# To aggregate or not to aggregate: Selective match kernels for image search


Giorgos Tolias[1,2] and Yannis Avrithis[2] and Hervé Jégou[1]

[1]INRIA, Rennes [2]NTUA


## Overview

- Build common model for existing approaches and derive our methods
- Bridge the gap between matching based (HE) and aggregated based (VLAD) methods
- Evaluate with full precision descriptors and approximate representation
- Combine with query expansion using the same principle of aggregation

## Set similarity functions

### Image representation

- $\mathcal{X} = \{x_1, \ldots, x_n\}$ set of $n$ $d$-dimensional local descriptors
- Descriptors assigned to cell $c$: $\mathcal{X}_c = \{x \in \mathcal{X} : q(x) = c\}$

### Set similarity function

$\mathcal{K}(\mathcal{X}, \mathcal{Y}) = \gamma(\mathcal{X})\,\gamma(\mathcal{Y}) \sum_{c \in \mathcal{C}} w_c \, \mathrm{M}(\mathcal{X}_c, \mathcal{Y}_c)$

- $\mathrm{M}$: cell similarity function
- $w_c$: visual word weighting *e.g.* idf
- Normalization factor $\gamma(\mathcal{X}) = \left(\sum_{c \in \mathcal{C}} w_c \, \mathrm{M}(\mathcal{X}_c, \mathcal{X}_c)\right)^{-1/2}$
- Self-similarity $\mathcal{K}(\mathcal{X}, \mathcal{X}) = 1$

## Existing methods

### Bag-of-Words (BoW)

- BoW - cosine similarity

$$\mathrm{M}(\mathcal{X}_c, \mathcal{Y}_c) = \#\mathcal{X}_c \times \#\mathcal{Y}_c = \sum_{x \in \mathcal{X}_c} \sum_{y \in \mathcal{Y}_c} 1 \qquad (1)$$

- Similarly with histogram intersection or max pooling

### Hamming Embedding (HE)

- Descriptor representation: visual word $q(x)$ - binary code $b_x$ of $B$ bits

$$\mathrm{M}(\mathcal{X}_c, \mathcal{Y}_c) = \sum_{x \in \mathcal{X}_c} \sum_{y \in \mathcal{Y}_c} w\left(\mathrm{h}(b_x, b_y)\right) \qquad (2)$$

- $\mathrm{h}$: Hamming distance
- $w$: weighting function $w(h) = e^{-h^2/\sigma^2}, h \leq \tau, \ 0,$ otherwise

### VLAD

- $V(\mathcal{X}_c) = \sum_{x \in \mathcal{X}_c} r(x)$, where $r(x) = x - q(x)$: residual of $x$
- Concatenation $\mathcal{V}(\mathcal{X}) \propto [V(\mathcal{X}_{c_1}), \ldots, V(\mathcal{X}_{c_k})]$ of $d$-dimensional vectors
- VLAD similarity: $\mathcal{V}(\mathcal{X})^\top \mathcal{V}(\mathcal{Y}) = \gamma(\mathcal{X})\,\gamma(\mathcal{Y}) \sum_{c \in \mathcal{C}} V(\mathcal{X}_c)^\top V(\mathcal{Y}_c)$

$$\mathrm{M}(\mathcal{X}_c, \mathcal{Y}_c) = V(\mathcal{X}_c)^\top V(\mathcal{Y}_c) = \sum_{x \in \mathcal{X}_c} \sum_{y \in \mathcal{Y}_c} r(x)^\top r(y) \qquad (3)$$

## Methods in the common model

| Model | $\mathrm{M}(\mathcal{X}_c, \mathcal{Y}_c)$ | $\phi(x)$ | $\sigma(u)$ | $\psi(z)$ | $\Phi(\mathcal{X}_c)$ |
|---|---|---|---|---|---|
| BoW (1) | $\mathrm{M_N}$ or $\mathrm{M_A}$ | 1 | $u$ | $z$ | $\#\mathcal{X}_c$ |
| HE (2) | $\mathrm{M_N}$ | $\hat{b}_x$ | $w\left(\frac{B}{2}(1-u)\right)$ | — | — |
| VLAD (3) | $\mathrm{M_N}$ or $\mathrm{M_A}$ | $r(x)$ | $u$ | $z$ | $V(\mathcal{X}_c)$ |
| SMK (4) | $\mathrm{M_N}$ | $\hat{r}(x)$ | $\sigma_\alpha(u)$ | — | — |
| ASMK (5) | $\mathrm{M_A}$ | $r(x)$ | $\sigma_\alpha(u)$ | $\hat{z}$ | $\hat{V}(\mathcal{X}_c)$ |
| SMK* (6) | $\mathrm{M_N}$ | $\hat{b}_x$ | $\sigma_\alpha(u)$ | — | — |
| ASMK* (7) | $\mathrm{M_A}$ | $r(x)$ | $\sigma_\alpha(u)$ | $\hat{b}(z)$ | $\hat{b}(V(\mathcal{X}_c))$ |

## Common model

### Non aggregated

$$\mathrm{M_N}(\mathcal{X}_c, \mathcal{Y}_c) = \sum_{x \in \mathcal{X}_c} \sum_{y \in \mathcal{Y}_c} \sigma\left(\phi(x)^\top \phi(y)\right)$$

- $\phi$: Descriptor representation (residual, binary, scalar)
- $\sigma$: Selectivity function (post-processing of similarity score)

### Aggregated

$$\mathrm{M_A}(\mathcal{X}_c, \mathcal{Y}_c) = \sigma\left\{\psi\left(\sum_{x \in \mathcal{X}_c} \phi(x)\right)^\top \psi\left(\sum_{y \in \mathcal{Y}_c} \phi(y)\right)\right\} = \sigma\left(\Phi(\mathcal{X}_c)^\top \Phi(\mathcal{Y}_c)\right)$$

- $\psi$: Post-processing of aggregated representation ($\ell_2$-normalization, power-law)
- $\Phi(\mathcal{X}_c)$: Aggregated representation of descriptors in cell $c$

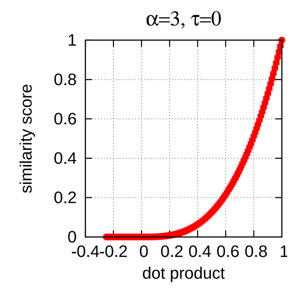
## Our methods

### Selective Match Kernel (SMK)

$$\mathrm{SMK}(\mathcal{X}_c, \mathcal{Y}_c) = \sum_{x \in \mathcal{X}_c} \sum_{y \in \mathcal{Y}_c} \sigma_\alpha\left(\hat{r}(x)^\top \hat{r}(y)\right) \qquad (4)$$

- Selectivity function $\boldsymbol{\sigma}$: Thresholded polynomial of the form

$$\sigma_\alpha(u) = \begin{cases} \mathrm{sign}(u)|u|^\alpha & \text{if } u > \tau \\ 0 & \text{otherwise} \end{cases}$$

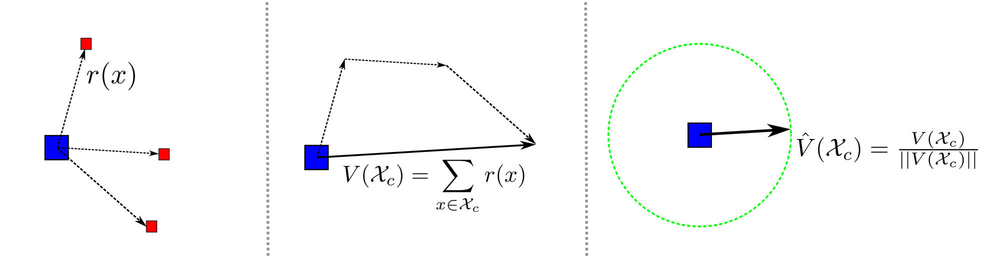



- $\boldsymbol{\phi}$: $\ell_2$-normalized residual: $\hat{r}(x) = r(x)/\|r(x)\|$

### Aggregated Selective Match Kernel (ASMK)

$$\mathrm{ASMK}(\mathcal{X}_c, \mathcal{Y}_c) = \sigma_\alpha\left(\hat{V}(\mathcal{X}_c)^\top \hat{V}(\mathcal{Y}_c)\right) \qquad (5)$$

- Aggregate residuals and $\ell_2$-normalize: $\Phi(\mathcal{X}_c) = \hat{V}(\mathcal{X}_c) = V(\mathcal{X}_c)/\|V(\mathcal{X}_c)\|$




- Selectivity function $\sigma_\alpha$ on single matches of aggregated representations

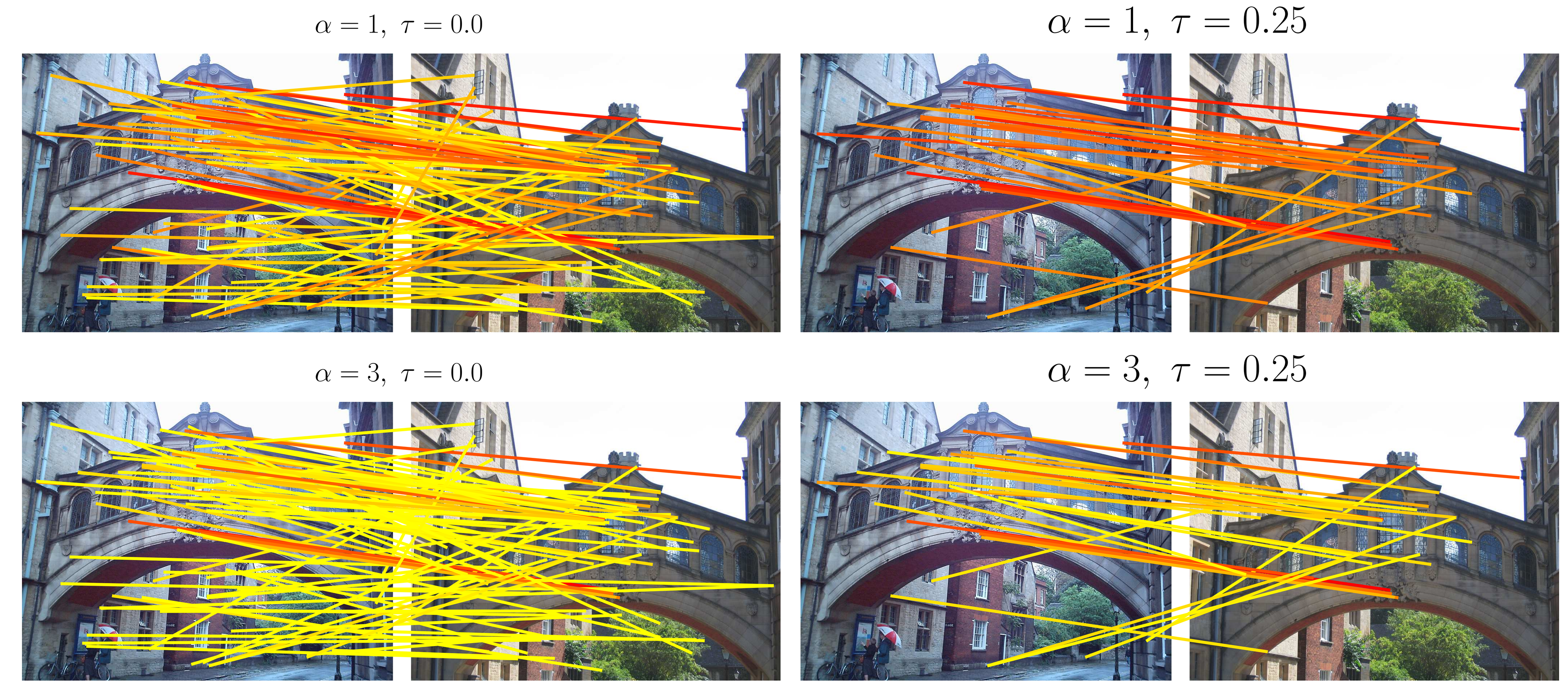
## Binarized counterparts

$$\mathrm{SMK}^\star(\mathcal{X}_c, \mathcal{Y}_c) = \sum_{x \in \mathcal{X}_c} \sum_{y \in \mathcal{Y}_c} \sigma_\alpha\left(\hat{b}_x^\top \hat{b}_y\right) \qquad (6)$$

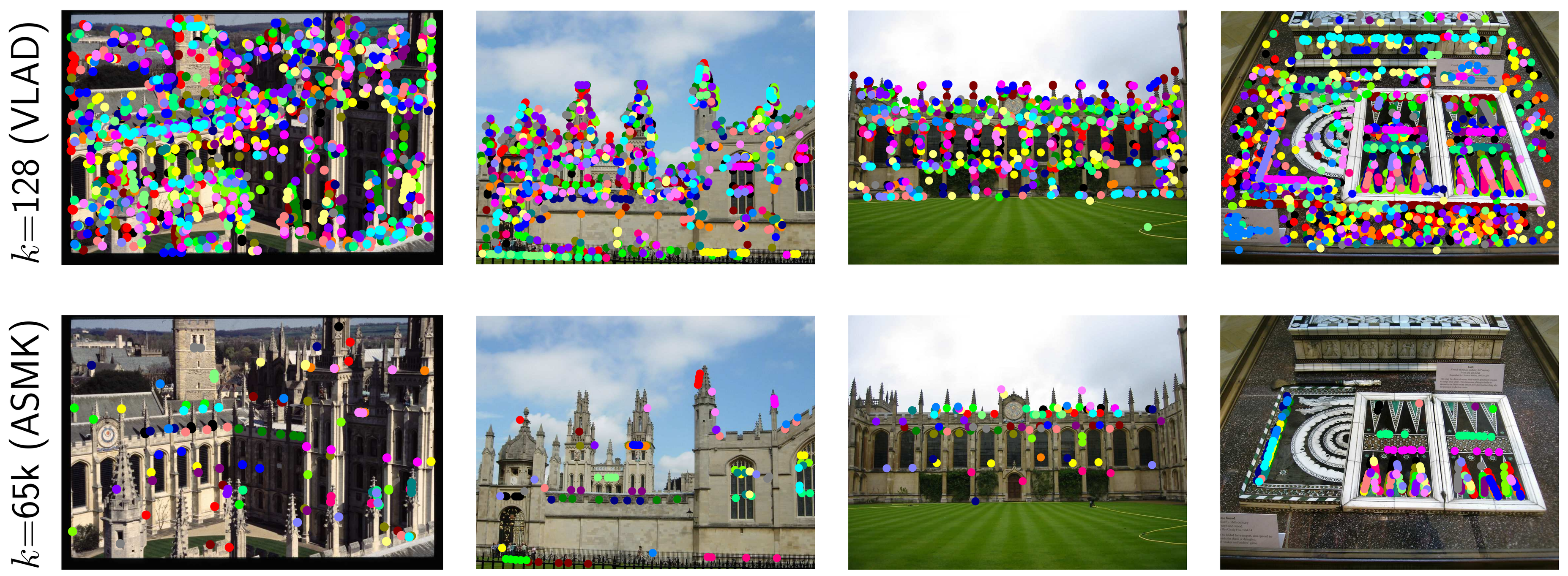
$$\mathrm{ASMK}^\star(\mathcal{X}_c, \mathcal{Y}_c) = \sigma_\alpha\left\{\hat{b}\left(\sum_{x \in \mathcal{X}_c} r(x)\right)^\top \hat{b}\left(\sum_{y \in \mathcal{Y}_c} r(y)\right)\right\} \qquad (7)$$

- $\hat{b}$: element-wise binarization function

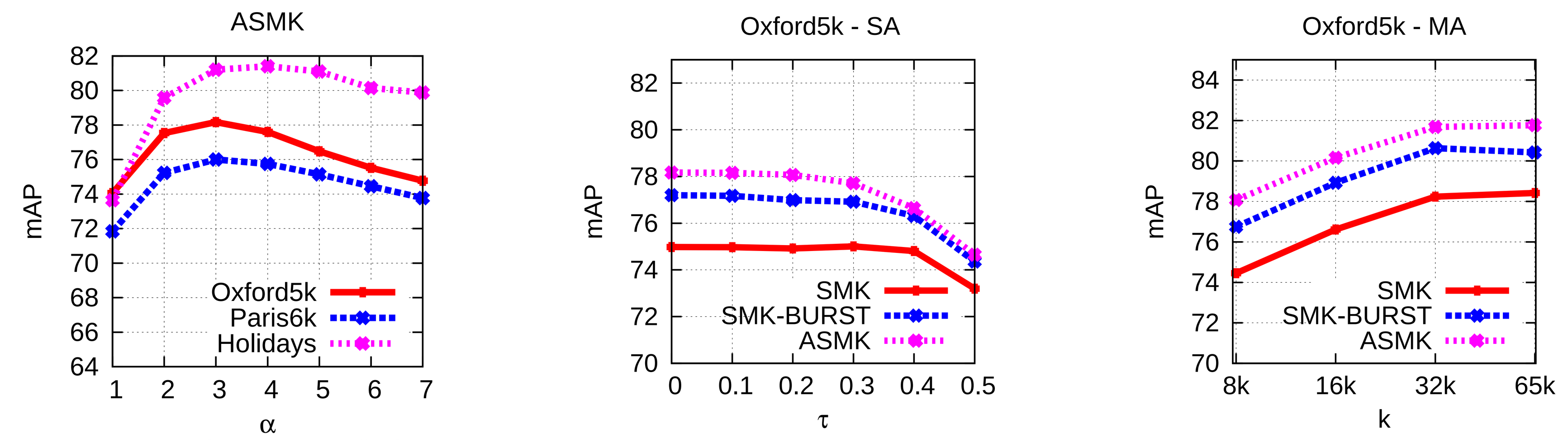
## Matching example




## Aggregation example




## Experiments




### Comparison with state of the art

| Dataset | MA | Oxf5k | Oxf105k | Par6k | Holidays |
|---|---|---|---|---|---|
| ASMK* | | 76.4 | 69.2 | 74.4 | 80.0 |
| ASMK* | × | 80.4 | **75.0** | 77.0 | 81.0 |
| ASMK | | 78.1 | - | 76.0 | 81.2 |
| ASMK | × | **81.7** | - | **78.2** | **82.2** |
| HE [Jégou *et al.* 10] | | 51.7 | - | - | 74.5 |
| HE [Jégou *et al.* 10] | × | 56.1 | - | - | 77.5 |
| HE-BURST [Jain *et al.* 10] | | 64.5 | - | - | 78.0 |
| HE-BURST [Jain *et al.* 10] | × | 67.4 | - | - | 79.6 |
| Fine vocabulary [Mikulík *et al.* 10] | × | 74.2 | 67.4 | 74.9 | 74.9 |
| AHE-BURST [Jain *et al.* 10] | | 66.6 | - | - | 79.4 |
| AHE-BURST [Jain *et al.* 10] | × | 69.8 | - | - | 81.9 |
| Rep. structures [Torri *et al.* 13] | × | 65.6 | - | - | 74.9 |

- Combined with query expansion
  - ASMK: 87.9 on Oxford5k (Fine vocabulary+QE: 84.9)
  - ASMK*: 85.0 on Oxford105k (Fine vocabulary+QE: 79.5)

### Memory ratio after-before aggregation

| $k$ | 8k | 16k | 32k | 65k |
|---|---|---|---|---|
| Oxf | 69 % | 78 % | 85 % | 89 % |
| Par | 68 % | 76 % | 82 % | 86 % |
| Hol | 55 % | 65 % | 73 % | 78 % |

- Aggregation reduces memory requirements and improves performance in all cases
- Aggregation handles burstiness in this context (large vocabularies)

Public code available for ASMK*


Contact: {gtolias, iavr}@image.ntua.gr, herve.jegou@inria.fr