed spatial matching

Giorgos Tolias and Yannis Avrithis
School of Electrical and Computer Engineering, National Technical University of Athens, Greece

Overview

- Fast re-ranking of top matching images in large scale retrieval.
- Inspired by Hough voting and pyramid matching.
- Relaxed and flexible matching model.
- Allow non-rigid motion and multiple matching surfaces.
- Linear in the number of correspondences.

Problem

Local shape – transformation space

- Scale and rotation invariant local feature $p \in P$

$$F(p) = \begin{bmatrix} M(p) & \mathbf{t}(p) \\ \mathbf{0}^{\mathrm{T}} & 1 \end{bmatrix}, \quad M(p) = \sigma(p)R(p).$$

- Set of candidate correspondences C according to proximity in descriptor space (eg. visual vocabulary) $C = \{(p,q) \in P \times Q : u(p) = u(q)\}$.
- Relative transformation for *correspondence* (assignment) $c = (p,q)$

$$F(c) = F(q)F(p)^{-1} = \begin{bmatrix} M(c) & \mathbf{t}(c) \\ \mathbf{0}^{\mathrm{T}} & 1 \end{bmatrix}.$$

- Parameter vector: 4-dof transformation (translation, relative log-scale, relative orientation)

$$f(c) = (x(c), y(c), \sigma(c), \theta(c)).$$

Compatibility of assignments

- For $c, c' \in C$, an *affinity* score $\alpha(c, c')$ measures their similarity in the *transformation space*
- One-to-one mapping: two assignments $c = (p,q)$, $c' = (p', q')$ are *compatible* if $p \neq p'$ and $q \neq q'$, and *conflicting* otherwise.

Goal

- Identify subset of pairwise compatible assignments that maximizes the total weighted, pairwise affinity.
- Estimate a total image similarity score – no transformation estimation needed.

Hough Pyramid Matching (HPM)

- *Hierarchical partition* $\mathcal{B} = \{B_0, \dots, B_{L-1}\}$ of transformation space $\mathcal{F}$ into L levels.
- *Histogram pyramid* of correspondences into bins $b \in B_\ell$ at level ℓ

$$h(b) = \{c \in C : f(c) \in b\}.$$

- Detect conflicting correspondences at each level; greedily choose the best one to keep; maintain the remaining in set of *erased* X. Histogram pyramid is now $\hat{h}(b) = h(b) \setminus X$.
- Isolated correspondences do not form a group; *group count* of bin b

$$g(b) = \max\{0, |\hat{h}(b)| - 1\}.$$

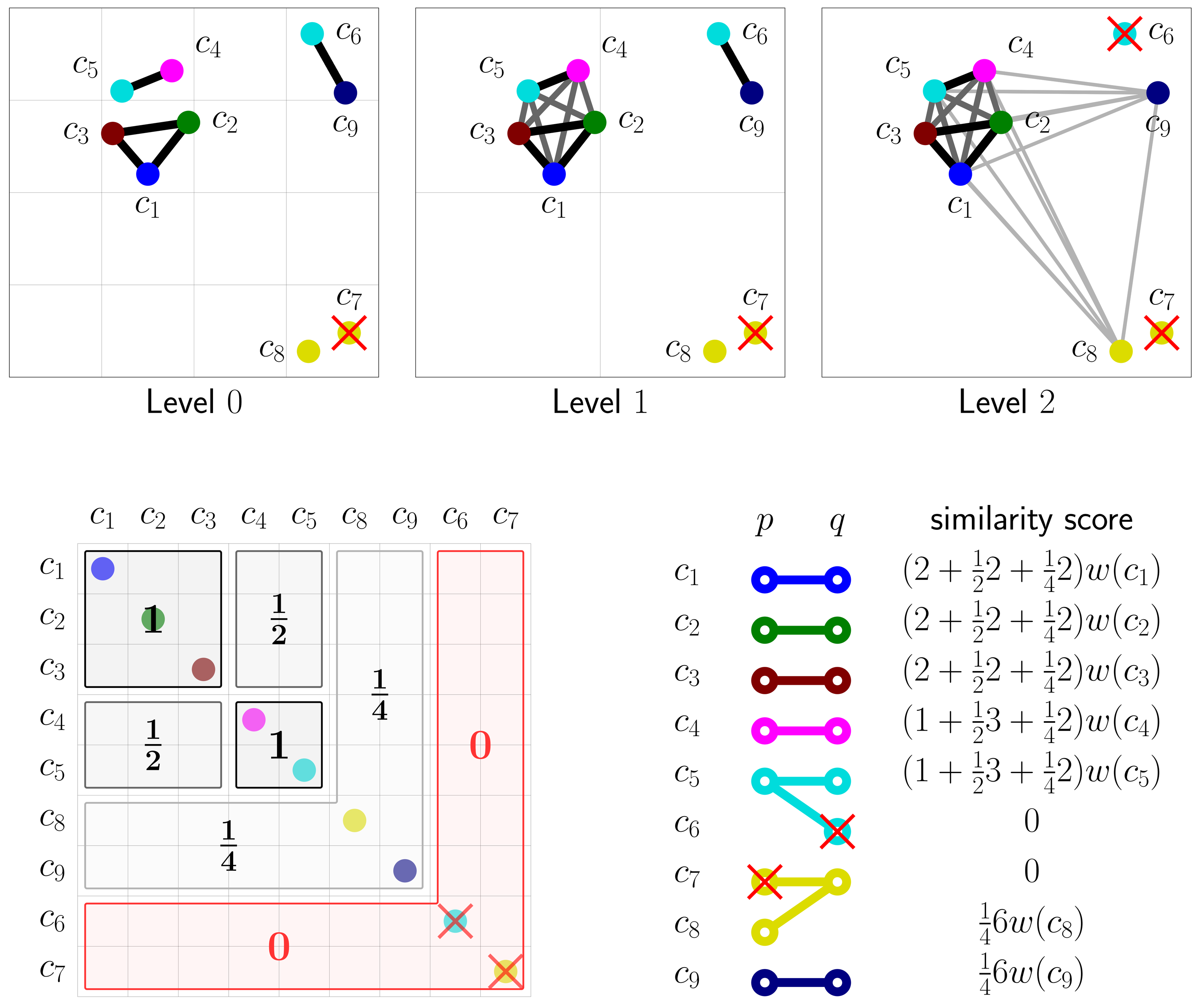
- Newly grouped correspondences with c at level ℓ are $g(b_\ell) - g(b_{\ell-1})$ and affinity at level ℓ is approximated with a non-increasing function. *Strength* of c up to level ℓ

$$s_\ell(c) = g(b_0) + \sum_{k=1}^{\ell} 2^{-k}\{g(b_k) - g(b_{k-1})\}.$$

- Image similarity score as a weighted sum of strengths at the top level

$$s(C) = \sum_{c \in C \setminus X} w(c) s_{L-1}(c).$$

Toy example



Experimental results

- **Memory usage reduction** by uniform quantization of local feature shape.
- Use 5 levels and 16 bins for each parameter – run length encoding for image id.

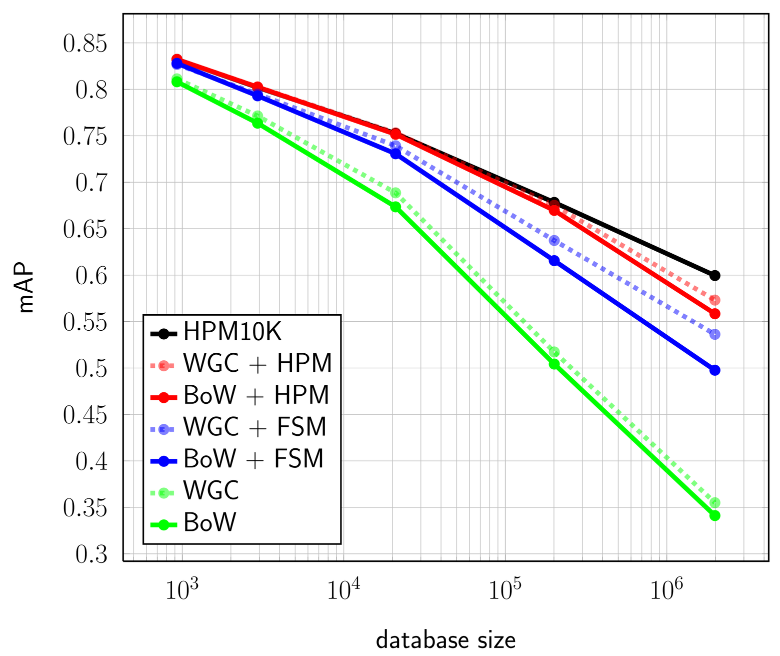
image id	x	y	$\log\sigma$	θ	total
16	4	4	4	4	32

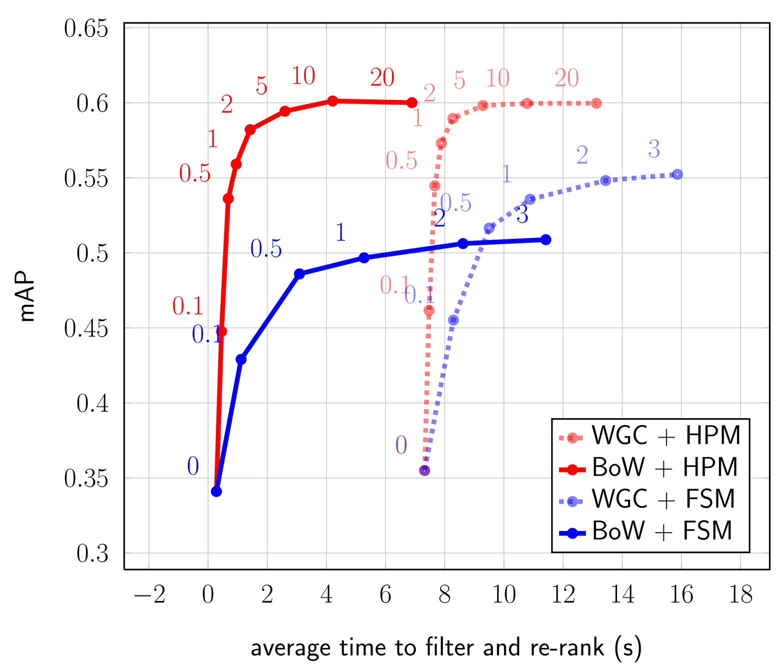
Inverted file memory usage per local feature, in bits.

- **mean Average Precision (mAP) for pyramid and flat matching** at different levels L with 2M distractors and re-ranking top 1000 images.

L	2	3	4	5	6
pyramid	0.473	0.498	0.536	0.556	**0.559**
flat	0.448	0.485	0.524	**0.534**	0.509

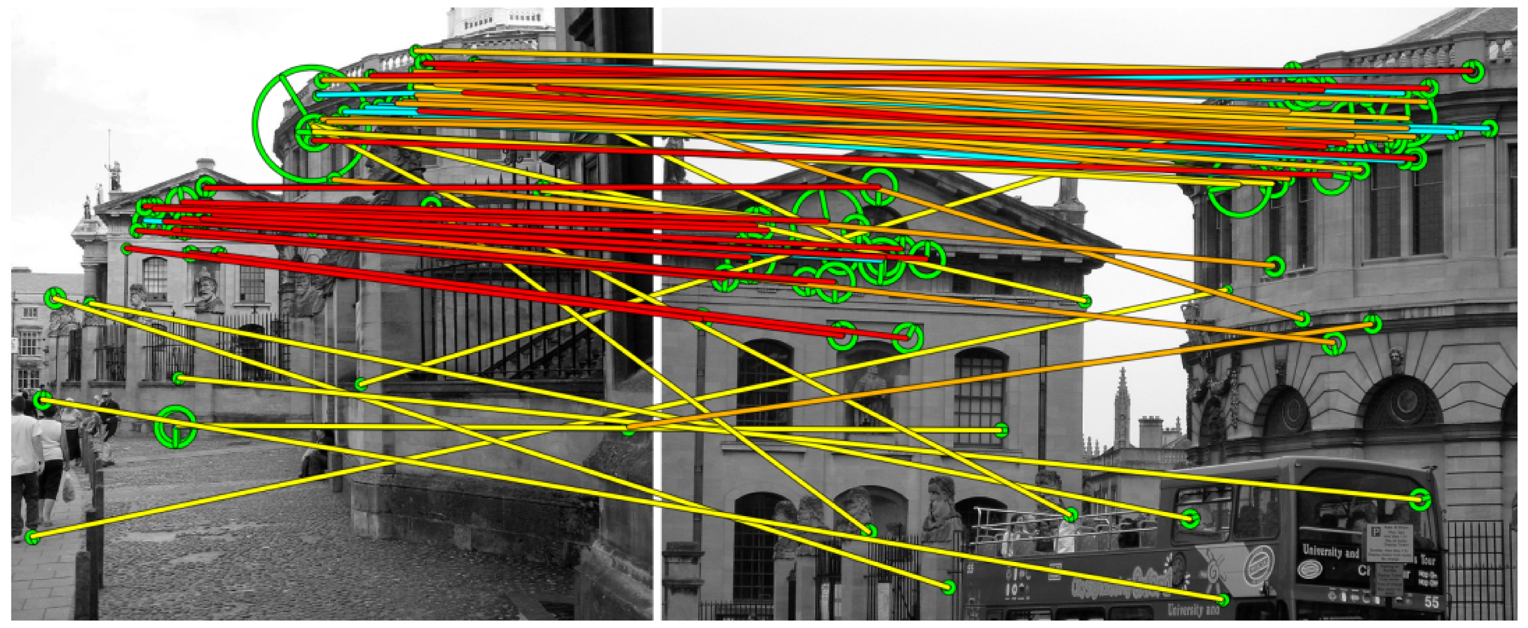
- **Large scale experiments** with up to 2M distractors.





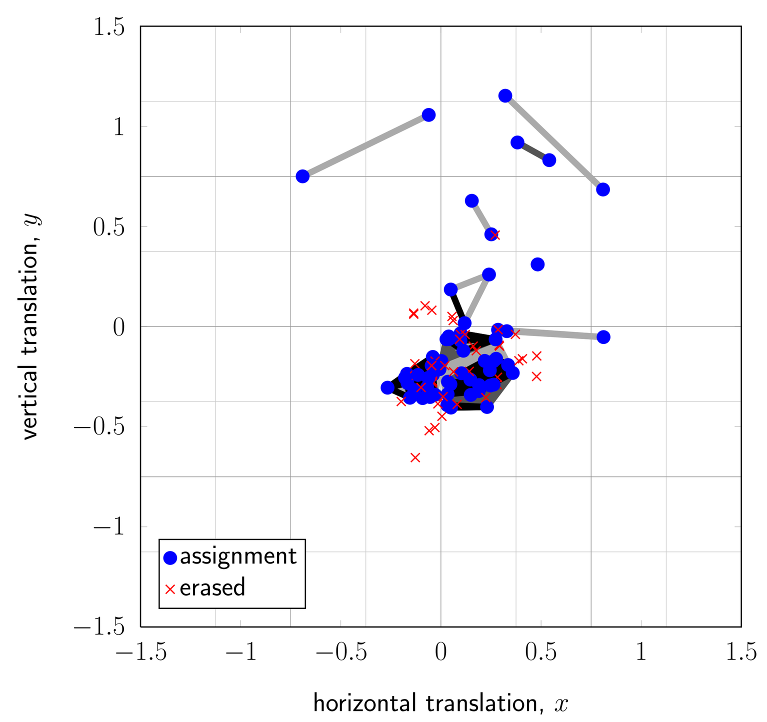
Matching examples

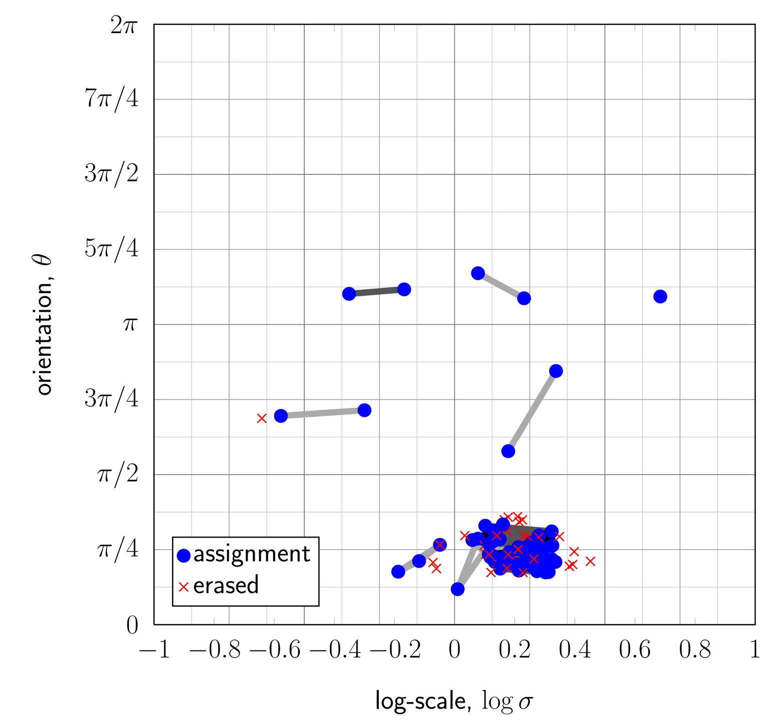
- **Matching with HPM (0.6ms).** All tentative correspondences are shown. The ones in cyan have been *erased*. The rest are colored according to strength, with red (yellow) being the strongest (weakest).



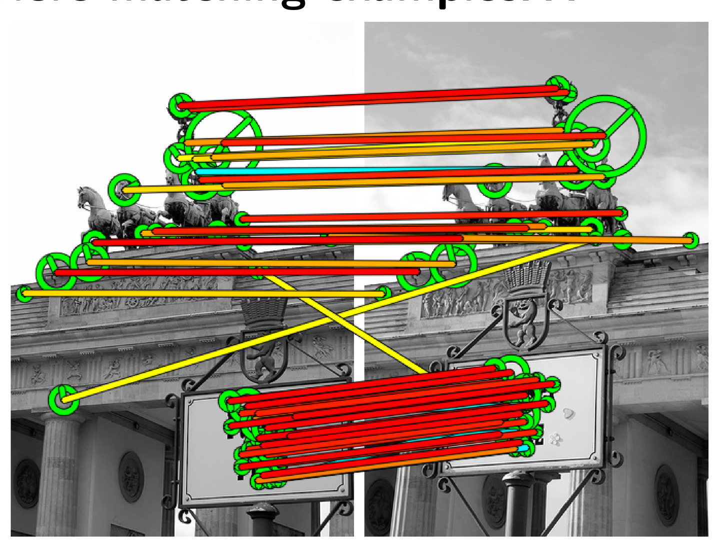
- **Matching with Fast Spatial Matching (7ms).** Inliers with a 4-dof model are shown in red.

- **Correspondences as votes in 4D transformation space**. Three tones of gray for *level affinity*.





- **More matching examples. . .**



Contact: {gtolias, iavr}@image.ntua.gr