The 20th Anniversary ACM Multimedia 2012

SymCity: Feature Selection by Symmetry for Large Scale Image Retrieval

Giorgos Tolias, Yannis Kalantidis and Yannis Avrithis

National Technical University of Athens

Introduction

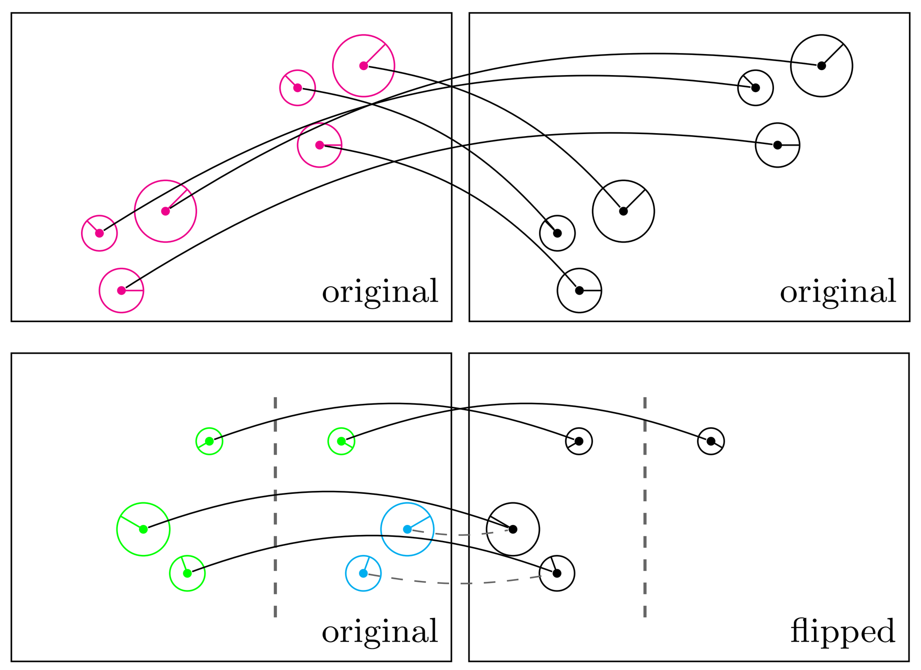
- **Scope:** search in a large corpus of images and retrieve a specific object
- **Challenge:** reduce memory requirements without sacrificing performance
- **Bag-of-Words (BoW):** good performance at low cost, but indexes each local feature separately
- **Geometry verification:** constantly better performance than BoW, with roughly same memory requirements
- **Compact representations:** much lower memory requirements, *e.g.* Fisher vectors [Perronnin et al. 2010], not compatible with geometry verification
- **Feature Selection:** currently only from multiple views
- **Our solution:** selection from single views via symmetry and repeating pattern detection

Related work: Feature selection from multiple views

- **Supervised (by geo-tag):**
  - informative feature selection [Schindler *et al.* 2007] [Li & Kosecka 2006]
  - foreground object detection [Gammeter *et al.* 2009]
  - scene map construction [Avrithis *et al.* 2010].
- **Unsupervised:** Spatial verification of multiple views [Turcot & Lowe 2009]

five different views matched by RANSAC | multiple view selection [Turcot & Lowe 2009] | single view selection (this work)

Feature selection from a single view



Self-matching

Flipped matching

Tentative Correspondences:

- Valid pairs: $C_v(X) = \{(x,y) \in X^2 : v(x,y)\}$
- Descriptor nearest neighbors: $N(x) = \{y \in X : y \in \mathcal{N}_X^k(x) \wedge d(x,y) \leq \delta\}$
- Tentative correspondences: $C_t(X) = C_d(X) \cap C_v(X)$
- **Flipped matching:** y': flipped counterpart of feature y.

$$\begin{aligned} C_v(X,Y) &= \{(x,y) \in X \times Y : v(x,y')\} \\ C_d(X,Y) &= \{(x,y) \in X \times Y : y \in N(x)\} \\ C_t(X,Y) &= C_d(X,Y) \cap C_v(X,Y) \end{aligned}$$

Solution 1: Spatial self-matching (SSM)

- Inspired by fast spatial matching (FSM) [Philbin *et al.* 2007]
- Hypothesis inliers: $I_C(h) = \{(x,y) \in C : \|\mathbf{p}(y) - h\mathbf{p}(x)\| < \epsilon\}$
- Seek best hypothesis per correspondence
- $H_C(x,y) = \{h \in t(C) : \|\mathbf{p}(y) - h\mathbf{p}(x)\| < \epsilon\}$
- Strength: $\alpha_C(c) = \max\{|I_C(h)| : h \in H_C(c)\}$
- Verified correspondences: $\alpha(C) = \{c \in C : \alpha_C(c) \geq \tau_\alpha\}$
- Select features of verified correspondences
- **Average running time on SymCity:** 95**ms**

procedure $\alpha \leftarrow \mathrm{SSM}(C, t; \tau_\alpha)$
input : correspondences C, transformations t
parameter: inlier threshold τ_α
output : inlier strengths α

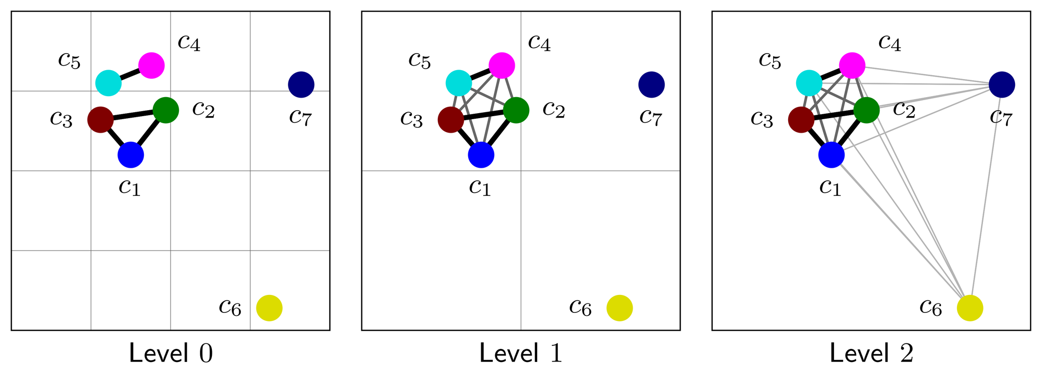
for $c \in C$ **do** ▷ initialize
  $inlier(c) \leftarrow$ FALSE ▷ mark as outlier
  $\alpha(c) \leftarrow 0$ ▷ zero strength
for $c \in C$ **do** ▷ for all hypotheses
  if $inlier(c)$ **then continue** ▷ skip hypothesis?
  $h \leftarrow t(c)$ ▷ current hypothesis
  $I \leftarrow I_C(h)$ ▷ current inliers (8)
  if $|I| < \tau_\alpha$ **then continue** ▷ verified hypothesis?
  for $c' \in I$ **do** ▷ for all inliers
    $inlier(c') \leftarrow$ TRUE ▷ mark as inlier
    $\alpha(c') \leftarrow \max(\alpha(c'), |I|)$ ▷ update strength
return α ▷ inlier strengths

Self-matching | Flipped matching

Selected features: Original (red), flipped (green) and back-projected (blue)

Solution 2: Hough pyramid self-matching (HPSM)

- Based on Hough pyramid matching [Tolias & Avrithis 2011]
- Same correspondences as in SSM but *linear* in the number of correspondences
- No inlier counting or transformation estimation
- Strength: geometrical consistency with all correspondences
- No one-to-one mapping as in original HPM
- **Average running time on SymCity:** 16.2**ms**

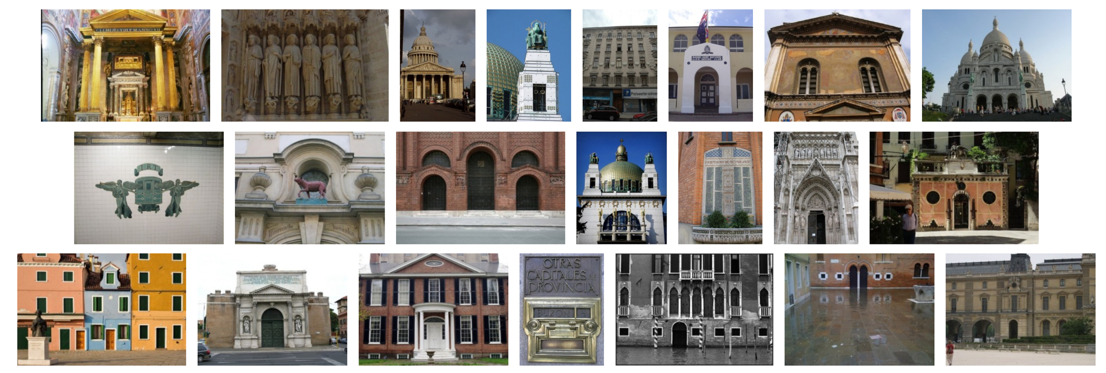


Correspondences in a single bin at level 0, reveal a symmetric feature group | All tentative correspondences, with red (yellow) being the strongest (weakest)

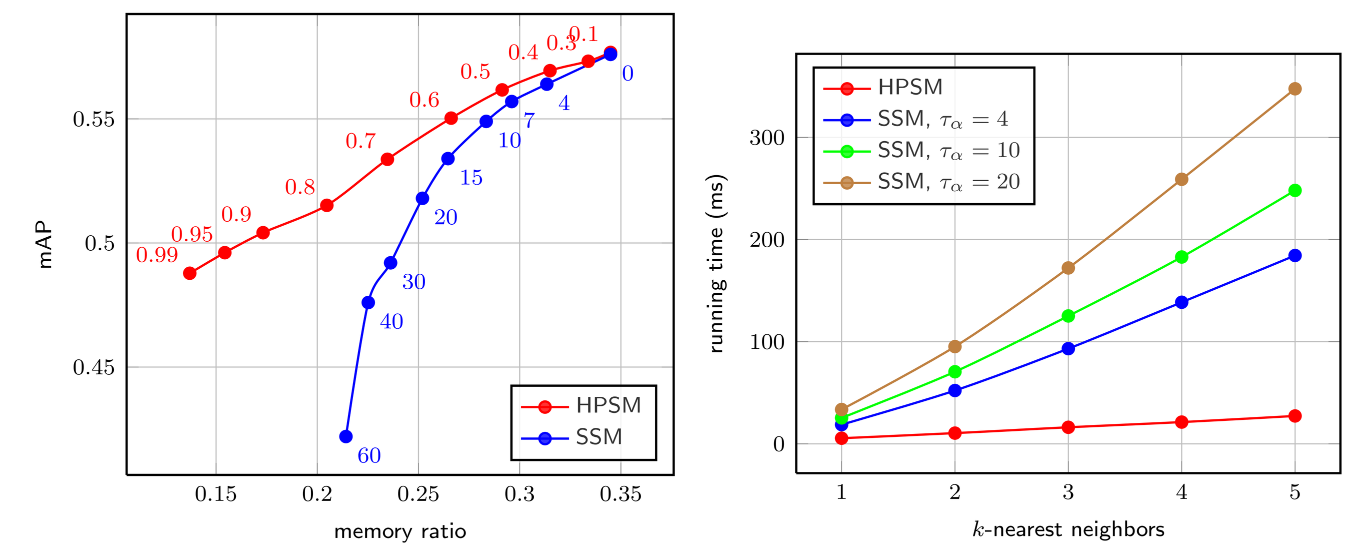
Selection examples

Experiments

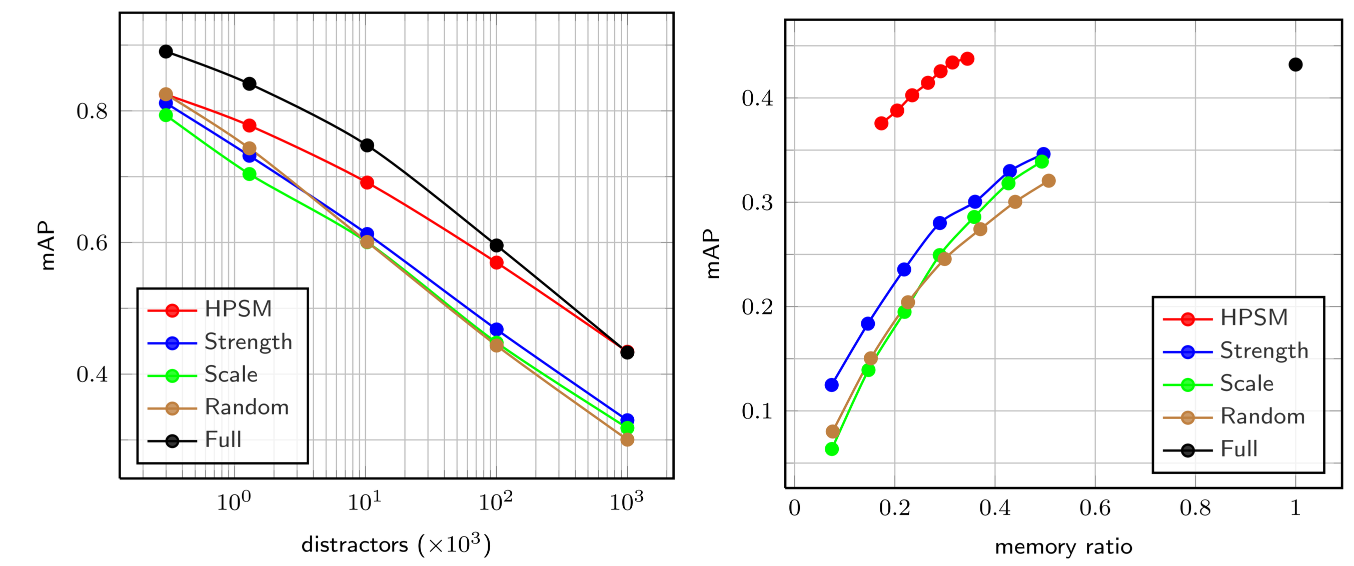
- **Datasets:** World Cities (WC) and new dataset SymCity
- **SymCity dataset:** 953 annotated photos from 299 groups; a single image from each group indexed in the database and the rest used as queries; publicly available



Sample images from the SymCity dataset



SSM vs HPSM using 100K distractors from WC



Large scale experiments using 1M distractors from WC