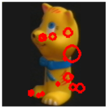
# Multiple-View Object Recognition in Band-Limited Distributed Camera Networks

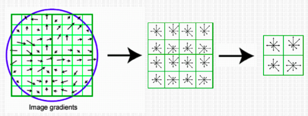
**Allen Y. Yang**
Subhransu Maji, Mario Christoudas, Trevor Darrell, Jitendra Malik, and Shankar Sastry

ICDSC, August 31, 2009

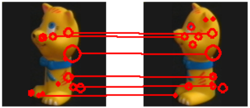
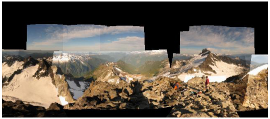
# Motivation: Object Recognition

- Affine invariant features, SIFT.

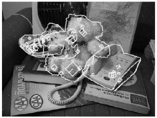





- SIFT Feature Matching [Lowe 1999, van Gool 2004]

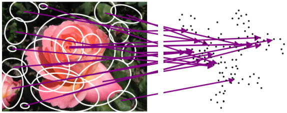
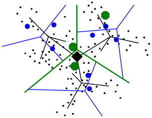



(a) Autostitch



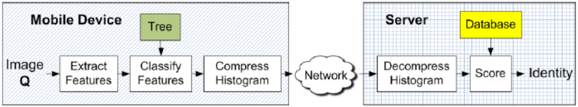
(b) Recognition

- Bag of Words [Nister 2006]

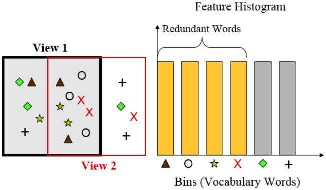



# Object Recognition in Band-Limited Sensor Networks

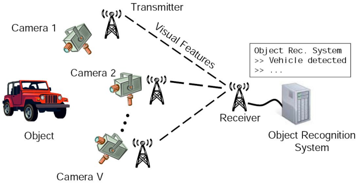
1. Compress scalable SIFT tree [Girod et al. 2009]




2. Multiple-view SIFT feature selection [Darrell et al. 2008]



# Problem Statement




1. $L$ camera sensors observe a single object in 3-D.
2. The mutual information between cameras are unknown, cross-sensor communication is prohibited.
3. On each camera, seek an encoding function for a **nonnegative, sparse** histogram $\mathbf{x}_i$

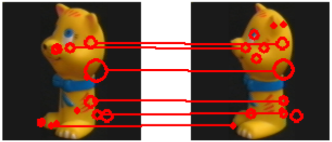
$$f : \mathbf{x}_i \in \mathbb{R}^D \mapsto \mathbf{y}_i \in \mathbb{R}^d$$

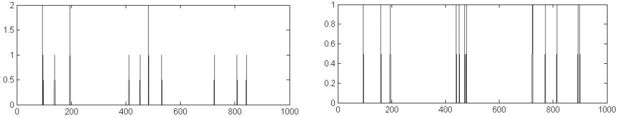
4. On the base station, upon receiving $\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_L$, **simultaneously recover**

$$\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_L,$$

and classify the object class in space.

# Key Observations





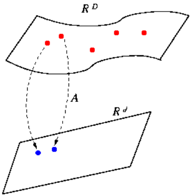
(a) Histogram 1 (b) Histogram 2

- All histograms are **nonnegative** and **sparse**.
- Multiple-view histograms share **joint sparse patterns**.
- Classification is based on the similarity measure in $\ell^2$-norm (linear kernel) or $\ell^1$-norm (intersection kernel).

# Compress SIFT Histograms: Random Projection

$$\mathbf{y} = A\mathbf{x}$$

Coefficients of $A \in \mathbb{R}^{d \times D}$ are drawn from zero-mean Gaussian distribution.




## Johnson-Lindenstrauss Lemma [Johnson & Lindenstrauss 1984, Frankl 1988]

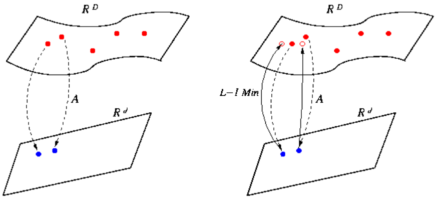
For $n$ number of point cloud in $\mathbb{R}^D$, given distortion threshold $\epsilon$, for any

$$d > O(\epsilon^2 \log n),$$

**a Gaussian random projection** $f(\mathbf{x}) = A\mathbf{x} \in \mathbb{R}^d$ **preserves pairwise** $\ell^2$**-distance**

$$(1-\epsilon)\|\mathbf{x}_i - \mathbf{x}_j\|_2^2 \leq \|f(\mathbf{x}_i) - f(\mathbf{x}_j)\|_2^2 \leq (1+\epsilon)\|\mathbf{x}_i - \mathbf{x}_j\|_2^2.$$

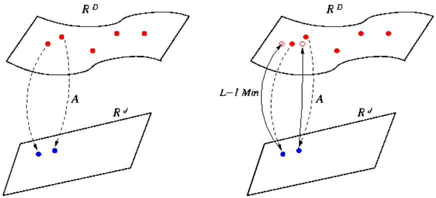
# From J-L Lemma to Compressive Sensing




(a) J-L lemma (b) Compressive sensing

1. **Problem I:** J-L lemma does not provide means to reconstruct **histogram hierarchy**.
2. **Problem II:** Gaussian projection **does not preserve $\ell^1$-distance** (for intersection kernels).
3. **Problem III:** Difficult (if not impossible) to incorporate **multiple-view** information.

# From J-L Lemma to Compressive Sensing




(a) J-L lemma (b) Compressive sensing

1. **Problem I:** J-L lemma does not provide means to reconstruct **histogram hierarchy**.
2. **Problem II:** Gaussian projection **does not preserve $\ell^1$-distance** (for intersection kernels).
3. **Problem III:** Difficult (if not impossible) to incorporate **multiple-view** information.

Compressive sensing provides principled solutions to the above problems.

# Compressive Sensing

## Noise-free case

Assume $\mathbf{x}_0$ is sufficiently $k$-sparse and mild condition on $A$,

$$(P_1): \quad \min \|\mathbf{x}\|_1 \text{ subject to } \mathbf{y} = A\mathbf{x}$$

recovers the exact solution.

# Compressive Sensing

## Noise-free case

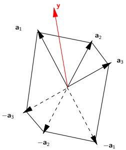
Assume $\mathbf{x}_0$ is sufficiently $k$-sparse and mild condition on $A$,

$$(P_1): \quad \min \|\mathbf{x}\|_1 \text{ subject to } \mathbf{y} = A\mathbf{x}$$

recovers the exact solution.

- Matching Pursuit [Mallat-Zhang 1993]

1. **Initialization:**
   - $\mathbf{y} = [A; -A]\tilde{\mathbf{x}}$, where $\tilde{\mathbf{x}} \geq 0$
   - $k \leftarrow 0$; $\tilde{\mathbf{x}} \leftarrow 0$; $\mathbf{r}^0 \leftarrow \mathbf{y}$; Sparse support $\mathcal{I} = \emptyset$



# Compressive Sensing

## Noise-free case

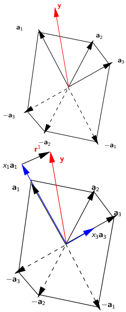
Assume $\mathbf{x}_0$ is sufficiently $k$-sparse and mild condition on $A$,

$$(P_1): \quad \min \|\mathbf{x}\|_1 \text{ subject to } \mathbf{y} = A\mathbf{x}$$

recovers the exact solution.

- Matching Pursuit [Mallat-Zhang 1993]

1. **Initialization:**
   - $\mathbf{y} = [A; -A]\tilde{\mathbf{x}}$, where $\tilde{\mathbf{x}} \geq 0$
   - $k \leftarrow 0$; $\tilde{\mathbf{x}} \leftarrow 0$; $\mathbf{r}^0 \leftarrow \mathbf{y}$; Sparse support $\mathcal{I} = \emptyset$
2. $k \leftarrow k + 1$:
   - $i = \arg\max_{j \notin \mathcal{I}} \{\mathbf{a}_j^T \mathbf{r}^{k-1}\}$
   - **Update**: $\mathcal{I} = \mathcal{I} \cup \{i\}$; $x_i = \mathbf{a}_i^T \mathbf{r}^{k-1}$; $\mathbf{r}^k = \mathbf{r}^{k-1} - x_i \mathbf{a}_i$



# Compressive Sensing

## Noise-free case

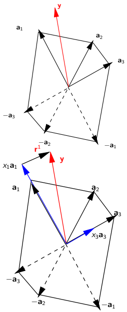
Assume $\mathbf{x}_0$ is sufficiently $k$-sparse and mild condition on $A$,

$$(P_1): \quad \min \|\mathbf{x}\|_1 \text{ subject to } \mathbf{y} = A\mathbf{x}$$

recovers the exact solution.

- Matching Pursuit [Mallat-Zhang 1993]

1. **Initialization:**
   - $\mathbf{y} = [A; -A]\tilde{\mathbf{x}}$, where $\tilde{\mathbf{x}} \geq 0$
   - $k \leftarrow 0$; $\tilde{\mathbf{x}} \leftarrow 0$; $\mathbf{r}^0 \leftarrow \mathbf{y}$; Sparse support $\mathcal{I} = \emptyset$
2. $k \leftarrow k+1$:
   - $i = \arg\max_{j \notin \mathcal{I}} \{\mathbf{a}_j^T \mathbf{r}^{k-1}\}$
   - **Update**: $\mathcal{I} = \mathcal{I} \cup \{i\}$; $x_i = \mathbf{a}_i^T \mathbf{r}^{k-1}$;
     $\mathbf{r}^k = \mathbf{r}^{k-1} - x_i \mathbf{a}_i$
3. **If**: $\|\mathbf{r}^k\|_2 > \epsilon$, **go to** STEP 2;
   **Else**: **output** $\tilde{\mathbf{x}}$



# Other Fast $\ell^1$-Min Routines

1. **Homotopy** Methods:
   - Polytope Faces Pursuit (PFP) [Plumbley 2006]
   - Least Angle Regression (LARS) [Efron-Hastie-Johnstone-Tibshirani 2004]
2. **Gradient Projection** Methods
   - Gradient Projection Sparse Representation (GPSR) [Figueiredo-Nowak-Wright 2007]
   - Truncated Newton Interior-Point Method (TNIPM) [Kim-Koh-Lustig-Boyd-Gorinevsky 2007]
3. **Iterative Thresholding** Methods
   - Soft Thresholding [Donoho 1995]
   - Sparse Reconstruction by Separable Approximation (SpaRSA) [Wright-Nowak-Figueiredo 2008]
4. **Proximal Gradient** Methods [Nesterov 1983, Nesterov 2007]
   - FISTA [Beck-Teboulle 2009]
   - Nesterov's Method (NESTA) [Becker-Bobin-Candés 2009]

## MATLAB Toolboxes

- SparseLab: `http://sparselab.stanford.edu/`
- $\ell^1$ Homotopy: `http://users.ece.gatech.edu/~sasif/homotopy/index.html`
- SpaRSA: `http://www.lx.it.pt/~mtf/SpaRSA/`

# Distributed Object Recognition in Smart Camera Networks

Outlines:

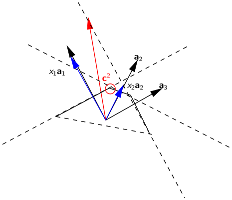
1. How to enforce nonnegativity to decode SIFT histograms?
2. How to enforce joint sparsity across multiple camera views?

# Enforcing Nonnegativity

- **Polytope Pursuit** Algorithms (MP, PFP, LARS):
  1. **Algebraically:** Do not add antipodal vertexes

$$\mathbf{y} = [A; \boxed{\text{-A}}]\tilde{\mathbf{x}}$$

  2. **Geometrically:** Pursuit on positive faces




- **Interior-Point** Algorithms (Homotopy, SpaRSA):
  Remove any sparse support that have negative coefficients.

# Sparse Innovation Model

- Definition (SIM):

$$\begin{aligned} \mathbf{x}_1 &= \tilde{\mathbf{x}} + \mathbf{z}_1, \\ &\vdots \\ \mathbf{x}_L &= \tilde{\mathbf{x}} + \mathbf{z}_L. \end{aligned}$$

$\tilde{\mathbf{x}}$ is called the **joint sparse** component, and $\mathbf{z}_i$ is called an **innovation**.

# Sparse Innovation Model

- Definition (SIM):

$$\begin{aligned} \mathbf{x}_1 &= \tilde{\mathbf{x}} + \mathbf{z}_1, \\ &\vdots \\ \mathbf{x}_L &= \tilde{\mathbf{x}} + \mathbf{z}_L. \end{aligned}$$

  $\tilde{\mathbf{x}}$ is called the **joint sparse** component, and $\mathbf{z}_i$ is called an **innovation**.

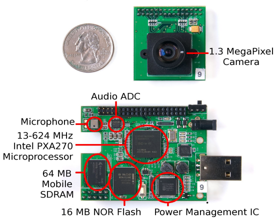
- Joint recovery of SIM

$$\begin{aligned} \begin{bmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_L \end{bmatrix} &= \begin{bmatrix} A_1 & A_1 & 0 & \cdots & 0 \\ \vdots & & \ddots & \ddots & \\ A_L & 0 & \cdots & 0 & A_L \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{x}} \\ \mathbf{z}_1 \\ \vdots \\ \mathbf{z}_L \end{bmatrix} \\ \Leftrightarrow \quad \mathbf{y}' &= A'\mathbf{x}' \in \mathbb{R}^{dL}. \end{aligned}$$

1. New histogram vector is **nonnegative** and **sparse**.
2. Joint sparsity $\tilde{\mathbf{x}}$ is automatically determined by $\ell^1$-min: No prior training, no assumption about fixing cameras.

# CITRIC: Wireless Smart Camera Platform

- CITRIC platform



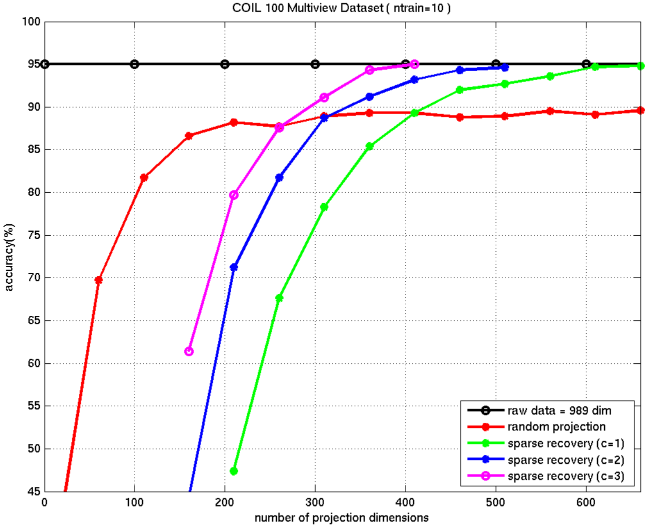

- Academic users:

- Available library functions
  1. Full support **Intel IPP Library** and **OpenCV**.
  2. **JPEG compression**: 10 fps.
  3. **Edge detector**: 3 fps.
  4. **Background Subtraction**: 5 fps.
  5. **SIFT detector**: 10 sec per frame.

# Experiment: COIL-100 object database

- **Database:** 100 objects, each provides 72 images captured with 5 degree difference.

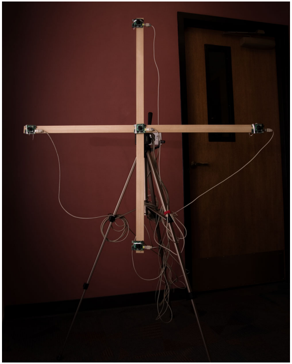

- **Setup:**
  - Dense sampling of overlapping $8 \times 8$ grids. PCA-SIFT descriptor.
  - 4-level hierarchical $k$-means ($k = 10$): Leaf-node histogram is 1000-D.
  - Classifier via intersection-kernel SVM: 10 random training images per class.

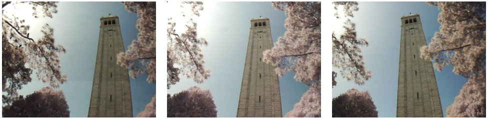

COIL 100 Multiview Dataset ( ntrain=10 )
accuracy(%)
number of projection dimensions
100
95
90
85
80
75
70
65
60
55
50
45
0
100
200
300
400
500
600
raw data = 989 dim
random projection
sparse recovery (c=1)
sparse recovery (c=2)
sparse recovery (c=3)

# Distributed Object Recognition in Band-Limited Smart Camera Networks

1. To harness the smart camera capacity, the system is separated in two components: **distributed feature extraction** and **centralized recognition**.
2. **Gaussian random projection** as universal dimensionality reduction function: J-L lemma.
3. $\ell^1$-**minimization** exploits two properties of SIFT histograms:
   - Sparsity.
   - Nonnegativity.
4. **Sparse innovation model** exploits joint sparsity of multiple-view histograms.
5. Complete system implemented on Berkeley CITRIC sensors.

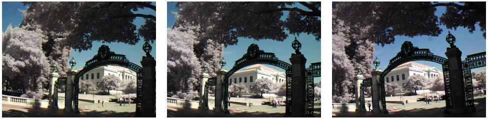
# Berkeley Multiple-view Wireless Database





(a) Campanile

(b) Bowles



(c) Sather Gate