Early burst detection for memory-efficient image retrieval

Miaojing Shi[1,3], Yannis Avrithis[2] and Hervé Jégou[3]

[1]Peking University [2]Univeristy of Athens, NTUA [3]INRIA, Rennes

Overview

- Explicitly detect visual bursts in an image at an early stage
- Aggregate bursty groups into meta-descriptors on database side
- Asymmetric scheme: do not aggregate on query side
- On par with state of the art, yet at much lower memory and query time

Representation and matching

Image representation

- $\mathcal{X}$: set of d-dimensional local descriptors per image
- $\mathcal{C}$: codebook of d-dimensional visual words or *cells*
- $\mathcal{X}_c$: descriptors assigned to cell $c \in \mathcal{C}$

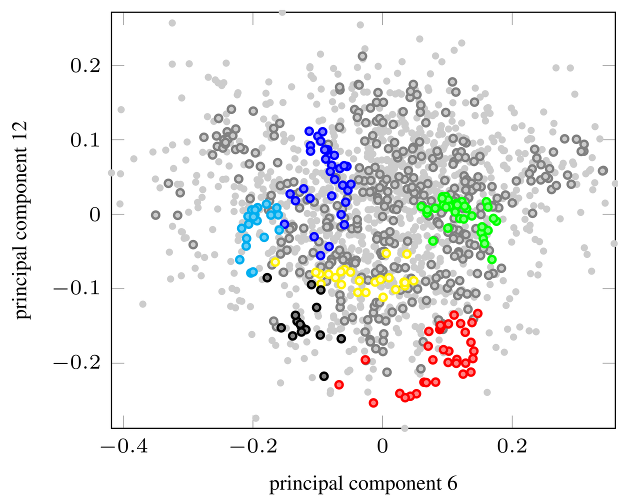
Similarity function

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

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

What are visual bursts?

(a) Six most populated bursts

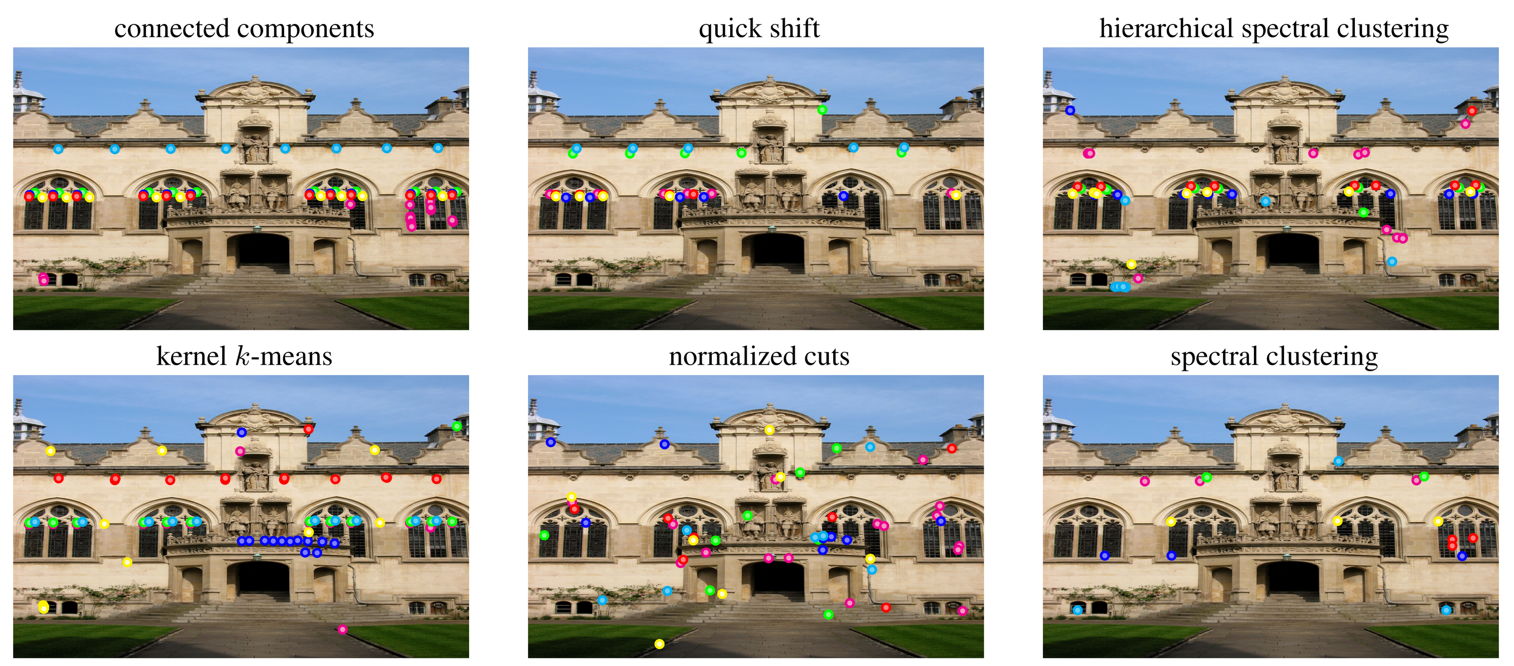


(b) Feature distribution after PCA analysis

Observation

- Not like text bursts: the descriptor space is continuous
- Bursts have arbitrary shape and overlap: unlikely to fit within codebook cells
- Possible sources: to structure in man-made scenes, texture in natural environments, multiple feature detector responses

Burst detection example



Feature kernel

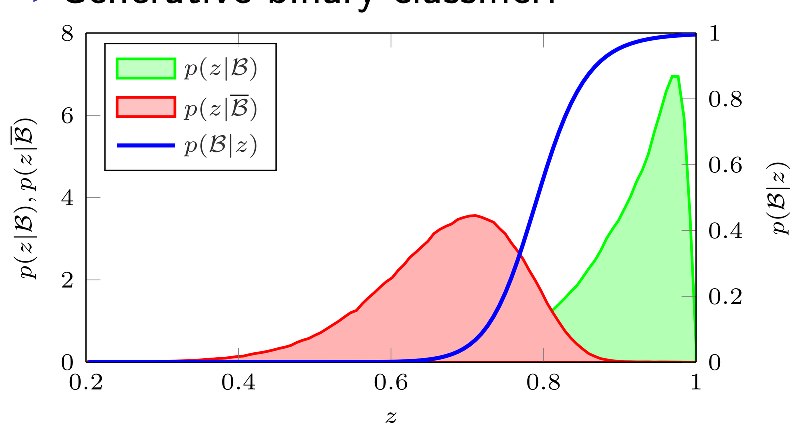
Feature kernel k

$$k(f,g) = k_u(u_f,u_g)\,k_s(s_f,s_g)\,k_\theta(\theta_f,\theta_g)$$

- f, g: local image features
- u_f: local descriptor; s_f: scale; θ_f: orientation

Descriptor kernel k_u

- Generative binary classifier:



$$k_u(x,y) = p(\mathcal{B}|\langle x,y\rangle)$$

$$p(\mathcal{B}|z) = \frac{p(z|\mathcal{B})p(\mathcal{B})}{p(z|\mathcal{B})p(\mathcal{B}) + p(z|\overline{\mathcal{B}})p(\overline{\mathcal{B}})}$$

- $\mathcal{B}$: pairs in the same burst
- $\overline{\mathcal{B}}$ pairs not in the same burst

Scale kernel k_s

- Gaussian:

$$k_s(s,t) = \exp\left\{-\lambda \log^2\left(\frac{s}{t}\right)\right\}$$

Orientation kernel k_θ

- von Mises:

$$k_\theta(\theta,\phi) = \frac{e^{\kappa\cos(\theta-\phi)} - e^{-\kappa}}{2\sinh(\kappa)}$$

Burst detection and aggregation

Representation

- $\mathcal{F} = \{f_1,\ldots,f_n\}$: set of n local image features
- K: $n\times n$ affinity matrix with $K_{ij} = k(f_i,f_j)$

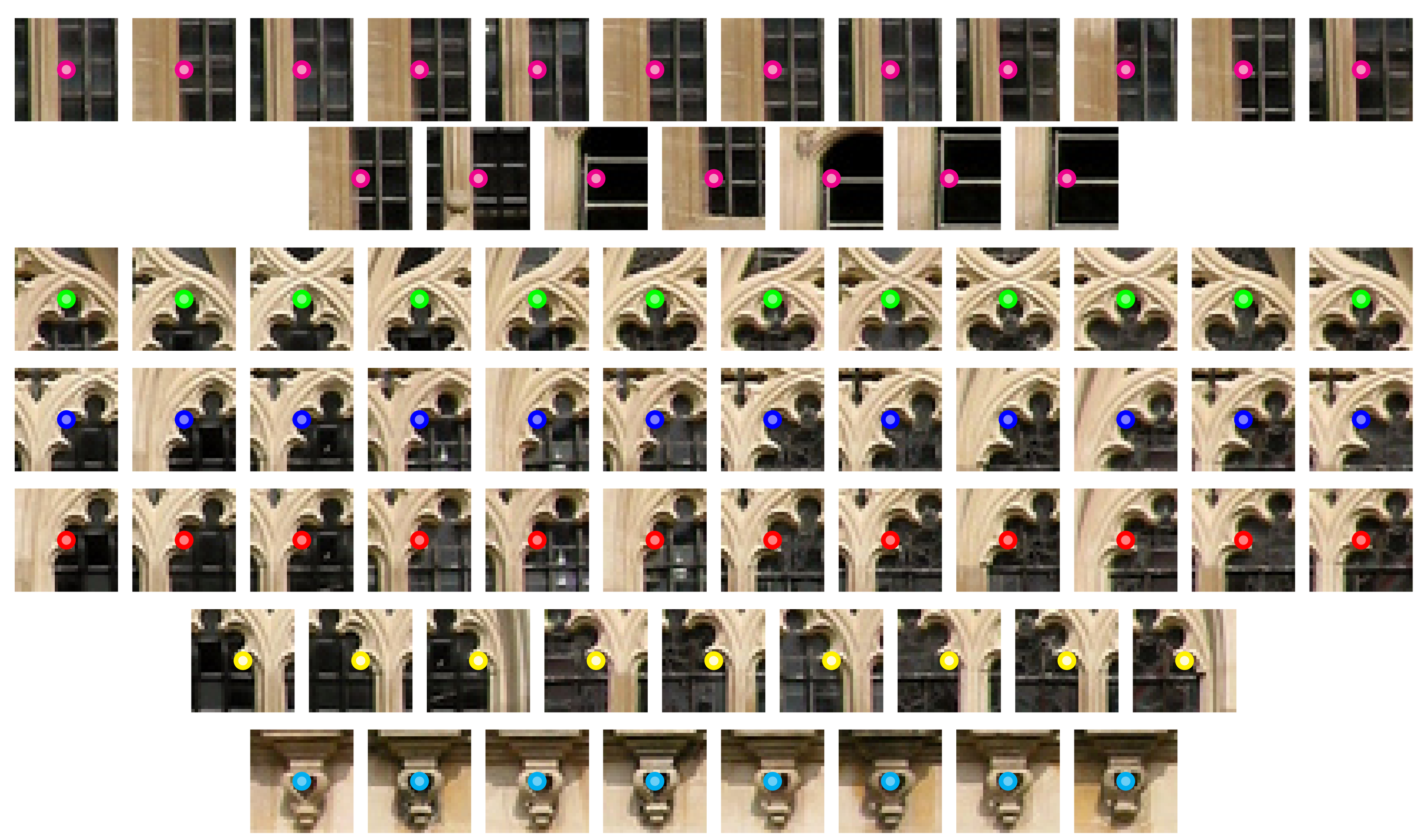
Burst detection

- Based on a kernel method, or operate on metric spaces
- Able to automatically determine (or to control with a parameter) the number of groups, such that non-matching features are not grouped
- The clear winner amongst all tested methods is connected component analysis

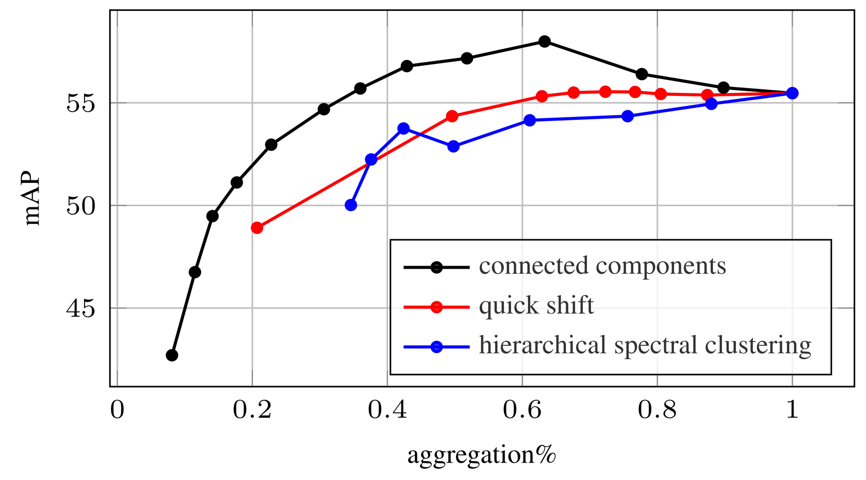
Burst aggregation

- Take the average of the descriptors in each bursty group and ℓ_2-normalize.

Burst detection using connected components



Experiments on VLAD

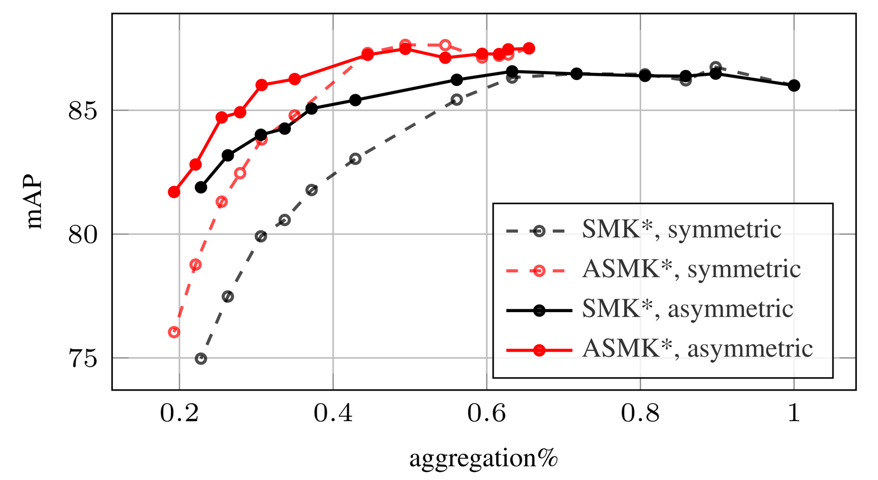


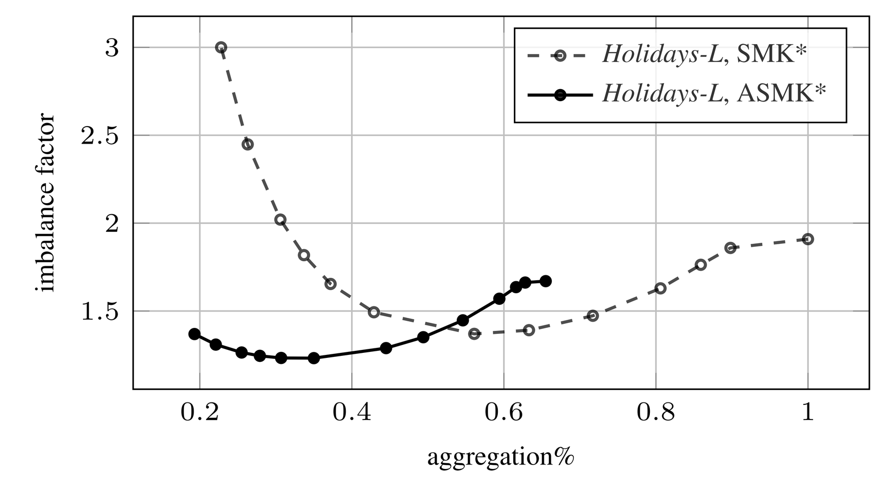
Performance on *Holidays-L* + *Flickr* 100k

aggregation%	1.000	0.764	0.638	0.556
$k=16$	41.3	42.7	44.1	45.0
$k=64$	46.3	47.5	48.3	48.8

- *Holidays-L*: Holidays dataset, 6.6M features
- aggregation%: ratio of aggregated to original features averaged over the dataset

Experiments on SMK*/ASMK*





Performance on *Holidays-L*, *Oxford* + *Flickr* 100k

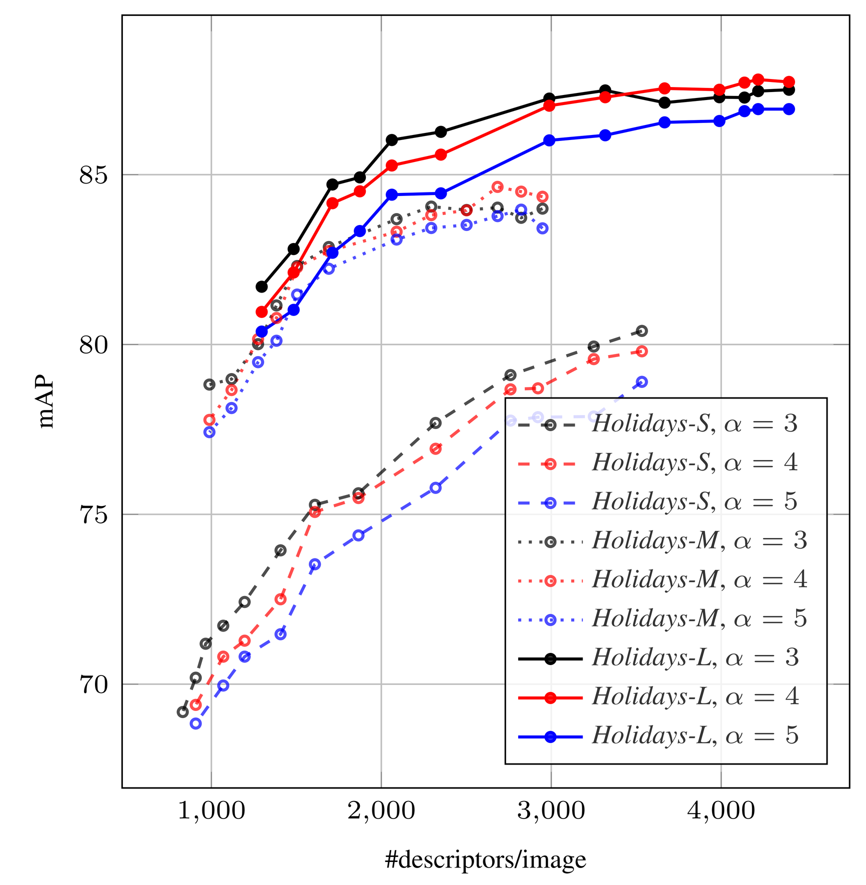
Dataset	*Holidays-L* 101k			*Oxford* 105k		
aggregation%	0.65	0.52	0.28	0.90	0.76	0.55
mAP	85.1	84.5	77.6	68.9	68.9	63.6

- Vocabulary size: 65k; first column is the baseline

Comparison to state of the art

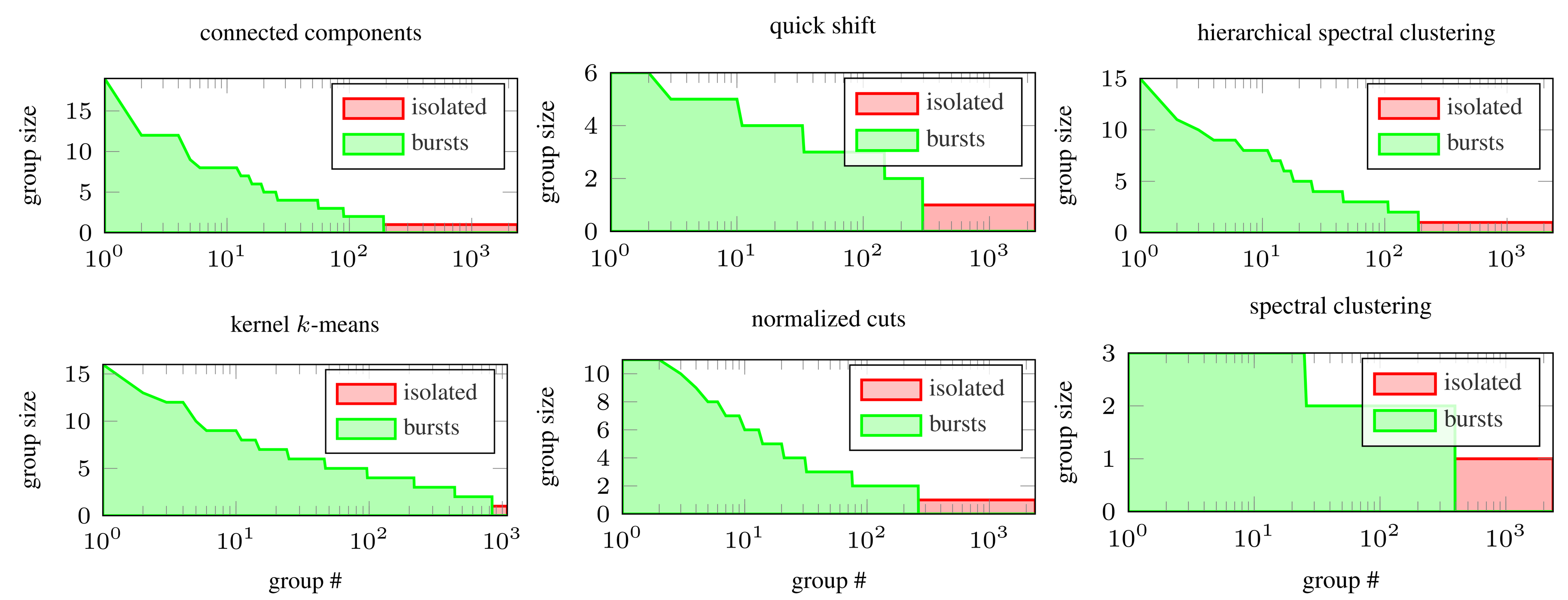
Dataset	MA	Hol.	Paris	Oxf.
BoW [27]		-	-	40.3
BoW [27]	✓	-	-	49.3
BoW [24]		-	-	55.8
Fine vocab. [22]		74.9	74.9	74.2
Multi-index [3]	✓	-	69.6	70.3
HE [15]		74.5	-	51.7
HE [15]	✓	77.5	-	56.1
AHE+burst [12]		79.4	-	66.0
AHE+burst [12]	✓	81.9	-	69.8
Query ad. [30]		81.4	70.3	73.9
Query ad. [30]	✓	82.1	73.6	78.0
aggregation%		78%	86%	89%
ASMK* [37]		80.0	74.4	76.4
ASMK* [37]	✓	81.0	77.0	80.4
This work	✓	**88.1**	**77.5**	**81.3**

Memory efficiency performance on *Holidays*



- *Holidays-S*: 4.4M features; *Holidays-M*: 4.3M features

Size distribution of burst groups



This work was supported by ERC grant VIAMASS no. 336054

Contact: miao0913@gmail.com, iavr@image.ntua.gr, rvj@fb.com