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 continuously (no quantization)
- Fast dominant orientation alignment without descriptor re-computation

Match kernels
Local descriptor representation: set of pixel attributes
\[ x = \{ x_1, x_2, \ldots, x_n \}, \ x \in \mathbb{R}^d \]
e.g. \( d = 1 \) for grayscale patches; \( d = 4 \) in this work
Match kernel to compare two patches
\[ K(x, y) = \beta(x) \beta(y) \sum_{x \in X} k(x, y) \]
Vector mapping \( \psi: \mathbb{R}^d \rightarrow \mathbb{R}^D \), such that \( k(x, y) = \psi(x) \psi(y) \)
Patch representation:
\[ X = \beta(X) \sum_{x \in X} \psi(x), \ \text{(such that } |X| = 1) \]
Match kernel computation via inner product
\[ K(X, Y) = \beta(X) \beta(Y) \sum_{x \in X} \sum_{y \in Y} \psi(x) \psi(y) = (X | Y) \]

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 such two patch representations approximates a match kernel of the form:
\[
K(X', Y') = \beta(X') \beta(Y') \sum_{x \in X'} \sum_{y \in Y'} m(x, \theta_x, \delta_x) k(\phi_x, \phi_y) k(\rho_x, \rho_y),
\]
\( m(\theta, \delta) \): magnitude and gradient at pixel \( x \)
\( (\phi, \rho) \): polar coordinates of pixel \( x \) with respect to the patch center
\( k(\phi, \rho) \): functions measuring consistency in gradients orientation, in positions on coordinate \( \phi \) and \( \rho \), respectively

Method

Our approach
- Individual pixel sampling
- Continuous embedding of pixel attributes
- Aggregation into dense pixel description

Weighting function for relative angles (normalized Von Mises)

\[
k_{\text{Von Mises}}(\Delta \theta) = \frac{\exp(k \cos(\Delta \theta)) - \exp(-k)}{2 \sinh(k)},
\]

Fourier series approximation:
\[
k_{\text{Von Mises}}(\Delta \theta) \approx \sum_{n=1}^{N_x} Y_n \cos(n \Delta \theta)
\]

Feature map for angle

Define a mapping from angle \( \theta \) to a vector \( \psi(\theta) \rightarrow \mathbb{R}^{N \times 1} \)
\( N \): number of truncated components/frequencies from the Fouries series approximation
\[
\psi(\theta) = (\gamma_{\theta_1}, \gamma_{\theta_2}, \gamma_{\theta_3}, \gamma_{\theta_4}, \gamma_{\theta_5}, \gamma_{\theta_6}, \gamma_{\theta_7}, \gamma_{\theta_8}, \gamma_{\theta_9}, \gamma_{\theta_10}, \gamma_{\theta_11}, \gamma_{\theta_12}, \gamma_{\theta_13}, \gamma_{\theta_14}, \gamma_{\theta_15}, \gamma_{\theta_16})
\]

Compare vectors of angle via inner product → approximate angle similarity
\[
\langle \psi(\theta_1), \psi(\theta_2) \rangle = Y_0 + \sum_{n=1}^{N} Y_n (\cos(n \phi_1) \cos(n \phi_2) + \sin(n \phi_1) \sin(n \phi_2))
\]

Embedding and aggregating pixel attributes

\( \theta \) and \( \phi \) are angles and are trivially mapped with \( (2) \).
The radius \( \rho \) of each pixel \( x \) is mapped to an angle by \( \rho = \rho_x \pi \), with \( \rho_x \in [0, 1] \) 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:
\[ X^1 = \beta(x^1) \sum_{x \in X} \psi(x) = \beta(x^1) \sum_{x \in X} m(x, \theta_x, \delta_x) k(\phi_x, \phi_y) k(\rho_x, \rho_y) \]
\[ (x^1 | y^1) = \sum_{x \in X} \psi(x) \psi(y) = \sum_{x \in X} \sum_{y \in Y} \psi(x) \psi(y) = K(X^1, Y^1) \]

References

Rotation matching
The method assumes upright objects

Evolution of pixel attributes when rotating patch by angle \( \phi_z \)

\[ X^1 = [X^1_1, X^1_2, X^1_3, \ldots, X^1_{4N}]^T \]
- \( X^1_1 \) associated to constant terms
- \( X^1_2 \) associated to cosine and sine terms for frequency \( N \)

\[ X^{1,N} = X^{1,0} \cos(n \phi_z) + X^{1,1} \sin(n \phi_z) \]
- \( X^{1,N} \): aggregated representation of rotated query
- \( X^{1,0} \): trigonometric polynomial with coefficients independent from \( \phi_z \)

Experiments
Hypothesis test on Brown patch dataset
False positive rate (%) at 95% recall. Learning type: N-none, US-unsupervised, S-supervised.

Nearst neighbors
Recall computed at R top ranked patches for 1000 randomly selected patch queries on Notredame dataset (800)