Kernel Descriptors for Visual Recognition

Liefeng Bo, Xiaofeng Ren, and Dieter Fox
University of Washington and Intel Labs Seattle

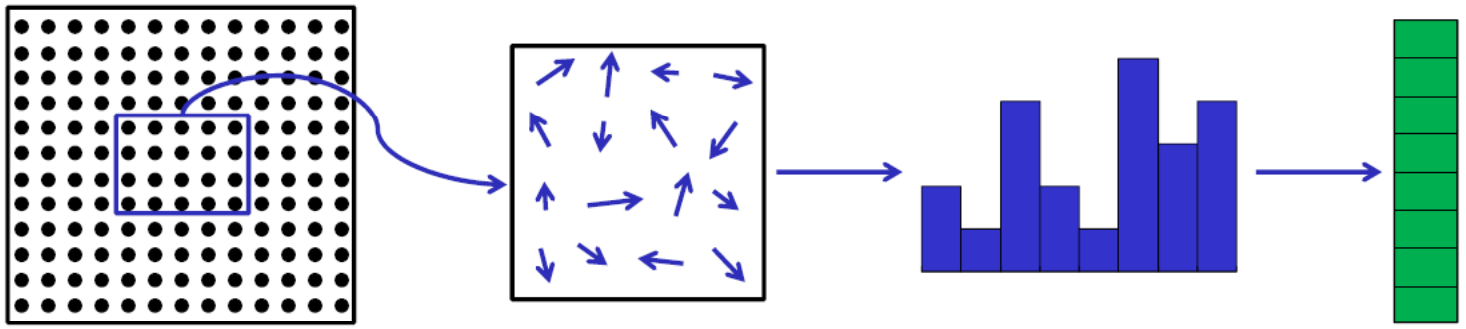
- **Orientation histograms (*e.g.* SIFT) are most successful in recognition**
- **A kernel view of features casts SIFT as a match kernel over patches**
- **Kernel Descriptors (KDES)**
  - **A principled way to design rich features to capture various visual attributes (*e.g.* using gradient, color and binary shape)**
  - **Learn compact features from match kernels via kernel approximation**
  - **Outperform SIFT and other sophisticated feature learning methods**
    Scene-15: 86.7%; Caltech101: 76.4%; CIFAR10: 76.0%

Novelty: **The first work on kernel-based low-level visual feature learning**

Poster ID
T40

Match Kernels over Image Patches

Most visual recognition systems are based on SIFT and HOG



Gradient Match Kernel

$$K_{\text{grad}}(P,Q) = \sum_{u \in P} \sum_{v \in Q} m_u m_v k_o(\theta_u, \theta_v) k_p(u,v)$$

- $K_{\text{grad}}(P,Q)$: image patch
- $m_u m_v$: normalized gradient magnitude
- $k_o(\theta_u, \theta_v)$: gradient orientation
- $k_p(u,v)$: pixel coordinates
- k_o, k_p: kernels

- ✓ **Includes SIFT and HOG as special cases;**
- ✓ **Avoid binning involved in SIFT or HOG;**
- ✓ **Makes it "easy" to design other match kernels:**

  pixel values →color match kernel; local binary pattern →shape match kernel.

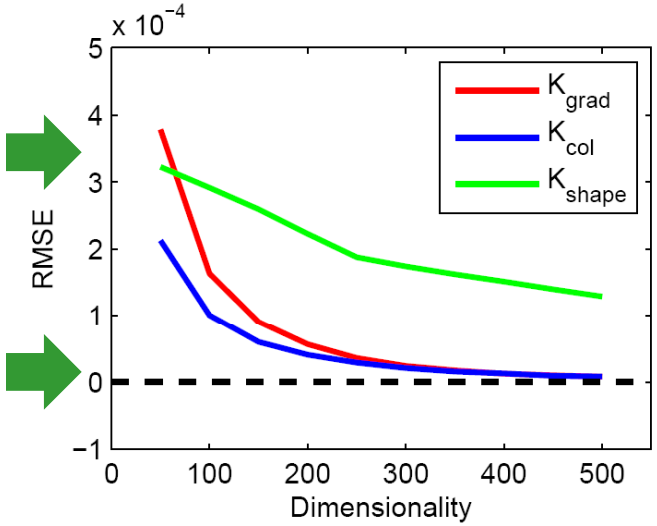
Kernel Descriptors

Why Kernel Descriptors? **(1) explicit low-dimensional visual features**
(2) efficient computation and storage

① **Uniformly and densely sample sufficient basis vectors to guarantee accurate approximation to match kernels;**

② **Learn explicit low-dimensional features based on sampled set using kernel principal component analysis (KPCA).**

***task-independent; ****accurate approximation**

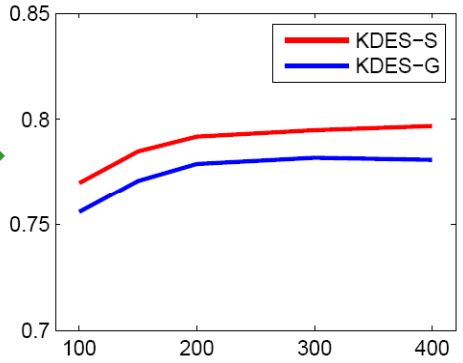


Gradient Kernel Descriptor

$$F_{\text{grad}}^{t}(P) = \sum_{i=1}^{d_o} \sum_{j=1}^{d_p} \alpha_{ij}^{t} \sum_{u \in P} m_u k_o(\theta_u, x_i) k_p(u, y_j)$$

Experiments

- **Free parameters in kernel descriptors are optimized on a subset of ImageNet.**
- **The resulting values are fixed in the following experiments.**



Scene-15

KDES:	**86.7%**
SIFT:	**82.2%**

Caltech-101

KDES:	**76.4%**	**CDBN[2]:**	**65.5%**
SPM[1]:	**64.4%**	**LCC[4]:**	**73.4%**

CIFAR10

KDES:	**76.0%**	**LCC[4]:**	**74.5%**
mcRBM-DBN[3]:	**71.0%**	**TCNN[5]:**	**73.1%**

[1] Lazebnik, Schmid, Ponce, CVPR '06 [2] Lee, Grosse, Ranganath, Ng, ICML '09
[3] Ranzato, Hinton, CVPR '10 [4] Yu, Zhang, ICML '10
[5] Le, Ngiam, Chen, Chia, Koh, Ng, NIPS '10