Kernel local descriptors with implicit rotation matching

Andrei Bursuc and Giorgos Tolias and Hervé Jégou
Inria, Rennes

Overview

- We consider kernelized patch encodings → vectorized patch representation
- Revisiting the SIFT descriptor with match kernels
- Jointly encode pixel gradient and position continously (no quantization)
- Fast dominant orientation alignment without descriptor re-computation

Match kernels

Local descriptor representation: set of pixel attributes

$$\mathcal{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_i, \dots\},\ \mathbf{x}_i \in \mathbb{R}^d$$

e.g. $d = 1$ for grayscale patches; $d = 4$ in this work
Match kernel to compare two patches

$$K(\mathcal{X}, \mathcal{Y}) = \beta(\mathcal{X})\beta(\mathcal{Y}) \sum_{\mathbf{x}\in\mathcal{X}} \sum_{\mathbf{y}\in\mathcal{Y}} k(\mathbf{x}, \mathbf{y})$$

Vector mapping $\psi : \mathbb{R}^d \to \mathbb{R}^D$, such that $k(\mathbf{x}, \mathbf{y}) = \langle \psi(\mathbf{x}) | \psi(\mathbf{y}) \rangle$
Patch representation:

$$\mathbf{X} = \beta(\mathcal{X}) \sum_{\mathbf{x}\in\mathcal{X}} \psi(\mathbf{x}), \qquad \text{(such that } \|\mathbf{X}\| = 1\text{)}$$

Match kernel computation via inner product

$$\mathcal{K}(\mathcal{X}, \mathcal{Y}) = \beta(\mathcal{X})\beta(\mathcal{Y}) \sum_{\mathbf{x}\in\mathcal{X}} \sum_{\mathbf{y}\in\mathcal{Y}} \langle \psi(\mathbf{x}) | \psi(\mathbf{y}) \rangle = \langle \mathbf{X} | \mathbf{Y} \rangle$$

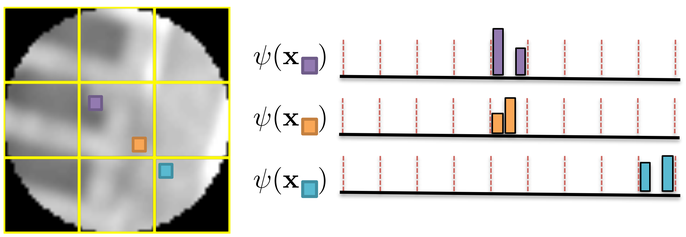
We seek to

Identify a kernel function for patch elements that reflects their resemblance in terms of gradients and their proximity in terms of spatial position.
Create vectorized local descriptor representation aggregating multiple pixel attributes, such that inner product of two such patch representations approximates a match kernel of the form:

$$\mathcal{K}(\mathbf{X}^\star, \mathbf{Y}^\star) = \beta(\mathcal{X})\beta(\mathcal{Y}) \sum_{\mathbf{x}\in\mathcal{X}} \sum_{\mathbf{y}\in\mathcal{Y}} \tilde{m}_\mathbf{x}\, \tilde{m}_\mathbf{y}\, k_\theta(\theta_\mathbf{x}, \theta_\mathbf{y})\, k_\varphi(\varphi_\mathbf{x}, \varphi_\mathbf{y})\, k_\rho(\tilde{\rho}_\mathbf{x}, \tilde{\rho}_\mathbf{y})$$
$$\approx \langle \mathbf{X}^\star | \mathbf{Y}^\star \rangle \qquad (1)$$

- $(m_\mathbf{x}, \theta_\mathbf{x})$: magnitude and gradient at pixel $\mathbf{x}$
- $(\varphi_\mathbf{x}, \rho_\mathbf{x})$: polar coordinates of pixel $\mathbf{x}$ with respect to the patch center
- $k_{\theta,\varphi,\rho}$: functions measuring consistency in gradients orientation, in positions on coordinate φ and on ρ, respectively

SIFT: kernel view



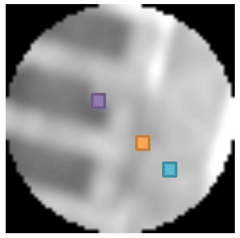
gradient sampling from grid *pixel gradient embedding*

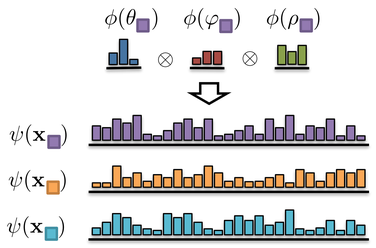
- Matching pixels with relatively similar positions and gradients orientations
- The hard assignment in the quantization process inserts some artifacts
- Losses in the selectivity of the similarity function

Method

Our approach

- Individual pixel sampling
- Continous embedding of pixel attributes
- Aggregation into dense pixel description



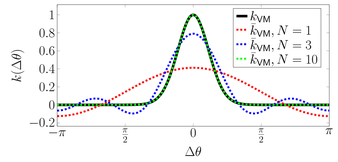


Weighting function for relative angles (normalized Von Mises)

$$k_{\text{VM}}(\Delta\theta) = \frac{\exp(\kappa\cos(\Delta\theta)) - \exp(-\kappa)}{2\sinh(\kappa)}$$

Fourier series approximation:

$$\bar{k}^N_{\text{VM}}(\Delta\theta) = \sum_{n=0}^{N} \gamma_n \cos(n\Delta\theta)$$



Feature map for angle

Define a mapping from angle θ to a vector $\phi(\theta)$:$[-\pi, \pi] \to \mathbb{R}^{2N+1}$
N: number of truncated components/frequencies from the Fouries series approximation

$$\phi(\theta) = (\sqrt{\gamma_0}, \sqrt{\gamma_1}\cos(\theta), \sqrt{\gamma_1}\sin(\theta), \dots, \sqrt{\gamma_N}\cos(N\theta), \sqrt{\gamma_N}\sin(N\theta))^\top \qquad (2)$$

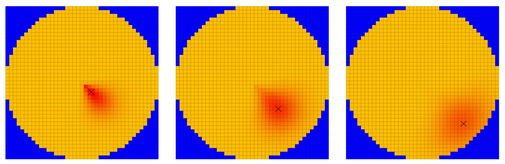
Compare vectors of angle via inner product → approximate angle similarity

$$\phi(\theta_1)^\top \phi(\theta_2) = \gamma_0 + \sum_{n=1}^{N} \gamma_n(\cos(n\theta_1)\cos(n\theta_2) + \sin(n\theta_1)\sin(n\theta_2))$$
$$= \sum_{n=0}^{N} \gamma_n \cos(n(\theta_1 - \theta_2)) = \bar{k}^N_{\text{VM}}(\theta_1 - \theta_2) \approx k_{\text{VM}}(\theta_1 - \theta_2)$$

Embedding and aggregating pixel attributes

θ and φ are angles and are trivially mapped with (2).
The radius ρ of each pixel $\mathbf{x}$ is mapped to an angle by $\tilde{\rho}_\mathbf{x} = \rho_\mathbf{x}\pi$, with $\rho_\mathbf{x} \in [0, 1]$ and then embedded with (2).

2D weighting function for 3 sample pixels. Colors reflect spatial similarity to other pixels. Red corresponds to maximum similarity



Final local descriptor representation:

$$\mathbf{X}^\star = \beta(\mathcal{X}^\star) \sum_{\mathbf{x}\in\mathcal{X}^\star} \psi(\mathbf{x}) = \beta(\mathcal{X}^\star) \sum_{\mathbf{x}\in\mathcal{X}^\star} \tilde{m}_\mathbf{x}\, \phi(\theta_\mathbf{x}) \otimes \phi(\varphi_\mathbf{x}) \otimes \phi(\tilde{\rho}_\mathbf{x})$$

$$\langle \mathbf{X}^\star | \mathbf{Y}^\star \rangle \propto \sum_{\mathbf{x}\in\mathcal{X}} \psi(\mathbf{x})^\top \sum_{\mathbf{y}\in\mathcal{Y}} \psi(\mathbf{y}) = \sum_{\mathbf{x}\in\mathcal{X}} \sum_{\mathbf{y}\in\mathcal{Y}} \psi(\mathbf{x})^\top \psi(\mathbf{y}) \approx \mathcal{K}(\mathbf{X}^\star, \mathbf{Y}^\star)$$

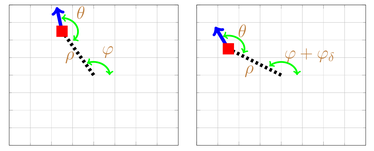
References

- L. Bo, X. Ren, and D. Fox. Kernel descriptors for visual recognition. In *NIPS*, 2010.
- H. Jégou, M. Douze, and C. Schmid. Hamming embedding and weak geometric consistency for large scale image search. In *ECCV*, 2008.
- G. Tolias, T. Furon, and H. Jegou. Orientation covariant aggregation of local descriptors with embeddings. In *ECCV*, 2014.
- A. Vedaldi and A. Zisserman. Efficient additive kernels via explicit feature maps. *Trans. PAMI*, 2012.

Rotation matching

The method assumes upright objects

Evolution of pixel attributes when rotating patch by angle φ_δ



Angle φ is the only pixel attribute that changes

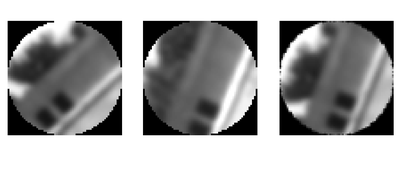
$$\mathbf{X}^\star = [\mathbf{X}_0^{\star\top}, \mathbf{X}_{1,c}^{\star\top}, \mathbf{X}_{1,s}^{\star\top}, \dots, \mathbf{X}_{N,c}^{\star\top}, \mathbf{X}_{N,s}^{\star\top}]^\top$$

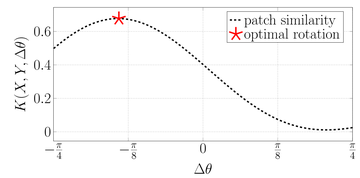
- $\mathbf{X}_0^\star$: associated to constant terms
- $\mathbf{X}_{N,c}^\star, \mathbf{X}_{N,s}^\star$: corresponding to *cosine* and *sine* terms for frequency N

$$\mathbf{X}^\star_{\delta,n,c} = \mathbf{X}^\star_{n,c}\cos n\varphi_\delta + \mathbf{X}^\star_{n,s}\sin n\varphi_\delta$$
$$\mathbf{X}^\star_{\delta,n,s} = -\mathbf{X}^\star_{n,c}\sin n\varphi_\delta + \mathbf{X}^\star_{n,s}\cos n\varphi_\delta$$

$$\langle \mathbf{X}^\star_\delta | \mathbf{Y}^\star \rangle = \langle \mathbf{X}^\star_0 | \mathbf{Y}^\star_0 \rangle + \sum_{n=1}^{N_\varphi} \cos(n\varphi_\delta)\ \left(\langle \mathbf{X}^\star_{n,c} | \mathbf{Y}^\star_{n,c} \rangle + \langle \mathbf{X}_{n,s} | \mathbf{Y}^\star_{n,s} \rangle\right)$$
$$+ \sum_{n=1}^{N_\varphi} \sin(n\varphi_\delta)\ \left(-\langle \mathbf{X}^\star_{n,c} | \mathbf{Y}^\star_{n,s} \rangle + \langle \mathbf{X}^\star_{n,s} | \mathbf{Y}^\star_{n,c} \rangle\right)$$

- $\mathbf{X}^\star_\delta$: aggregated representation of rotated query
- $\langle \mathbf{X}^\star_\delta | \mathbf{Y}^\star \rangle$: trigonometric polynomial with coefficients independent from φ_δ





Patch A *Patch B* *Aligned Patch A* *Similarities for rotations of Patch A*

Experiments

Hypothesis test on Brown patch dataset

False positive rate (%) at 95% recall. Learning type: N-none, US-unsupervised, S-supervised.

Train	Test	RootSIFT	RootSIFT-PCA	Simonyan	Brown	$KD_{2,3,1}$	$KD_{3,3,1}$	$KD_{3,2,2}$-PCA
ND	Lib	29.64	19.34	**12.42**	16.85	21.58	20.06	13.17
Yos	Lib		19.95	14.58	18.27			**14.53**
ND	Yos	26.69	18.37	10.08	13.55	11.07	9.66	**8.31**
Lib	Yos		19.52	11.18	N/A			**9.65**
Lib	ND	22.06	13.90	7.22	N/A	7.99	7.00	**6.36**
Yos	ND		13.98	6.82	11.98			**6.50**
Mean		26.14	17.51	10.38	15.16	13.55	12.24	**9.75**
Dimensions		128	80	73-77	29-36	105	147	80
Learning		N	US	S	S	N	N	US

Nearest neighbors

Recall computed at R top ranked patches for 1000 randomly selected patch queries on Notredame dataset (80d)

R	1	5	10	100	1000	10000
RootSIFT	8.5	24.4	33.0	62.9	79.7	90.6
RootSIFT-PCA	8.8	23.9	32.7	61.4	78.4	90.4
$KD_{3,3,1}$ (No rotations)	9.1	24.7	34.6	64.9	80.8	91.3
$KD_{3,3,1}$ (16 rotations)	**8.8**	**26.2**	**37.3**	**68.3**	**84.4**	**93.1**
$KD_{3,3,1} - PCA$	9.4	24.9	35.2	66.4	82.9	92.4

Contact: {andrei.bursuc,giorgos.tolias,herve.jegou}@inria.fr