Asymmetric Feature Maps with Application to Sketch Based Retrieval

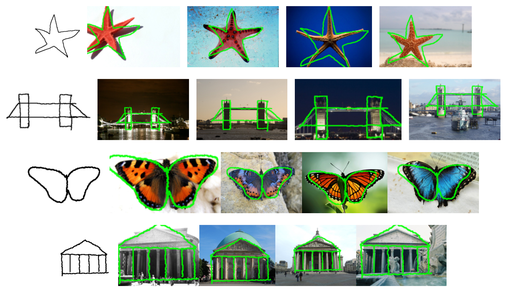
Giorgos Tolias and Ondřej Chum
Visual Recognition Group, CTU in Prague

CVPR July 21-26 2017 HONOLULU

CTU CZECH TECHNICAL UNIVERSITY IN PRAGUE

Problem formulation

- Large image collection & query in sketch form
- Retrieve images with shapes similar to the sketch
- Localize the matching object
- We achieve this with a single vector representation



Kernel descriptors

- Sketch $\mathcal{Q}$: binary edge map
- Image $\mathcal{P}$: edge detection to obtain edge map
- Edge map: set of edge pixels
- Pixel attributes: position, edge orientation, edge magnitude

Sketch (edge map) similarity

- Match kernel comparing all edge pixel pairs
- Stationary kernels to compare location and edge orientation
- Weigh by edge strength

$$\mathcal{S}(\mathcal{P},\mathcal{Q}) = \sum_{p\in\mathcal{P}}\sum_{q\in\mathcal{Q}} p_w q_w k_x(\lambda_x) k_y(\lambda_y) k_\phi(\lambda_\phi)$$

Match kernel evaluation is costly (enumerate all pairs)!

Employ approximation of stationary kernels

$$\hat{k}(\lambda) \approx \sum_{\omega\in\Omega} \alpha_\omega \cos(\omega\lambda), \text{ with } \alpha_\omega \in \mathbb{R}_0^+$$

giving rise to feature maps for p and q, with $\lambda = p - q$

$$\Psi_\omega(p) = \left(\sqrt{\alpha_\omega}\cos(\omega p),\ \sqrt{\alpha_\omega}\sin(\omega p)\right)^\top$$
$$\Psi_\omega(q) = \left(\sqrt{\alpha_\omega}\cos(\omega q),\ \sqrt{\alpha_\omega}\sin(\omega q)\right)^\top$$

and the sketch (kernel) descriptor

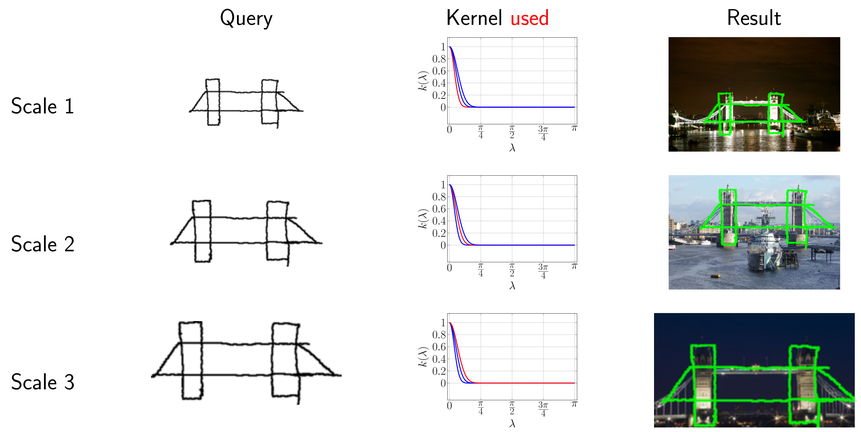
$$\mathbf{V}(\mathcal{P}) = \sum_{p\in\mathcal{P}} p_w \Psi(p_x) \otimes \Psi(p_y) \otimes \Psi(p_\phi)$$

where Ψ comprises Ψ_ω for $\omega \in \Omega$. Frequency set Ω can consist of:

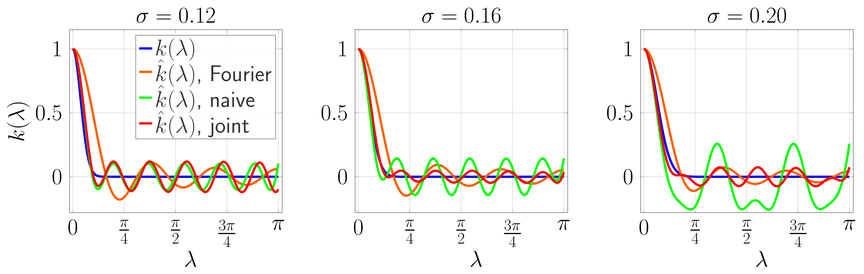
- Harmonic frequencies (fixed for all kernels) [3]
- Learned frequencies (better approximation) LDFM [1]

Asymmetric feature maps (AFM) for multi-scale matching

Different spatial kernel is required for different scales



Linear program to learn weights and joint frequencies for multiple kernels: our extension to LDFM [1]



Asymmetric feature maps → Evaluate different kernels with the same database vector

$$\hat{\Psi}_\omega(q) = \left(\alpha_\omega\cos(\omega q),\ \alpha_\omega\sin(\omega q)\right)^\top \quad \text{Query side embedding}$$
$$\hat{\Psi}'_\omega(p) = \left(\cos(\omega p),\ \sin(\omega p)\right)^\top \quad \text{Database side embedding}$$

Efficient contour matching and localization

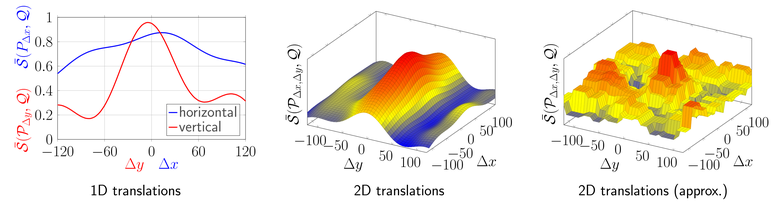
Efficient similarity computation over multiple translations

$$\mathcal{S}(\mathcal{P}_{\Delta x},\mathcal{Q}) = \sum_{\omega\in\Omega_x} \left(\beta_\omega\cos(\omega\Delta x) + \gamma_\omega\sin(\omega\Delta x)\right),$$

with coefficients β_ω and γ_ω given by sub-vector inner products

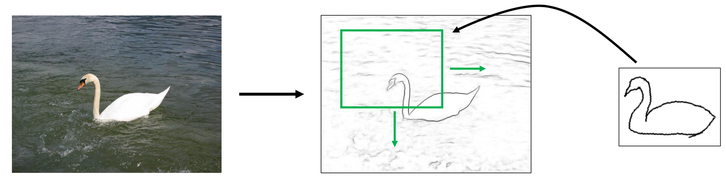
$$\beta_\omega = \mathbf{V}^c_\omega(\mathcal{P})^\top\mathbf{V}^c_\omega(\mathcal{Q}) + \mathbf{V}^s_\omega(\mathcal{P})^\top\mathbf{V}^s_\omega(\mathcal{Q})$$
$$\gamma_\omega = \mathbf{V}^s_\omega(\mathcal{P})^\top\mathbf{V}^c_\omega(\mathcal{Q}) - \mathbf{V}^c_\omega(\mathcal{P})^\top\mathbf{V}^s_\omega(\mathcal{Q})$$

Localization by composition of 1D translation search or direct 2D search

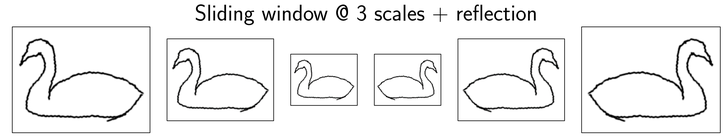


1D translations | 2D translations | 2D translations (approx.)

Sketch retrieval as exhaustive contour matching



Sliding window @ 3 scales + reflection

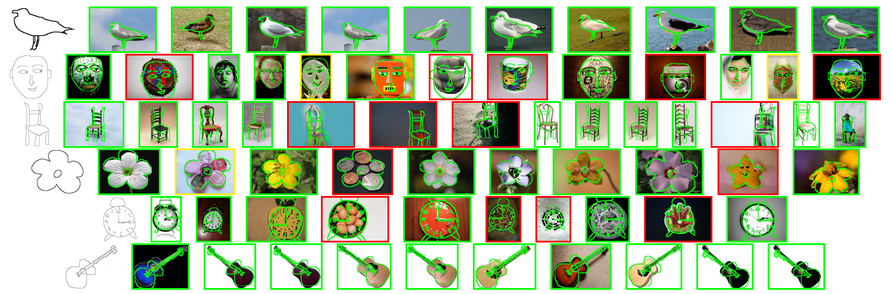


Best performance among hand-crafted descriptors

Achievements:

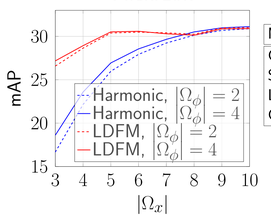
- Needs to store 300-600 bytes per image
- Working @ 1.2M images in 2 sec on a single code (including reflections)

Examples of retrieval and localization (search in 1.2M images)



Experiments on sketch-base image retrieval

Flickr15k



mAP on Flickr15k

Method	mAP	Method	mAP
GF-HOG	12.2	$\bar{\mathcal{S}}_x + \bar{\mathcal{S}}_y \rightarrow \bar{\mathcal{S}}_{xy}$ (1k)	26.7
SHELO	12.3	$\bar{\mathcal{S}}_x + \bar{\mathcal{S}}_y \rightarrow \bar{\mathcal{S}}_{xy}$ (5k)	29.2
LKS	24.5	$\bar{\mathcal{S}}_{xy}$	30.4
GF-HOG	18.2	$\bar{\mathcal{S}}_{xy}$ + QE3	**57.9**

Performance on 1.2M images

Method	Dim	Time	DB	P@5	@10	@25
$\bar{\mathcal{S}}_{xy}$ (1.2M)AFM [2]	(8,3)	55.4	15.3	43.2	40.9	37.2
$\bar{\mathcal{S}}_{xy}$ (1.2M)AFM [2]	(5,2)	20.2	3.3	25.8	24.7	22.5
$\bar{\mathcal{S}}_{xy}$ (1.2M)	(8,3)	55.4	5.1	50.1	46.7	42.0
$\bar{\mathcal{S}}_{xy}$ (1.2M)	(5,2)	20.2	1.1	45.8	44.1	38.5
$\bar{\mathcal{S}}_x + \bar{\mathcal{S}}_y \rightarrow \bar{\mathcal{S}}_{xy*}$ (50k)	(6,3)	3.5	2.8	49.7	47.4	41.3
$\bar{\mathcal{S}}_> \rightarrow \bar{\mathcal{S}}_< \rightarrow \bar{\mathcal{S}}_{xy*}$ (50k)	(6,3)	2.5	2.8	49.6	47.3	41.0
$\bar{\mathcal{S}}_> \rightarrow \bar{\mathcal{S}}_< \rightarrow \bar{\mathcal{S}}_{xy*}$ (50k)†	(6,3)	2.5	0.7	50.3	47.3	41.5
$\bar{\mathcal{S}}_x + \bar{\mathcal{S}}_y \rightarrow \bar{\mathcal{S}}_{xy*}$ (50k)	(5,2)	2.5	1.1	45.8	44.2	38.4
$\bar{\mathcal{S}}_> \rightarrow \bar{\mathcal{S}}_< \rightarrow \bar{\mathcal{S}}_{xy*}$ (50k)	(5,2)	1.7	1.1	45.7	44.2	38.3
$\bar{\mathcal{S}}_> \rightarrow \bar{\mathcal{S}}_< \rightarrow \bar{\mathcal{S}}_{xy*}$ (50k)†	(5,2)	1.7	0.3	45.6	43.5	38.0

- Compact setting $|\Omega_x| = 5$, $|\Omega_\phi| = 2$: 243D image/sketch descriptor
- High performance setting $|\Omega_x| = 6$, $|\Omega_\phi| = 3$: 605D image/sketch descriptor

References:

[1] O. Chum. Low dimensional explicit feature maps. In *ICCV*, 2015.
[2] G. Tolias, A. Bursuc, T. Furon, and H. Jégou. Rotation and translation covariant match kernels for image retrieval. *CVIU*, 140:9–20, 2015.
[3] A. Vedaldi and A. Zisserman. Efficient additive kernels via explicit feature maps. *Trans. PAMI*, 34(3):480–492, Mar. 2012.

Contact: {giorgos.tolias, chum}@cmp.felk.cvut.cz