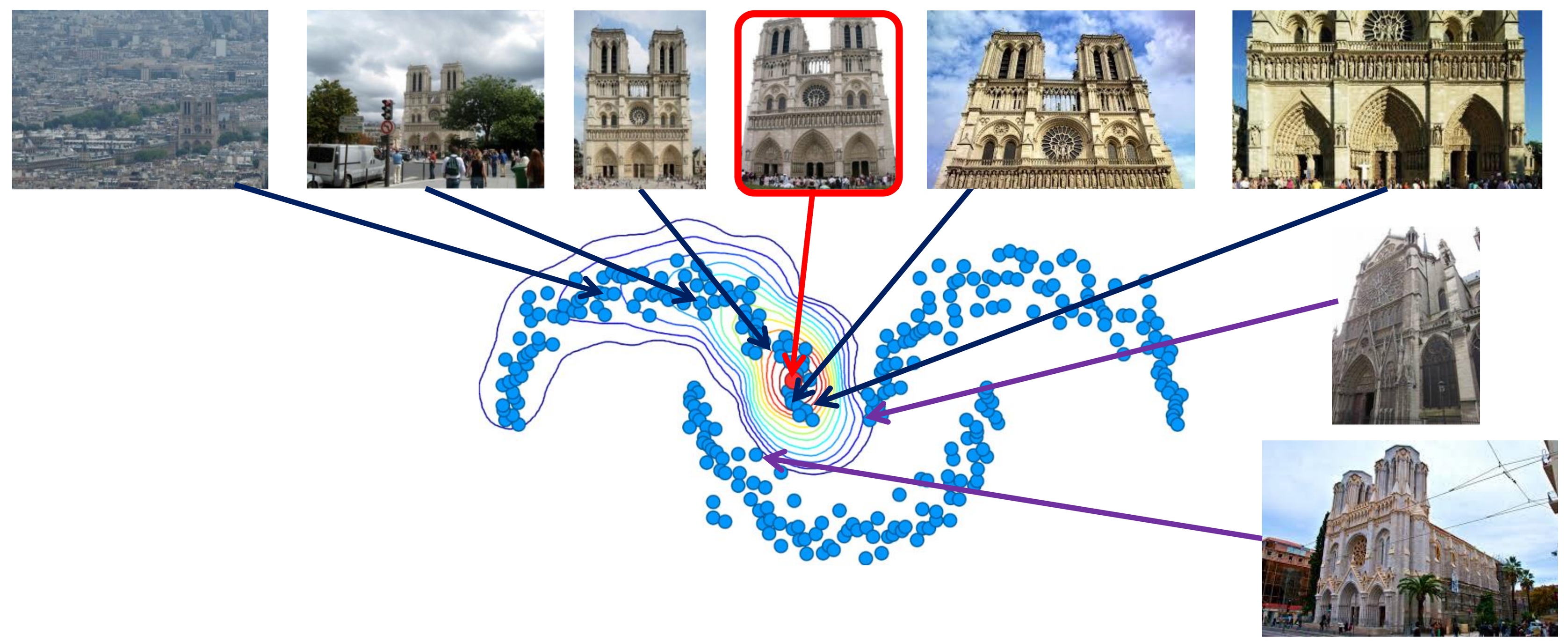
Fast Spectral Ranking for Similarity Search

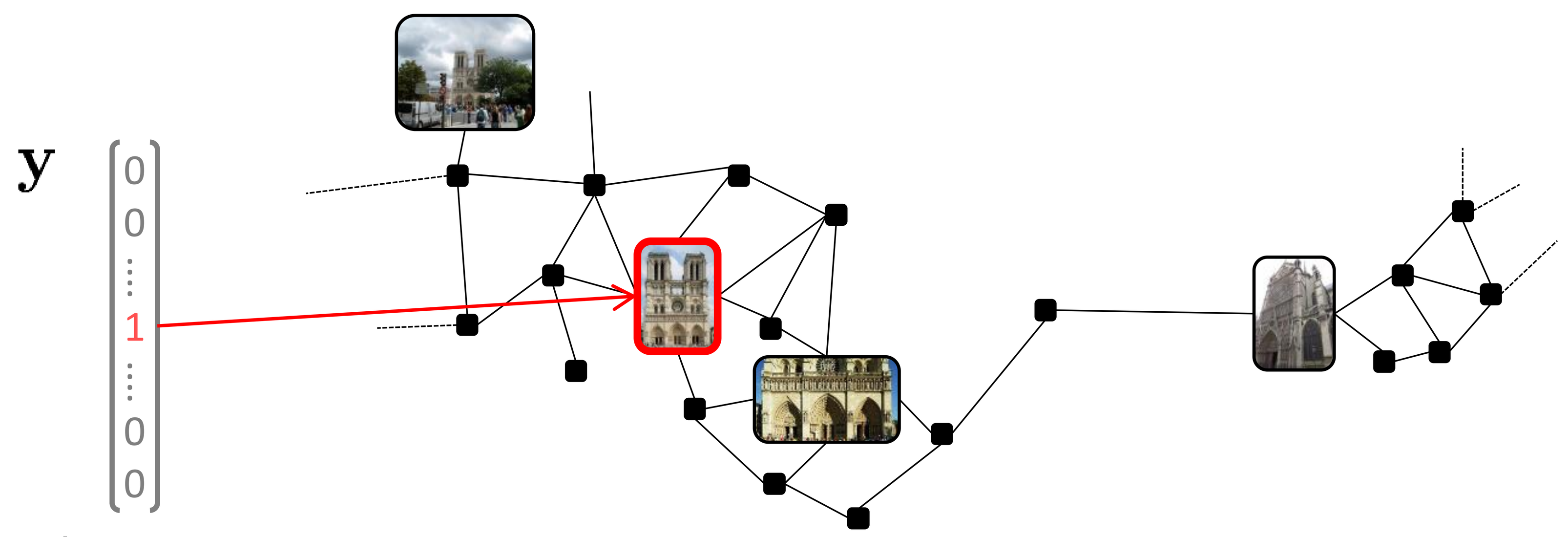
Ahmet Iscen[1], Yannis Avrithis[2], Giorgos Tolias[1], Teddy Furon[2], Ondřej Chum[1]
[1]Visual Recognition Group, CTU in Prague, [2]Inria

Manifold Search



- Euclidean distance is only **locally** a good metric
- Related images form non-linear manifolds
- Manifold search retrieves them but is expensive [2]
- We dramatically reduce the online (query time) cost

Background



Query indicator vector

- Database represented by adjacency matrix $W \in \mathbb{R}^{n\times n}$, symmetrically normalized as $\mathcal{W} := D^{-1/2} W D^{-1/2}$, where $D := \operatorname{diag}(W\mathbf{1})$
- Regularized Laplacian given by $\mathcal{L}_\alpha := (I - \alpha\mathcal{W})/(1-\alpha)$, where $\alpha \in [0,1)$
- Query: $n \times 1$ *observation vector* $\mathbf{y}$, determined by k-NN of query descriptor
- Search [2]: $n \times 1$ *similarity vector* $\mathbf{x}$ by conjugate gradient method

$$\mathcal{L}_\alpha \mathbf{x} = \mathbf{y}$$

Accelerating manifold search

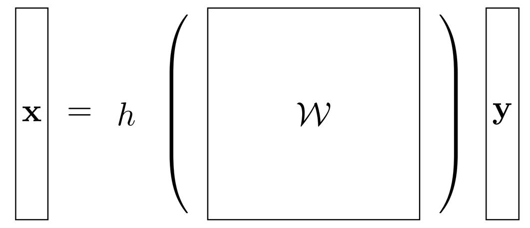
- Reduce manifold search to dot product search:

$$\mathbf{x} = \mathcal{L}_\alpha^{-1}\mathbf{y}$$

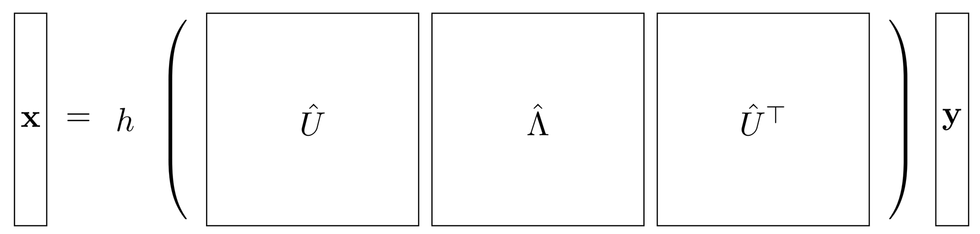
- Challenge: $\mathcal{L}_\alpha$ is sparse, but its inverse is not!
- Use a low-rank approximation of $\mathcal{L}_\alpha^{-1}$ to compute $\mathbf{x}$ without ever computing $\mathcal{L}_\alpha^{-1}$
- Formulate as $\mathbf{x} = h_\alpha(\mathcal{W})\mathbf{y}$, where *transfer function*

$$h_\alpha(\mathcal{W}) := (1-\alpha)(I - \alpha\mathcal{W})^{-1}$$

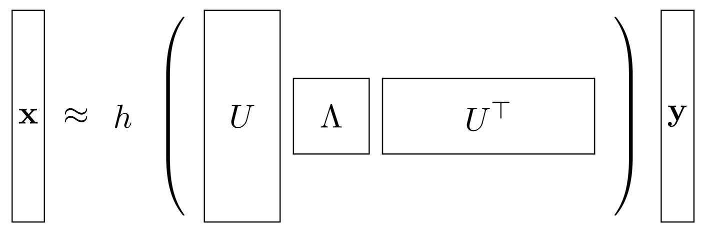
Fast Spectral Ranking



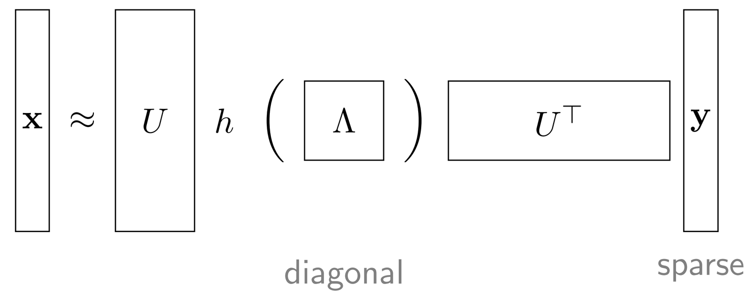
generalize h_α to an arbitrary transfer function h



eigenvalue decomposition of $\mathcal{W}$

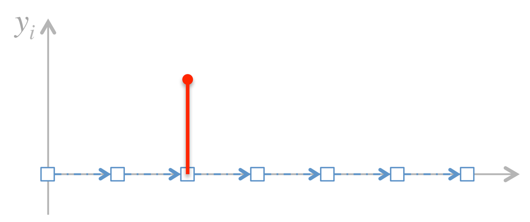


low-rank approximation

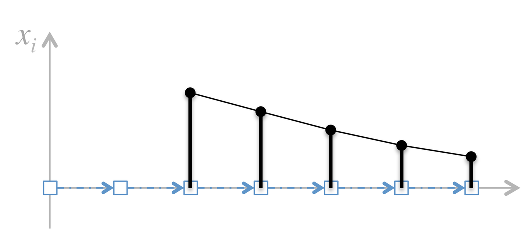


dot-product search (under conditions on h and Λ)

Smoothing in the frequency domain

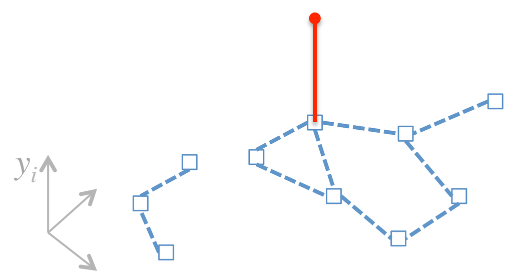


(a) Input signal $\mathbf{y}$

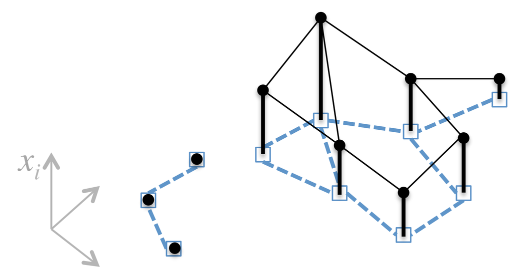


(b) Output signal $\mathbf{x}$

Low-pass filtering (moving average) of an impulse over the integers

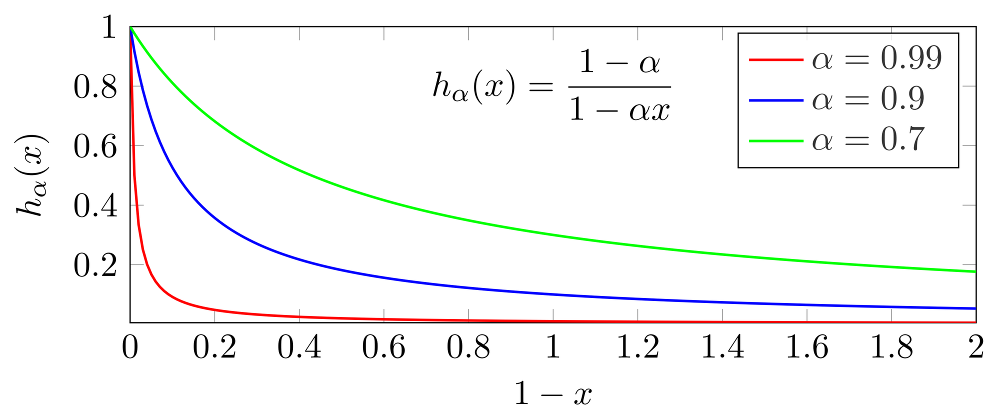


(c) Input signal $\mathbf{y}$



(d) Output signal $\mathbf{x}$

Low-pass filtering over a weighted undirected graph: impulse is query



h_α is a 'low-pass filter'; $1-x$ represents eigenvalues of $\mathcal{L}_\alpha$

Low-dimensional manifold representation

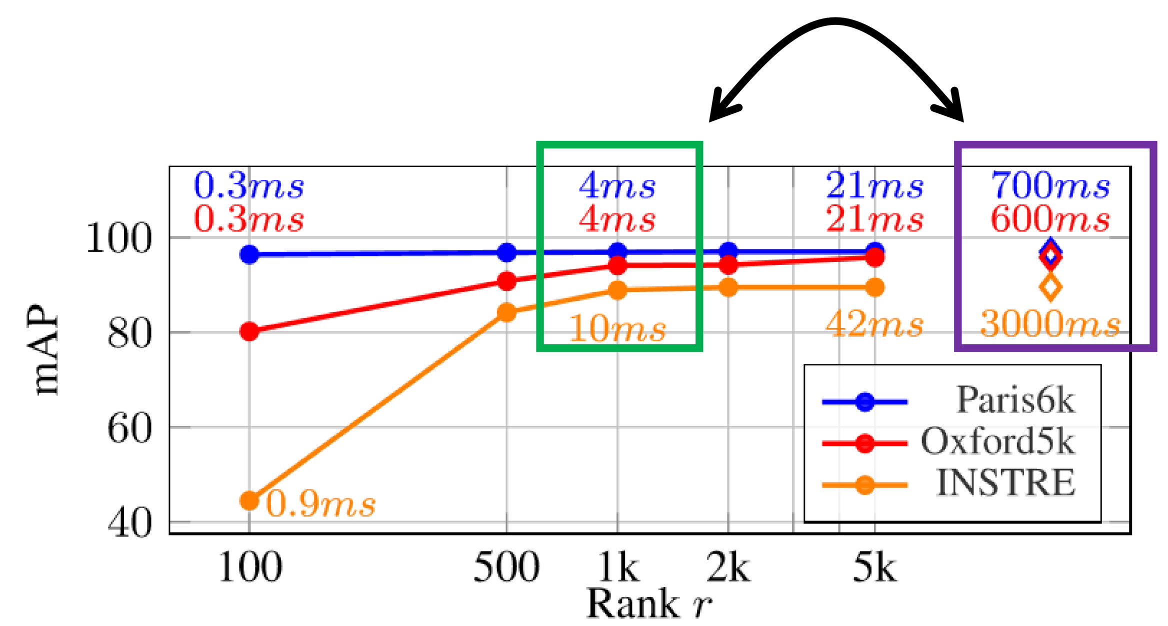
$$\begin{aligned}\mathbf{x} &= \mathcal{L}_\alpha^{-1}\mathbf{y} \\ &\approx U h_\alpha(\Lambda) U^\top \mathbf{y} \\ &= \Phi^\top \Phi \mathbf{y}\end{aligned}$$

- Dataset representations $\Phi := h_\alpha(\Lambda)^{1/2} U^\top \in \mathbb{R}^{n\times r}$
- Query $\Phi\mathbf{y}$ is a linear combination of dataset representations
- Manifold search is reduced to dot product search

Image retrieval

- Regional diffusion on Oxford (5k), Paris (6k) and Instre (27k)
- Each image is represented by 21 regions/vectors on average.
- Each region represented by 2048-D ResNet descriptor [1]
- Graph size: Oxford ($\approx$100k), Paris ($\approx$100k), Instre ($\approx$500k)

Same performance as **[2]**, two orders of magnitude faster



False negatives?

Top-ranked negative images (incorrectly labeled and correctly labeled)
Fixed annotations: See Paper # 2730

References

[1] A. Gordo, J. Almazan, J. Revaud, and D. Larlus. End-to-end learning of deep visual representations for image retrieval. *IJCV*, 124, 2017.
[2] A. Iscen, G. Tolias, Y. Avrithis, T. Furon, and O. Chum. Efficient diffusion on region manifolds: Recovering small objects with compact CNN representations. In *CVPR*, 2017.

Contact: ahmet.iscen@cmp.felk.cvut.cz