3D Detection from D-RGB data

Presented by
Caleb Jordan
Olga Russakovsky

Outline

- Generalized view of feature kernels
  - L. Bo, X. Ren and D. Fox "Kernel descriptors for visual recognition" NIPS 2010
- Large-scale RGBD dataset
  - K. Lai, L. Bo, X. Ren and D. Fox "A Large-scale hierarchical multi-view RGB-D object detection" ICRA 2010
- Applying hierarchical feature kernels to RGBD data
  - L. Bo, K. Lai, X. Ren and D. Fox "Object recognition with hierarchical kernel descriptors" CVPR 2011

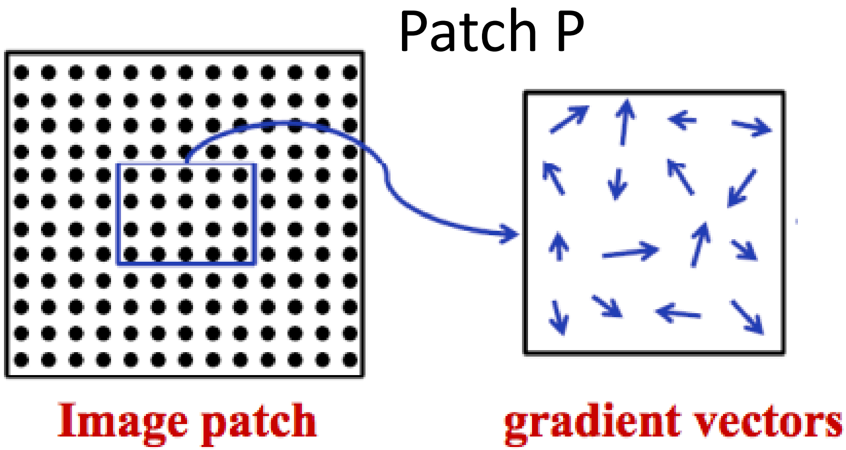
Outline

- **Generalized view of feature kernels**
  - **L. Bo, X. Ren and D. Fox "Kernel descriptors for visual recognition" NIPS 2010**
- Large-scale RGBD dataset
  - K. Lai, L. Bo, X. Ren and D. Fox "A Large-scale hierarchical multi-view RGB-D object detection" ICRA 2010
- Applying hierarchical feature kernels to RGBD data
  - L. Bo, K. Lai, X. Ren and D. Fox "Object recognition with hierarchical kernel descriptors" CVPR 2011

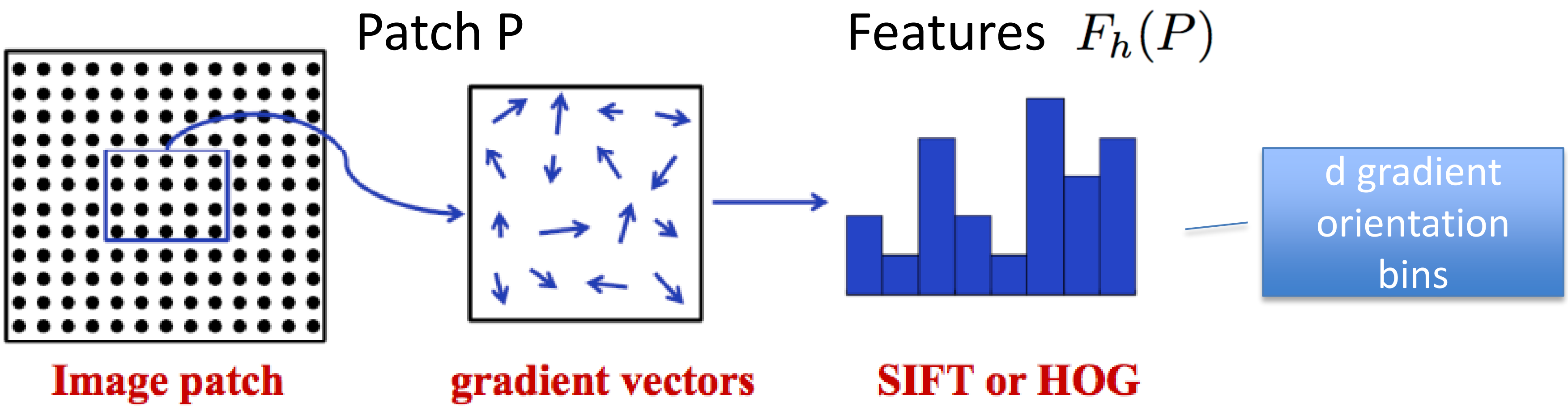
Kernel descriptors for visual recognition (NIPS ‘10)

- Kernel view of low-level image descriptors based on orientation histograms (HOG, SIFT)
  - plus cool tricks and approximations
- Features validated on RGB image classification
  - Scene-15, Caltech-101, CIFAR10, CIFAR10-ImageNet

Prior work: orientation histograms (SIFT, HOG)



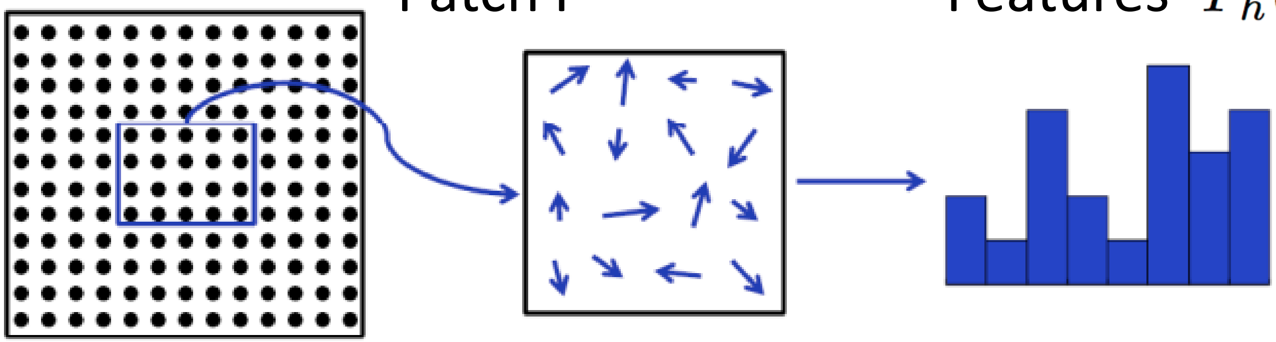
Prior work: orientation histograms (SIFT, HOG)



Prior work: orientation histograms (SIFT, HOG)

Patch P

Features $F_h(P)$



d gradient orientation bins

Image patch

gradient vectors

SIFT or HOG

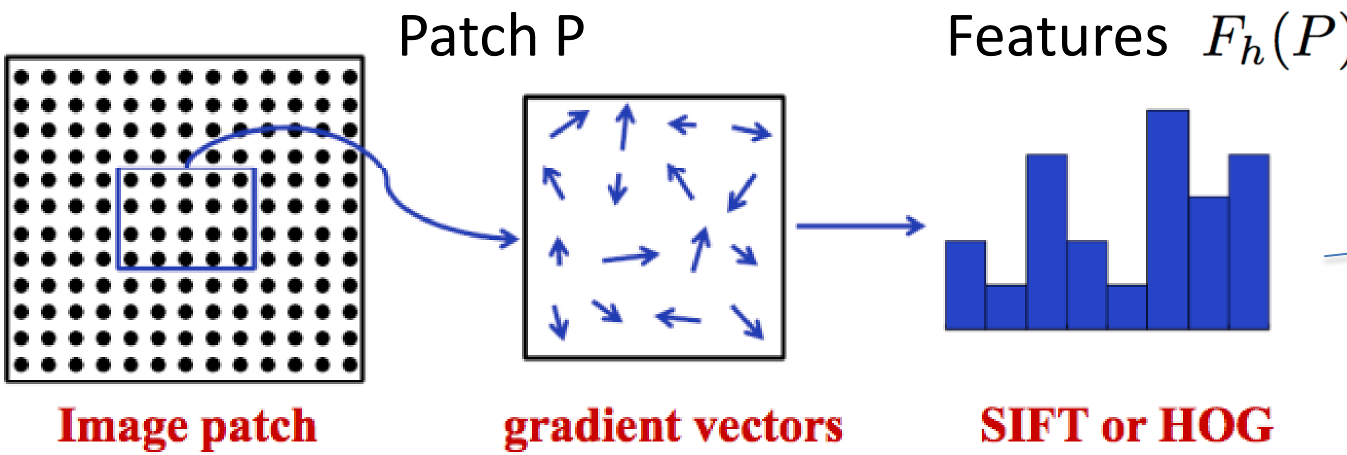
Sum over pixels

$$F_h(P) = \sum_{z \in P} \widetilde{m}(z)\delta(z)$$

d-dimensional indicator vector for d gradient orientation bins

Gradient magnitude

Kernelizing orientation histograms



d gradient orientation bins

Similarity between patch P and patch Q

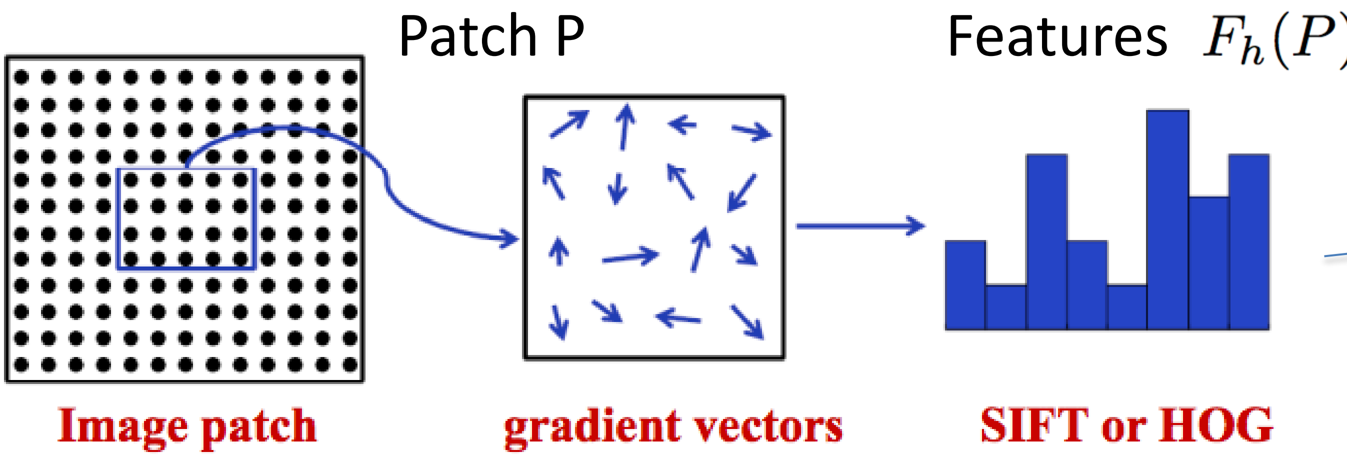
Sum over every pair of pixels

$$K_h(P,Q) = F_h(P)^\top F_h(Q) = \sum_{z \in P} \sum_{z' \in Q} \tilde{m}(z)\tilde{m}(z')\delta(z)^\top \delta(z')$$

Gradient magnitude

Orientation

Kernelizing orientation histograms



d gradient orientation bins

Similarity between patch P and patch Q

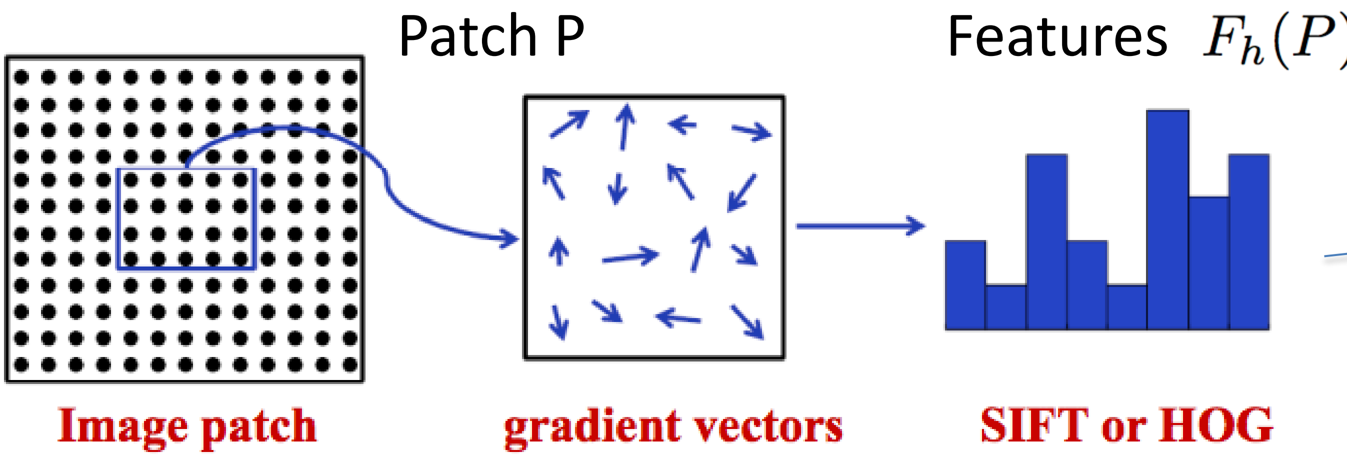
Sum over every pair of pixels

$$K_{\text{grad}}(P,Q) = \sum_{z \in P} \sum_{z' \in Q} \widetilde{m}(z)\widetilde{m}(z')k_o(\widetilde{\theta}(z),\widetilde{\theta}(z'))k_p(z,z')$$

Gradient magnitude

Orientation

Kernelizing orientation histograms



d gradient orientation bins

Similarity between patch P and patch Q

Sum over every pair of pixels

Position

$$K_{\mathrm{grad}}(P,Q) = \sum_{z \in P} \sum_{z' \in Q} \widetilde{m}(z)\widetilde{m}(z')k_o(\widetilde{\theta}(z), \widetilde{\theta}(z'))k_p(z, z')$$

Gradient magnitude

Orientation

Pixel-level similarity measures

Position $$k_p(z, z') = \exp(-\gamma_p \|z - z'\|^2)$$

Pixel-level similarity measures

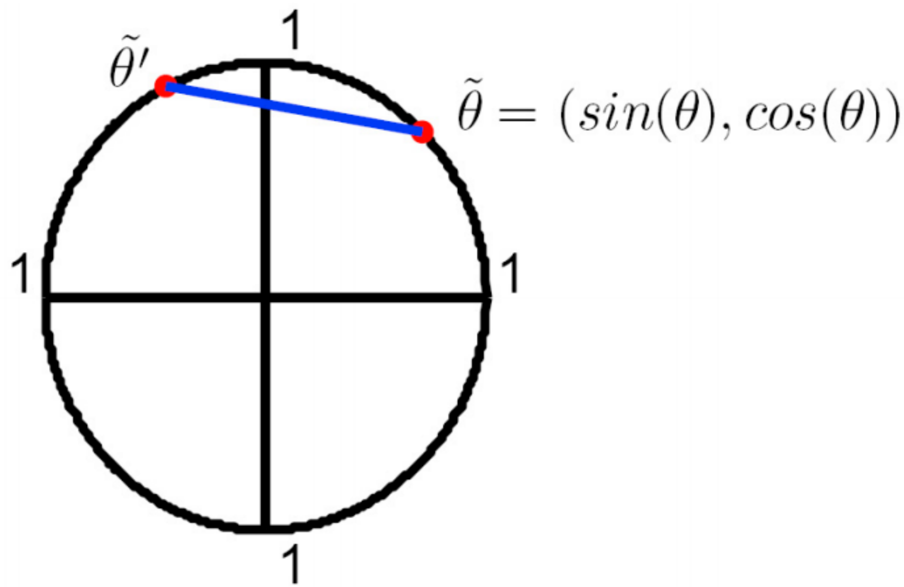
Position $$k_p(z, z') = \exp(-\gamma_p \|z - z'\|^2)$$

Orientation $$k_o(\widetilde{\theta}(z), \widetilde{\theta}(z')) = \exp(-\gamma_o \|\widetilde{\theta}(z) - \widetilde{\theta}(z')\|^2)$$

Pixel-level similarity measures

Position $k_p(z, z') = \exp(-\gamma_p \|z - z'\|^2)$

Orientation $k_o(\widetilde{\theta}(z), \widetilde{\theta}(z')) = \exp(-\gamma_o \|\widetilde{\theta}(z) - \widetilde{\theta}(z')\|^2)$



Pixel-level similarity measures

Position $$k_p(z, z') = \exp(-\gamma_p \|z - z'\|^2)$$

Orientation $$k_o(\widetilde{\theta}(z), \widetilde{\theta}(z')) = \exp(-\gamma_o \|\widetilde{\theta}(z) - \widetilde{\theta}(z')\|^2)$$

Pixel-level similarity measures

Position $$k_p(z, z') = \exp(-\gamma_p \|z - z'\|^2)$$

Orientation $$k_o(\widetilde{\theta}(z), \widetilde{\theta}(z')) = \exp(-\gamma_o \|\widetilde{\theta}(z) - \widetilde{\theta}(z')\|^2)$$

Color $$k_c(c(z), c(z')) = \exp\left(-\gamma_c \|c(z) - c(z')\|^2\right)$$

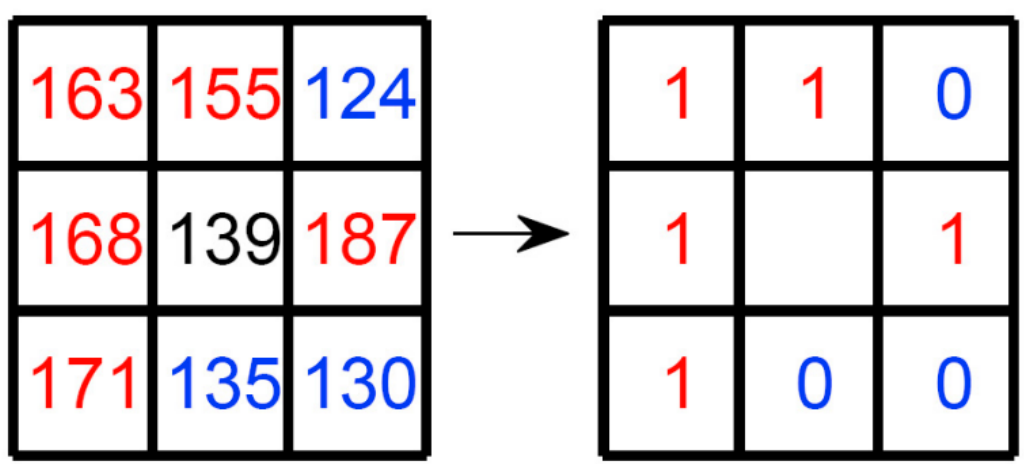
Pixel-level similarity measures

Position $k_p(z, z') = \exp(-\gamma_p \|z - z'\|^2)$

Orientation $k_o(\widetilde{\theta}(z), \widetilde{\theta}(z')) = \exp(-\gamma_o \|\widetilde{\theta}(z) - \widetilde{\theta}(z')\|^2)$

Color $k_c(c(z), c(z')) = \exp\left(-\gamma_c \|c(z) - c(z')\|^2\right)$

Shape $k_b(b(z), b(z')) = \exp(-\gamma_b \|b(z) - b(z')\|^2)$



Approximating pixel-level similarity measures

Orientation

$$k_o(\widetilde{\theta}(z), \widetilde{\theta}(z')) = \exp(-\gamma_o \|\widetilde{\theta}(z) - \widetilde{\theta}(z')\|^2)$$

Approximating pixel-level similarity measures

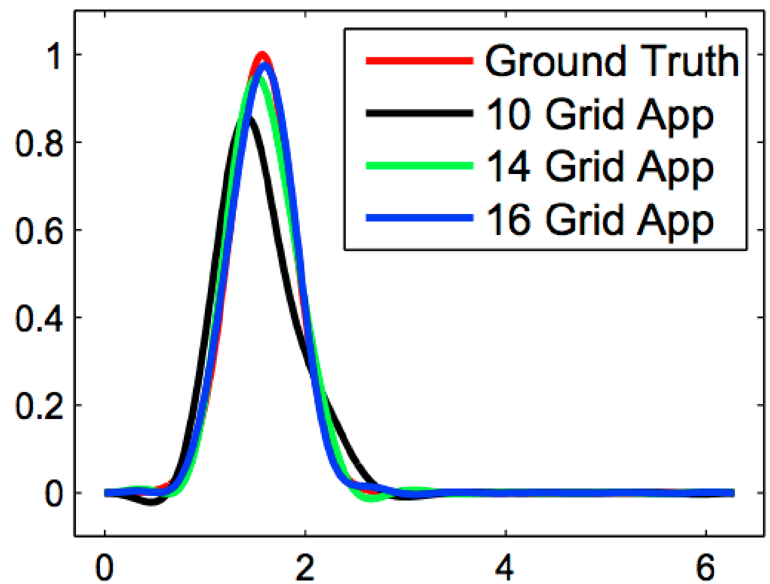
Orientation

$$k_o(\widetilde{\theta}(z), \widetilde{\theta}(z')) = \exp(-\gamma_o \|\widetilde{\theta}(z) - \widetilde{\theta}(z')\|^2)$$
$$= \phi_o(\widetilde{\theta}(z))^\top \phi_o(\widetilde{\theta}(z'))$$

Approximating pixel-level similarity measures

Orientation

$$k_o(\widetilde{\theta}(z), \widetilde{\theta}(z')) = \exp(-\gamma_o \|\widetilde{\theta}(z) - \widetilde{\theta}(z')\|^2)$$
$$= \phi_o(\widetilde{\theta}(z))^\top \phi_o(\widetilde{\theta}(z'))$$



Pixel-level similarity measures

Position $k_p(z, z')$

Orientation $k_o(\widetilde{\theta}(z), \widetilde{\theta}(z'))$

Color $k_c(c(z), c(z'))$

Shape $k_b(b(z), b(z'))$

Patch-level similarity measures

Orientation

$$K_{\text{grad}}(P,Q) = \sum_{z \in P} \sum_{z' \in Q} \widetilde{m}(z)\widetilde{m}(z') \boxed{k_o(\widetilde{\theta}(z), \widetilde{\theta}(z'))} \boxed{k_p(z, z')}$$

Color

$$K_{\text{col}}(P,Q) = \sum_{z \in P} \sum_{z' \in Q} \boxed{k_c(c(z), c(z'))} \boxed{k_p(z, z')}$$

Shape

$$K_{\text{shape}}(P,Q) = \sum_{z \in P} \sum_{z' \in Q} \widetilde{s}(z)\widetilde{s}(z') \boxed{k_b(b(z), b(z'))} \boxed{k_p(z, z')}$$

Approximating patch-level similarity measures

Orientation

$$K_{\mathrm{grad}}(P,Q)=\sum_{z\in P}\sum_{z'\in Q}\widetilde{m}(z)\widetilde{m}(z')\boxed{k_o(\widetilde{\theta}(z),\widetilde{\theta}(z'))}\boxed{k_p(z,z')}$$

Approximating patch-level similarity measures

Orientation

$$K_{\text{grad}}(P,Q) = \sum_{z \in P} \sum_{z' \in Q} \widetilde{m}(z)\widetilde{m}(z') \boxed{k_o(\widetilde{\theta}(z), \widetilde{\theta}(z'))} \boxed{k_p(z, z')}$$

$$= F_{\text{grad}}(P)^\top F_{\text{grad}}(Q)$$

Approximating patch-level similarity measures

Orientation

$$K_{\text{grad}}(P,Q) = \sum_{z \in P} \sum_{z' \in Q} \tilde{m}(z)\tilde{m}(z') \boxed{k_o(\tilde{\theta}(z), \tilde{\theta}(z'))} \boxed{k_p(z, z')}$$

$$= F_{\text{grad}}(P)^\top F_{\text{grad}}(Q)$$

$$\boxed{k_o(\tilde{\theta}(z), \tilde{\theta}(z')) = \phi_o(\tilde{\theta}(z))^\top \phi_o(\tilde{\theta}(z'))}$$

$$\boxed{k_p(z, z') = \phi_p(z)^\top \phi_p(z')}$$

Approximating patch-level similarity measures

Orientation

$$K_{\text{grad}}(P,Q) = \sum_{z\in P}\sum_{z'\in Q} \widetilde{m}(z)\widetilde{m}(z')\boxed{k_o(\widetilde{\theta}(z),\widetilde{\theta}(z'))}\boxed{k_p(z,z')}$$

$$= F_{\text{grad}}(P)^\top F_{\text{grad}}(Q)$$

$$\boxed{k_o(\widetilde{\theta}(z),\widetilde{\theta}(z')) = \phi_o(\widetilde{\theta}(z))^\top \phi_o(\widetilde{\theta}(z'))}$$

$$\boxed{k_p(z,z') = \phi_p(z)^\top \phi_p(z')}$$

$$F_{\text{grad}}(P) = \sum_{z\in P} \widetilde{m}(z)\phi_o(\widetilde{\theta}(z)) \otimes \phi_p(z)$$

Approximating patch-level similarity measures

Orientation

$$K_{\text{grad}}(P,Q)=\sum_{z\in P}\sum_{z'\in Q}\widetilde{m}(z)\widetilde{m}(z')\boxed{k_o(\widetilde{\theta}(z),\widetilde{\theta}(z'))}\boxed{k_p(z,z')}$$

$$=F_{\text{grad}}(P)^\top F_{\text{grad}}(Q)$$

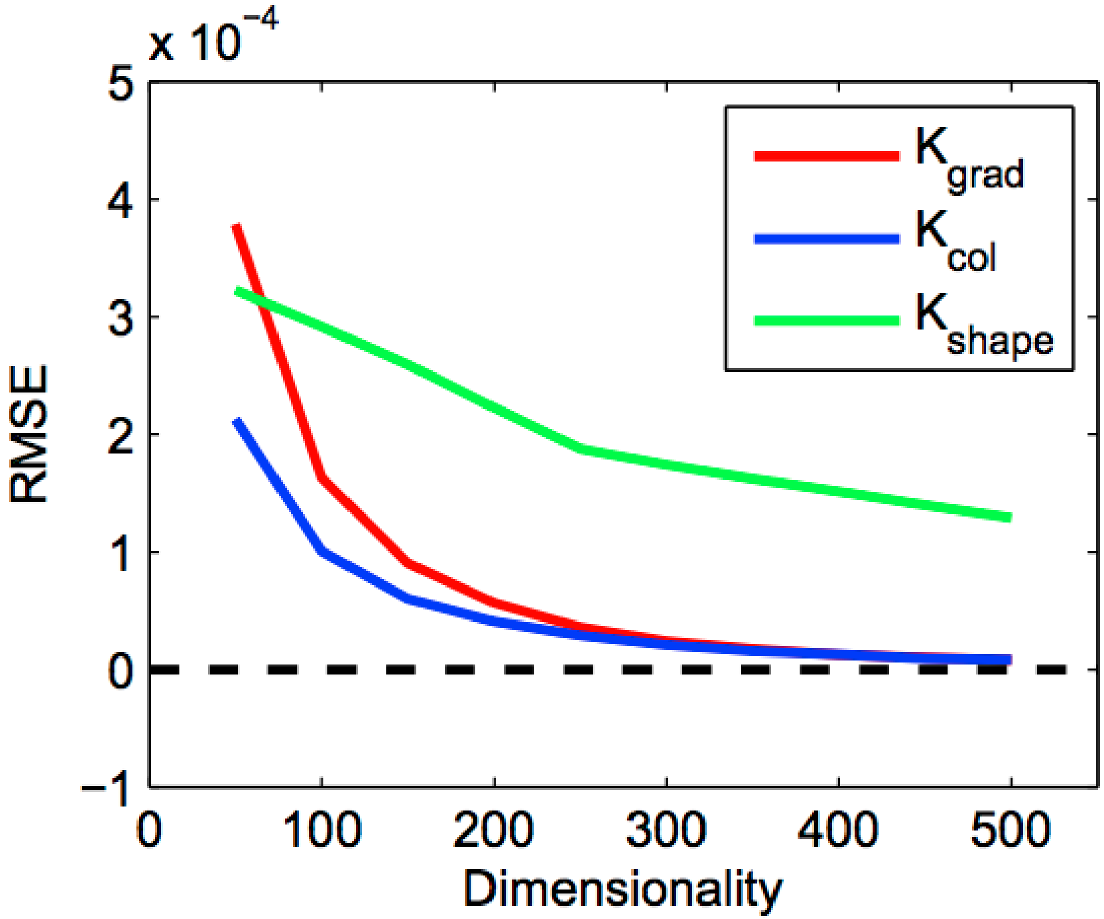
$$\boxed{k_o(\widetilde{\theta}(z),\widetilde{\theta}(z'))=\phi_o(\widetilde{\theta}(z))^\top\phi_o(\widetilde{\theta}(z'))}$$

$$\boxed{k_p(z,z')=\phi_p(z)^\top\phi_p(z')}$$

$$F_{\text{grad}}(P)=\sum_{z\in P}\widetilde{m}(z)\phi_o(\widetilde{\theta}(z))\otimes\phi_p(z)$$

… KPCA fancy math **(with code online!)**

Approximating patch-level similarity measures



Efficiency (in 2010)

On 300x300 image with MATLAB CPU implementation:

SIFT: 0.4s

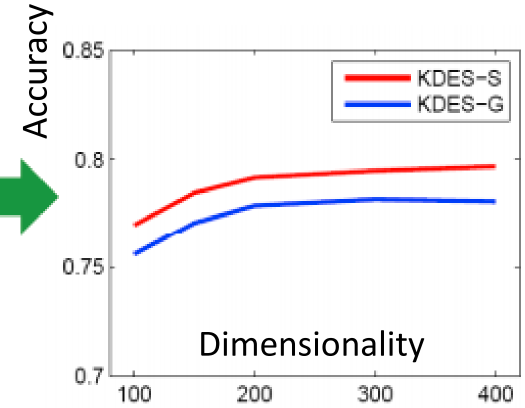
Gradient kernel: 1.5s

Shape kernel: 4s

Shamelessly stolen from their NIPS spotlight

Experiments (on RGB image classification)

- 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

Kernel descriptors for visual recognition (NIPS ‘10)

- Kernel view of low-level image descriptors based on orientation histograms (HOG, SIFT)
  - plus cool tricks and approximations
- Features validated on RGB image classification
  - Scene-15, Caltech-101, CIFAR10, CIFAR10-ImageNet
- Psst: code is online

Outline

- Generalized view of feature kernels
  - L. Bo, X. Ren and D. Fox "Kernel descriptors for visual recognition" NIPS 2010
- **Large-scale RGBD dataset**
  - **K. Lai, L. Bo, X. Ren and D. Fox "A Large-scale hierarchical multi-view RGB-D object detection" ICRA 2010**
- Applying hierarchical feature kernels to RGBD data
  - L. Bo, K. Lai, X. Ren and D. Fox "Object recognition with hierarchical kernel descriptors" CVPR 2011

A new RGB-D dataset

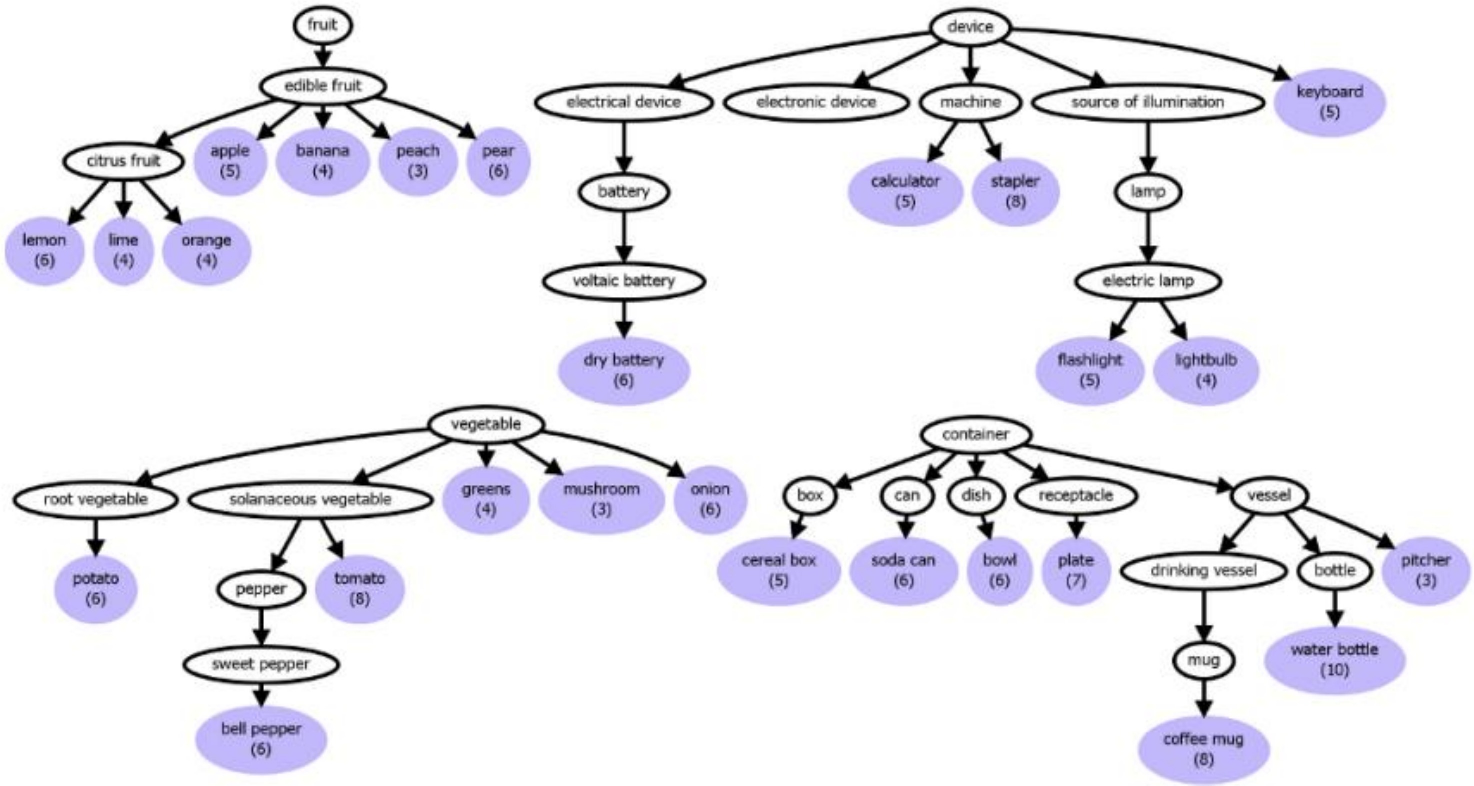
- RGB-D – RGB + depth values per frame
- 300 distinct objects, video sequence of ~250 frames for full revolutions at 30, 45, 60 degree elevation.

- (Previous was 3D Object Category Dataset by Savarese et al: 8 categories x 10 objects x 24 distinct views.)

http://www.cs.washington.edu/rgbd-dataset

Object categories

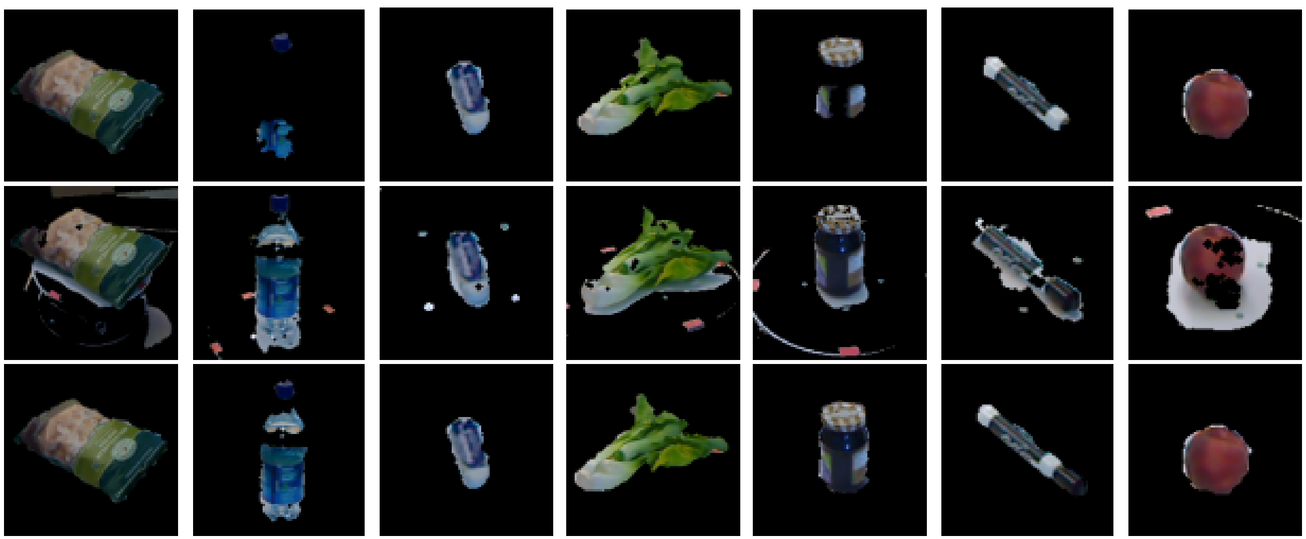
Category hierarchies



51 leaf nodes, 3-14 object instances per leaf category

RGB-D extras

- Images segmented using depth and color information



Segmented using RGB

Using depth

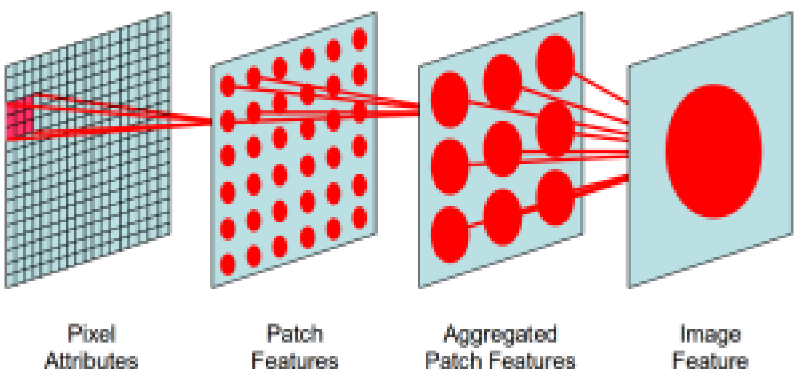
Combined

Outline

- Generalized view of feature kernels
  - L. Bo, X. Ren and D. Fox “Kernel descriptors for visual recognition” NIPS 2010
- Large-scale RGBD dataset
  - K. Lai, L. Bo, X. Ren and D. Fox “A Large-scale hierarchical multi-view RGB-D object detection” ICRA 2010
- **Applying hierarchical feature kernels to RGBD data**
  - **L. Bo, K. Lai, X. Ren and D. Fox “Object recognition with hierarchical kernel descriptors” CVPR 2011**

Hierarchical Kernel Descriptors (HKDES)

- Applies the concept of kernel descriptors recursively to create image-level features from pixel-level features.



Related Work

- Models with hierarchies or levels: deep belief nets, convolutional deep belief nets, convolutional neural networks
- Also sparse coding, spatial pyramid
- Parallel approach works on optimizing low-level match kernels

Recall first level kernels

- Gradient kernel

$$K_{\text{grad}}(P,Q) = \sum_{z \in P} \sum_{z' \in Q} \tilde{m}_z \tilde{m}_{z'} \boxed{k_o(\tilde{\theta}_z, \tilde{\theta}_{z'})} \boxed{k_p(z, z')}$$

- Color kernel

$$K_{\text{col}}(P,Q) = \sum_{z \in P} \sum_{z' \in Q} \boxed{k_c(c_z, c_{z'})} \boxed{k_p(z, z')}$$

Add a second level kernel

- Gradient kernel

$$\boxed{K_{\text{grad}}(P,Q)} = \sum_{z\in P}\sum_{z'\in Q} \tilde{m}_z \tilde{m}_{z'} \boxed{k_o(\tilde{\theta}_z, \tilde{\theta}_{z'})}\boxed{k_p(z,z')}$$

- Color kernel

$$\boxed{K_{\text{col}}(P,Q)} = \sum_{z\in P}\sum_{z'\in Q} \boxed{k_c(c_z, c_{z'})}\boxed{k_p(z,z')}$$

- Patch-level color + gradient kernel

$$K(\overline{P},\overline{Q}) = \sum_{A\in\overline{P}}\sum_{A'\in\overline{Q}} \widetilde{W}_A \widetilde{W}_{A'} \boxed{k_F(F_A, F_{A'})}\boxed{k_C(C_A, C_{A'})}$$

Kernel over patch-level gradient features

Kernel over patch-level color features

Results for CIFAR-10

HKDES results for RGB data on CIFAR-10 dataset

Features	KDES [1]	HKDES (this work)
Color	53.9	**63.4**
Shape	68.2	**69.4**
Gradient	66.3	**71.2**
Combination	76.0	**80.0**

Comparison to non-hierarchical

Method	Accuracy
Logistic regression [25]	36.0
Support Vector Machines [1]	39.5
GIST [25]	54.7
SIFT [1]	65.6
fine-tuning GRBM [24]	64.8
GRBM two layers [24]	56.6
mcRBM [25]	68.3
mcRBM-DBN [25]	71.0
Tiled CNNs [16]	73.1
improved LCC [31]	74.5
KDES + EMK + linear SVMs [1]	76.0
Convolutional RBM [4]	78.9
K-means (Triangle, 4k features) [4]	79.6
HKDES + linear SVMs (this work)	**80.0**

Comparison to state-of-the-art

Results for RGB-D!

Comparing RGB to depth for HKDES

Method	Category	Instance
Color HKDES (RGB)	60.1 ± 2.1	58.4
Shape HKDES (RGB)	72.6 ± 1.9	74.6
Gradient HKDES (RGB)	70.1 ± 2.9	75.9
Combination of HKDES (RGB)	76.1 ± 2.2	79.3
Color HKDES (depth)	61.8 ± 2.4	28.8
Shape HKDES (depth)	65.8 ± 1.8	36.7
Gradient HKDES (depth)	70.8 ± 2.7	39.3
Combination of HKDES (depth)	75.7 ± 2.6	46.8
Combination of all HKDES	$\mathbf{84.1\pm2.2}$	**82.4**

Comparing other approaches using a combination of color and depth information

Approaches	Category	Instance
Linear SVMs [15]	81.9 ± 2.8	73.9
Nonlinear SVMs [15]	83.8 ± 3.5	74.8
Random Forest [15]	79.6 ± 4.0	73.1
Combination of all HKDES	$\mathbf{84.1\pm2.2}$	**82.4**

Outline

- Generalized view of feature kernels
  - L. Bo, X. Ren and D. Fox “Kernel descriptors for visual recognition” NIPS 2010
- Large-scale RGBD dataset
  - K. Lai, L. Bo, X. Ren and D. Fox “A Large-scale hierarchical multi-view RGB-D object detection” ICRA 2010
- Applying hierarchical feature kernels to RGBD data
  - L. Bo, K. Lai, X. Ren and D. Fox “Object recognition with hierarchical kernel descriptors” CVPR 2011