Automatic Target Recognition via Sparse Representations

Outline

- Research objective / ATR challenges
- Background on Compressed Sensing (CS)
- ATR in the CS framework
  - Dictionary and sensing matrix
  - Detection
  - Recognition
- Results
- Conclusions

ATR Challenges / Objective

- **ATR systems must search, detect and identify targets of interest in cluttered environments in real or almost real time mode.**
- **Compressed Sensing theory offers robust algorithms that may allowed to meet the accuracy and speed criteria of ATR requirements while working in noisy environment.**

Why Compressed Sensing?

- **Sampling at the Nyquist rate is expensive and redundant when dealing with compressible signals.**
  - **Nyquist:** $f_s \geq 2 f_{bw}$
- **Compressive Sensing (new emerging theory):**
  - **Transform to low dimensional measurement domain.**
  - **Perform filtering / recognition / processing in the lower dimensional domain.**
  - **Translates into faster algorithms and inexpensive hardware require for processing.**

Compressed Sensing

$$y = \Phi A x$$

Mx1
Test vector

$$\begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} = \begin{pmatrix} f_{11} & \cdots & f_{1k} \\ \vdots & \ddots & \vdots \\ f_{m1} & \cdots & f_{mk} \end{pmatrix} \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{k1} & \cdots & a_{kn} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

Nx1
Sparse vector with **S** nonzero entries

MxK
Sensing/Random Matrix

KxN
Dictionary / Basis

$$M = O(S(\log(N/S))$$

$$M << N$$

Solve undetermined system of equations by exploiting the sparse nature of the signal

Recognition Problem as a Sparse Representation

- **Given test sample (*y*), identify y based on a set of training data: *y= Dx***

  ***w*here**

  - ***D* is a matrix containing all the training samples for all classes or individuals**
  - ***x* is the sparse vector containing the identity of *y***

  - **Of special interest is when the system *y= Dx* is undetermined (no unique solution).**
  - **Solve the following optimization problem:**

    - ***(l_0)*** $\hat{x}_0 = \min \| x \|_0 \ \ s.t. \ \ Dx = y$

    - **Sparsest solution to the l_0 optimization problem for an undetermined system is a highly non-convex combinatorial problem**

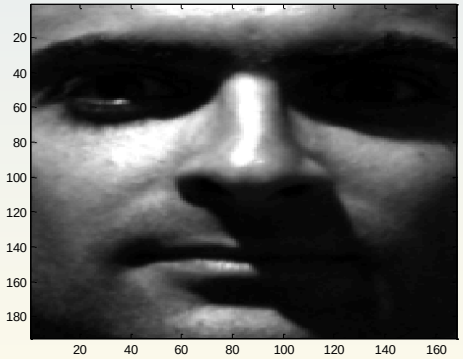
Solution

- **Theory of CS reveals that if the solution x_0 is sparse enough, then the solution of the l_0 minimization problem is equivalent to the following:**

  - **(l_1)** $\hat{x}_1 = \min \| x \|_1 \;\; s.t. \;\; Dx = y$

  - **The l_1 minimization problem can be solved by standard linear programming methods.**
  - **L1 solution can be achieved with algorithm such as:**
    - **Orthogonal Matching Pursuit (OMP)**
    - **Basis Pursuit/LASSO**
    - **Elastic Net**

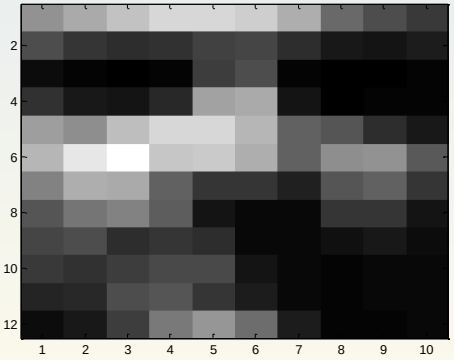
Face Recognition



(***y***) Test image belonging to class 11

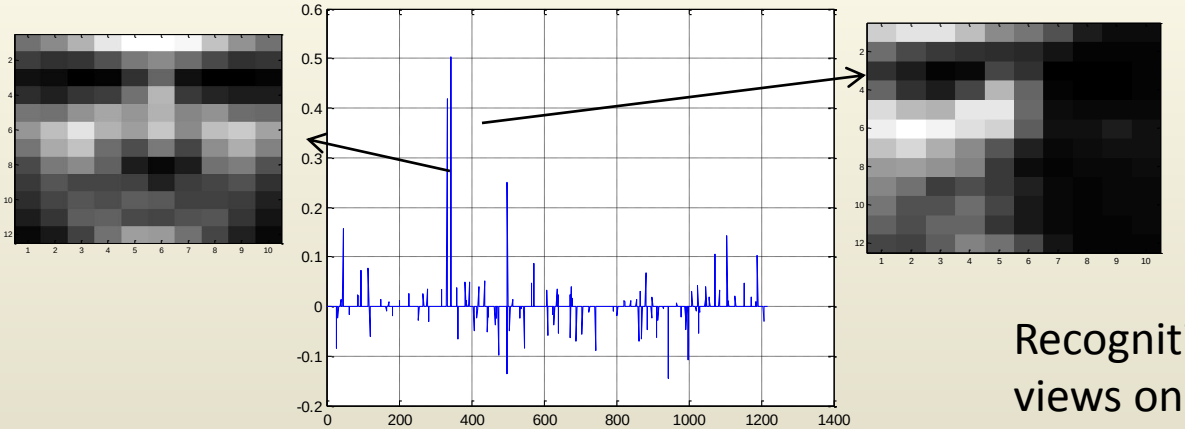
192x168

12x10



Downsampled test input

Sparse Solution (***x***)



Maximum coefficients belonging to the same class (11)

Recognition of frontal views only

Work by J. Wright

ATR in the CS Framework

Test Input $y = \Phi A x$ Target identifier

Gaussian Matrix

Dictionary composed of target images every $\Delta\theta°$

- **Resulting dictionary ΦA is an overcomplete dictionary.**
  - **Overcompleteness (or redundancy) brings flexibility and generality to the detection/classification problem.**
  - **Very few elements (maybe one) from the dictionary will be needed to represent the input signal.**

Database

- **Imagery collected by the US Army Night Vision and Electronic Sensors Directorate (NVESD).**
  - **8 targets at all aspect angles (azimuth) in range steps of 500 meters from 500 to 5000 meters during both day and night.**
  - **Two modalities: Mid-wave IR and Visible**
  - **Target location within each scene is provided for ground truth data as part of the database.**
- **1, 2 and 3 Km data sets were selected for this training and testing.**

Target to Detect and Recognize

Class 1

Pickup

Class 2

SUV

Class 3

BTR70 – Armored Personnel Carrier

Class 4

BRDM2 – Infantry Scout Vehicle

Class 5

BMP2 – Armored Personnel Carrier

Class 6

T72 – Main Battle Tank

Class 7

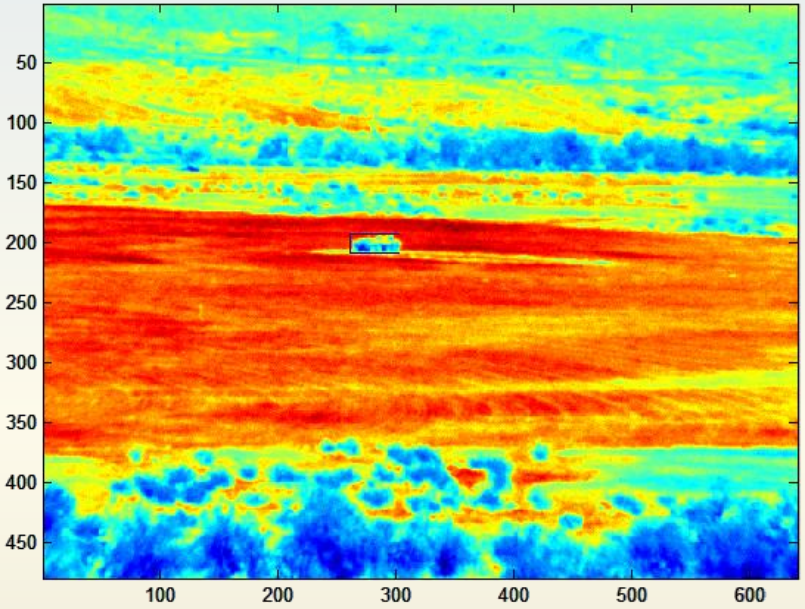
ZSU23-4 - Anti-Aircraft Weapon

Class 8

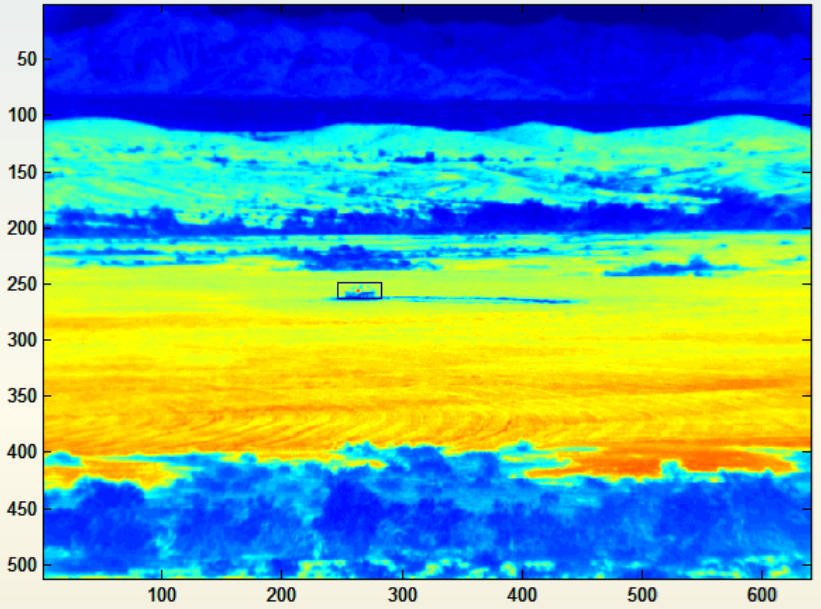
2S3 – Self-Propelled Howitzer

Target at 2Km

Visible

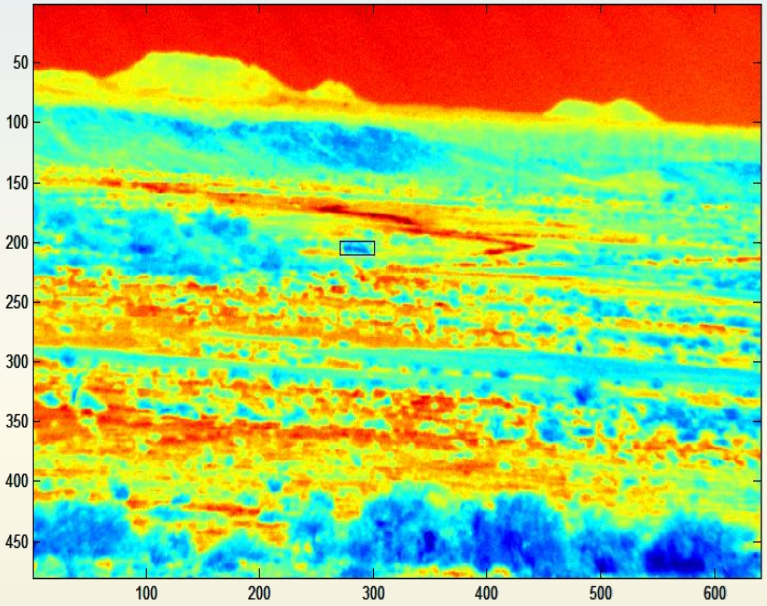


MWIR

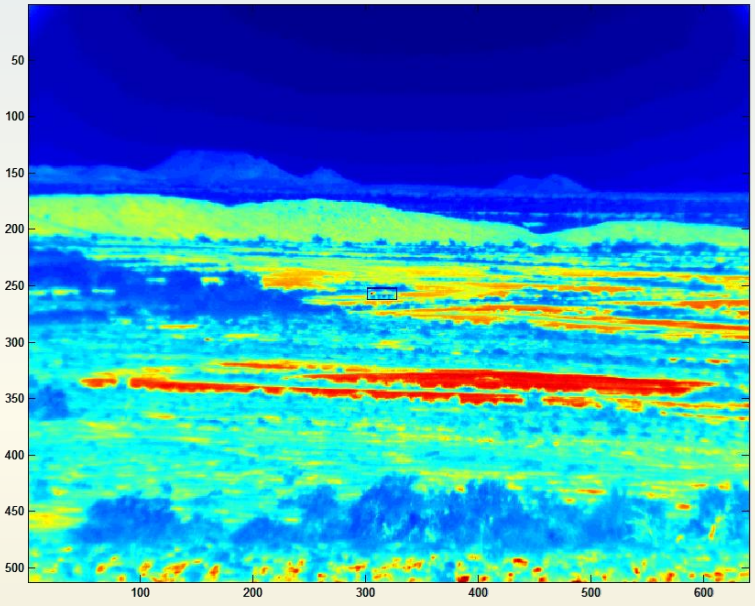


Target at 3Km

Visible



MWIR



Dictionary

- **Three components related to building the dictionary: target aspect resolution, target size and number of random measurements.**
  - **Targets and sensor characteristics are known, so it is easy to calculate the ROI (window size) where random sampling is to be performed.**
  - **Dictionary is built in $\Delta\theta°$ steps in order to cover the target from 0° to 360°**
    - **Each target class contains N different samples from different azimuth angles.**
  - **Random measurements $M>S*\log(N)$**
- **Background clutter(randomly selected from each target image already in the dictionary) is included as an additional class.**

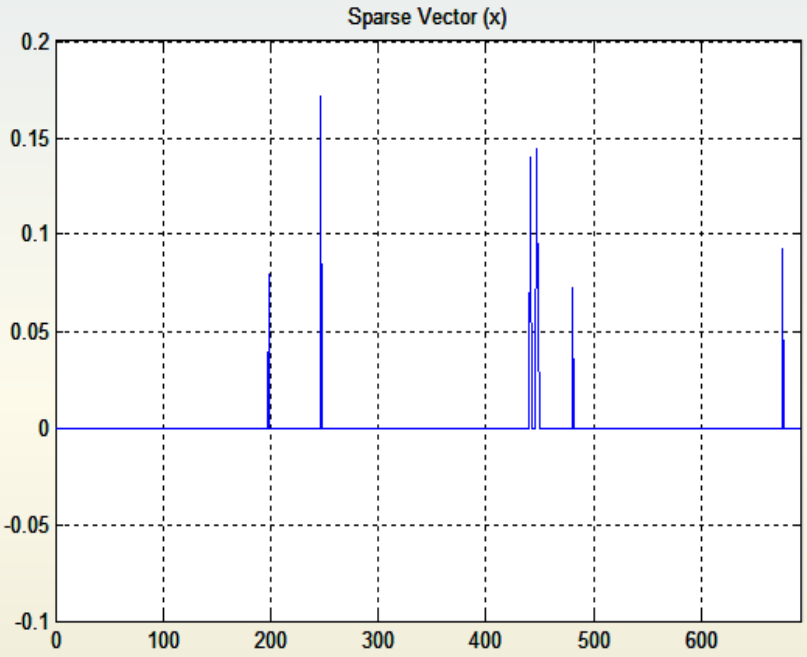
Recognition

- **LASSO:** $$\min_{x} \frac{1}{2} \| y - Ax \|_{l2}^{2} + \lambda \| x \|_{l1}$$
- **Robust, stable algorithm that computes the sparse solution(*x*).**

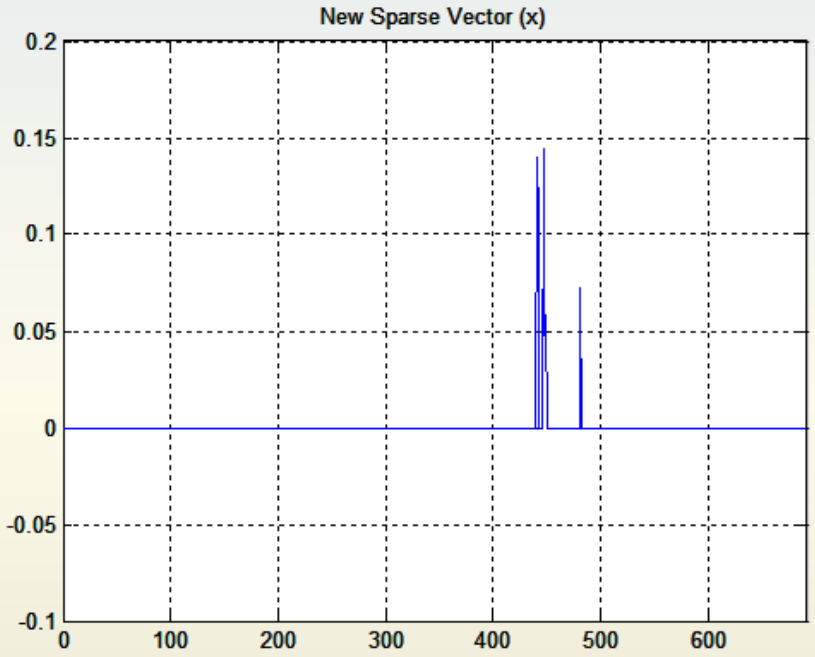
$$x_k = \begin{Bmatrix} x_k, & \sum_{i=1:M} x_i > \tau_2 \\ 0, & \sum_i x_i \le \tau_2 \end{Bmatrix}, x = x_{11}, x_{12}, \ldots\ldots x_{1M}, \ldots\ldots\ldots x_{k1}, x_{k2}, \ldots\ldots x_{kM}$$

- **Small coefficients in x indicate an invalid answer. Restricting the coefficients of $x > \tau_2$ increases the validity of SCI reducing false alarms**
- ***min*$||y-Ax_k||_2$, $SCI(x_i)>\tau_3$**

Sparse Vector(*x*) Modified



SCI(x_7)= 0.7



SCI(x_7)= 1

$$SCI(x_i) = \frac{k\|x_i\|_1}{\|x\|_1} - 1 \Big/ (k-1)$$

Results

Results from IR imagery at 3 Km, Pd=0.98, Pc=0.96, Pm=0.02, Pf-=0.04.

	truck	SUV	BTR70	BRDM2	BMP2	T72	ZSU23-4	2S3	clutter
truck	455	3	22	0	0	0	0	0	20
SUV	0	499	0	0	0	0	0	0	1
BTR70	0	0	491	2	0	0	0	0	7
BRDM2	5	0	1	463	0	1	0	4	26
BMP2	0	0	0	0	498	0	0	0	2
T72	0	0	0	0	0	499	0	0	1
ZSU23-4	0	0	0	0	0	0	495	0	5
2S3	8	0	0	12	0	7	4	451	18
clutter	0	4	0	0	5	4	0	5	482

Five hundred test samples for each target were utilized to generate results

Results

Results from Visible imagery at 3 Km, Pd=0.97, Pc=0.96, Pm=0.02, Pf=0.03.

	truck	SUV	BTR70	BRDM2	BMP2	T72	ZSU23-4	2S3	clutter
truck	489	1	0	1	2	0	2	0	5
SUV	0	499	0	0	0	0	0	0	1
BTR70	5	0	491	0	0	2	0	0	2
BRDM2	2	0	6	489	0	0	0	0	3
BMP2	4	0	2	0	482	0	0	0	12
T72	0	0	0	0	0	494	0	0	6
ZSU23-4	2	0	2	0	0	0	459	0	37
2S3	2	0	0	0	0	0	0	468	30
clutter	7	0	2	1	1	0	1	1	487

Conclusions

- **Proved the feasibility of an ATR algorithm built upon CS theory.**
  - **Targets can be detected and recognized at various ranges (1, 2 and 3km).**
- **Methodology is surprisingly excellent, especially at the longer range of 3 Km. where a human operator would not only have difficulty discerning the target but even more trouble distinguishing an armored vehicle from a truck.**

Issues

- **Algorithm assumes that data is available to construct robust dictionaries.**
  - **Target rotational information not always available.**
  - **What can we do if data is limited?**
- **Training data and test data are not always aligned.**
  - **Need invariant target representations.**

Naval Air Warfare Center Weapons Division – China Lake, CA.

Face Recognition and Learning via SIFT-OMP

Outline

- Project Overview
- ATR based on Compressed Sensing (CS) ideas
  - Sparse representation
  - Scale invariant transformation feature (SIFT)
  - Joint classification and learning

Face Recognition - Challenges/Objective

- Marines(Intelligence Plans and Policy Branch) are interested in face recognition systems that are suitable for portable devices.
- Challenges: facial expression, illumination, pose, scale changes, glasses, hats……
- Multiple samples for the varying conditions would be needed if a dictionary was to be implemented utilizing the previous approach.
  - –Single sample per person (SSPP) is a more realistic scenario.
- Objective is to develop a feature invariant algorithm for joint dictionary learning and classification

GOALS

- **Achieve a robust invariant face recognition system**
  - **Affine invariance: rotation, sheer, scale and translation**
  - **Illumination invariance**
  - **Appearance invariance**
- **Recognition based on a single sample per person (SSPP)**
- **Joint dictionary learning and classification**
  - **Take advantage of un-labeled data to improve/update dictionary knowledge**
- **Improve on current state of the art approaches to face recognition**

Sparse Representations

$$Y = AX$$

$$\begin{pmatrix} y_{11} & \cdots & y_{1n} \\ \vdots & \ddots & \vdots \\ y_{k1} & \cdots & y_{kn} \end{pmatrix} = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{k1} & \cdots & a_{kn} \end{pmatrix} \begin{pmatrix} x_{11} & \cdots & x_{1n} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nn} \end{pmatrix}$$

KxN
SIFT descriptors

KxN
Dictionary
SIFT descriptors

Nx1
Sparse vector
with **S**
nonzero entries

$$K = O(S(\log(N/S))$$

$$K << N$$

Solve undetermined system of equations by exploiting the sparse nature of the signal

APPROACH-RECOGNITION

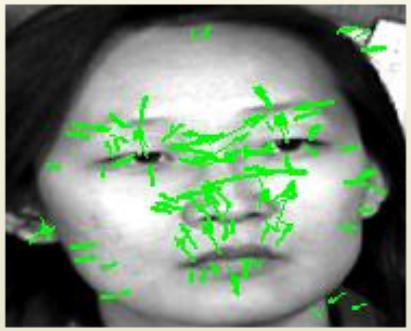
- **Sparse representation approach**
- **Recognition**
  - **Scale Feature Invariance Transformation (SIFT) dictionaries**
  - **SIFT descriptors are computed for each training image to form a sub-dictionary for each class.**
  - **Recognition is achieved by solving for the sparse solution via ℓ_1 minimization.**
  - **Voting system that selects the class with the most matching SIFT descriptors.**

APPROACH-LEARNING

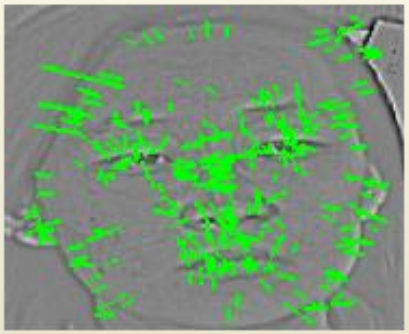
- **Adaptive learning is the natural process utilized by humans.**
  - **If we are sure that an object belongs to class “1” and it provides new information about the object, then we learned to associate the new information with class “1”.**
  - **SIFT descriptors for the same face from two different images(pose changes) will provide overlapping and non-overlapping information.**
    - **The new information needs to be incorporated into the dictionary.**
- **System must decide when and how to learn.**
  - **When: Variational Inference, K-Means and descriptive statistics.**
  - **How: Orthogonal learning**

FEATURE DESCRIPTOR (SIFT)

- SIFT detects and extracts local feature descriptors that are reasonable invariant to:
  - Changes in scale, 2D translation and rotation illumination, image noise and viewpoint.
- SIFT keypoint detector is not invariant to illumination changes.
- Highpass filtering + adaptive thresholding ensures adequate number of descriptors per image.

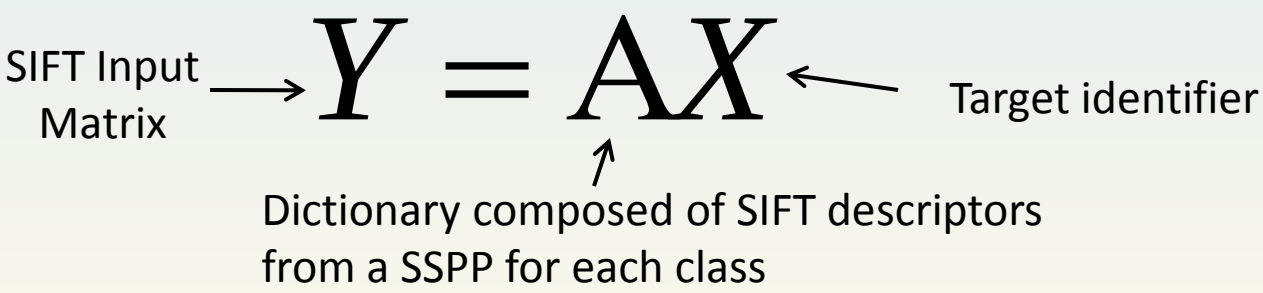


Keypoint selection with original SIFT method, 120 points



Keypoint selection with filtering + adaptive thresholding, 290 points

ATR in the CS Framework



- **A is an overcomplete dictionary containing SIFT samples for all target classes.**
  - **Seeking to identify the single group in A that approximates Y**
  - **Ideally only one group from the dictionary is needed to represent the input signal.**
  - **Solve the following L1 optimization problem.**

$$X_1 = \min \| X \|_1 \quad s.t. \quad AX = Y$$

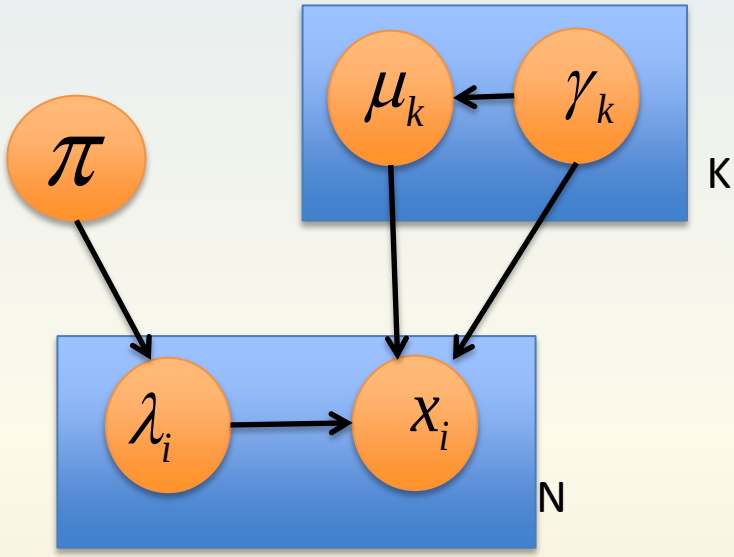
WHEN TO LEARN?

- Decision to learn is based on the ℓ_1 minimization output from all the classes.
- Max(SCI) is not sufficient to make the learning decision.
  - The wrong decision leads to corruption of dictionary class.
  - Learning should only be engaged when uncertainty in the answer is minimized.

WHEN TO LEARN?-APPROACHES

- Non-parametric approach - Variational Inference
  - Approximates a posterior distribution that represents a set of N observed data points.
- Kmeans – Assumes that number of clusters are known.
  - Iterative partitioning approach that minimizes the sum, over all clusters, of the within-cluster sums of point-to-cluster-centroid distances.
- Ad-hoc approach – Descriptive statistics (kurtosis)
  - Heavy tailed distributions are associated with high kurtosis

Variational Inference



Bayesian Network for a Probabilistic Model

$$P(x/\lambda,\{\mu_k\},\{\gamma_k\}) = \prod N(X/\mu_k,\gamma_k^{-1})^{\delta(\lambda,k)}$$

$$P(\mu_k,\gamma_k) = N(\mu_k/m,(\beta\gamma_k)^{-1})Gamma(\gamma_k/a,b)$$

$$P(\pi) = Dirichlet(\pi/\alpha)$$

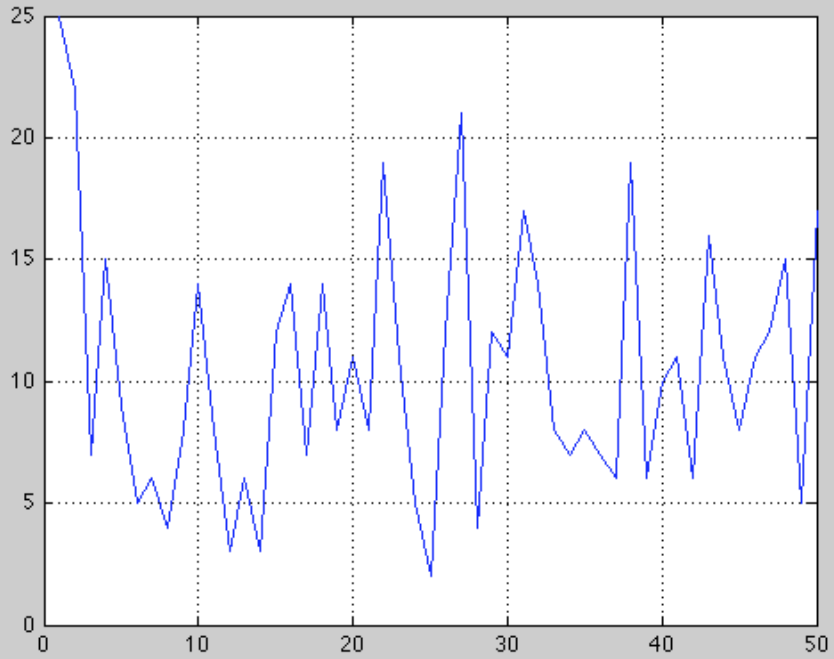
$$P(\lambda_i) = Multinomial(\pi)$$

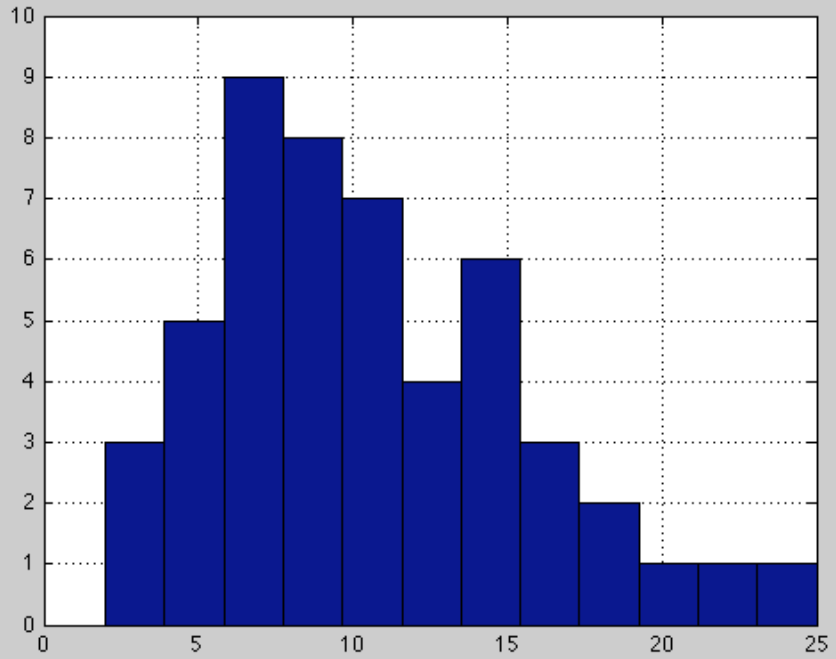
-Need to estimate the posterior distribution for each parameter from the data itself

$$P(H/X) = P(D/H)P(H)$$

$$P(\mu_k,\gamma_k/X,m,\beta^{-1}) \propto P(X/\mu_k,\gamma_k)P(\mu_k,\gamma_k/m,\beta^{-1})$$

How many clusters? VB vs. Kmeans

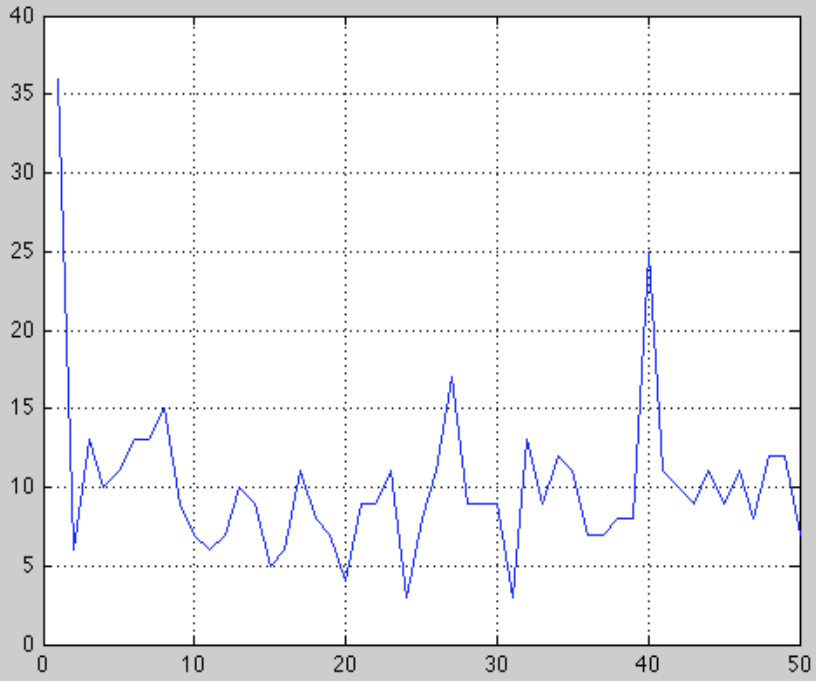


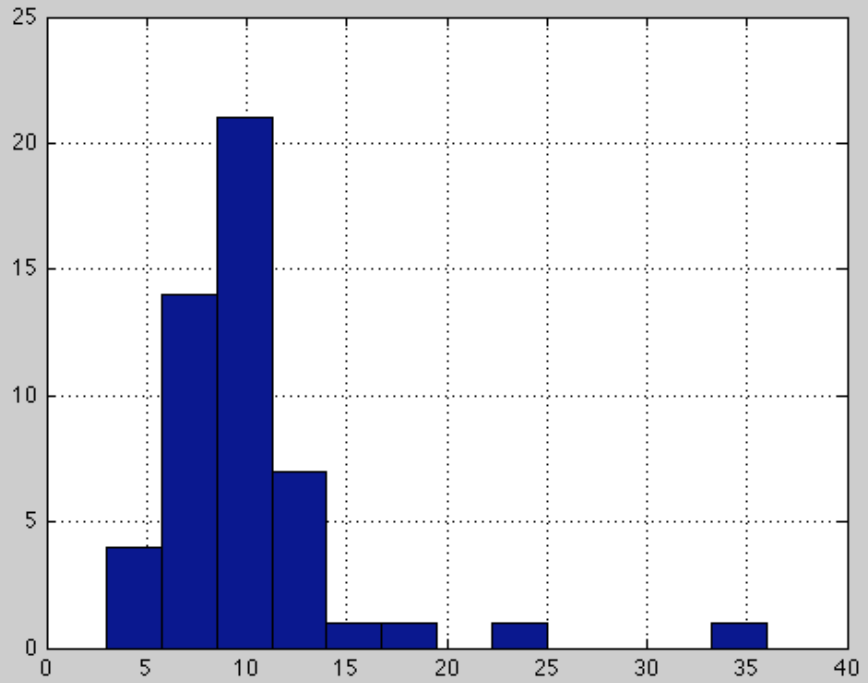


Variational Inference => One cluster, μ= 10.4, σ = 5.2

Kmeans=> Two clusters, 32, 18

How many clusters? VB vs. Kmeans

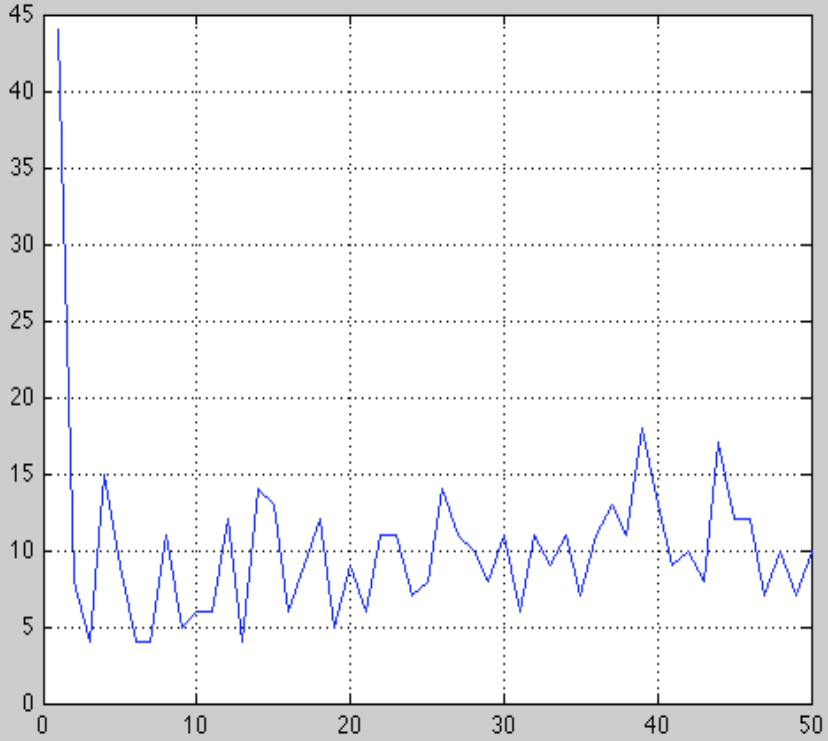


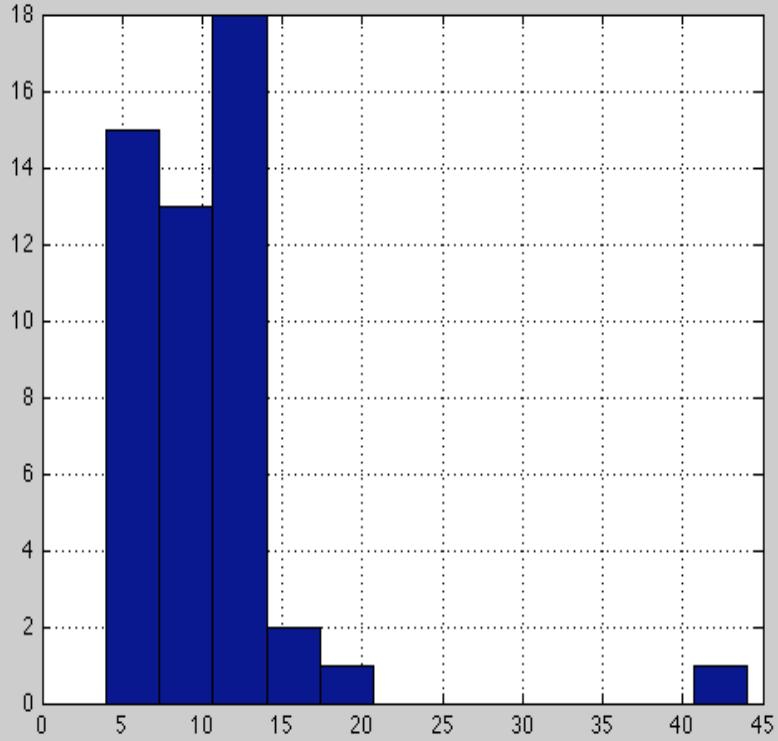


Variational Inference => Two clusters, n_1= 2, n_2= 48, μ_1= 30, μ_2= 9.2, σ_1= 5.8, σ_2= 2.8

Kmeans=> Two clusters , n_1 =22 , n_2= 28

How many clusters? VB vs. Kmeans





Variational Inference : Two clusters , n_1= 1, n_2= 49, μ_1= 44, μ_2= 9.4, σ_1= = 0.01, σ_2 = 3.3

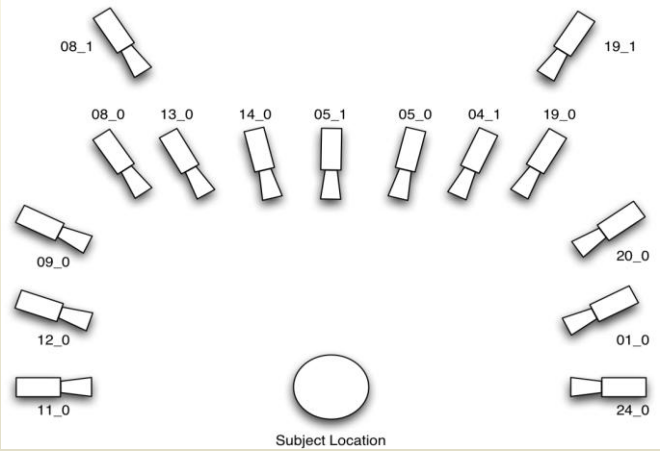
Kmeans=> Two clusters, n_1= 26 , n_2= 24, C1= 7.2, C2= 14

HOW TO LEARN?

- **Need to find new information which is not in the dictionary class. Need Orthogonal information to sub-dictionary class.**
- **Reverse Orthogonal Matching Pursuit (OMP) is currently utilized to select descriptors that are not in the selected class sub-dictionary**
  - **Solve for the complement of the ℓ_1 solution**

Data

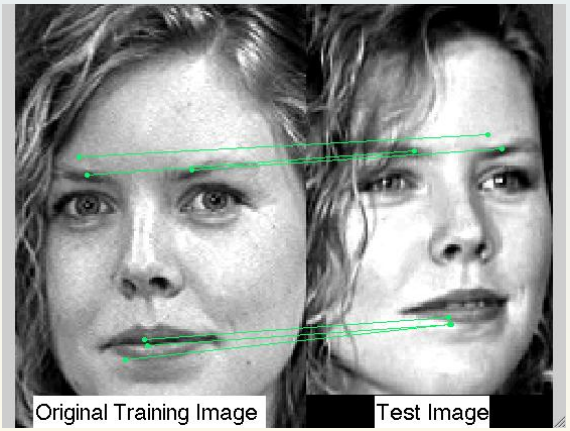
- **The Georgia Tech and the CMU Multi-PIE databases were used to evaluate the algorithm.**
  - **The GT database contains 50 subjects with images that include variations in illumination, facial expression, and appearance. The faces were captured at different scales and orientations.**
- **The Multi-PIE database contains 337 subjects, captured under 19 illumination conditions and 15 view points in four recording sessions for a total of more than 750,000 images.**

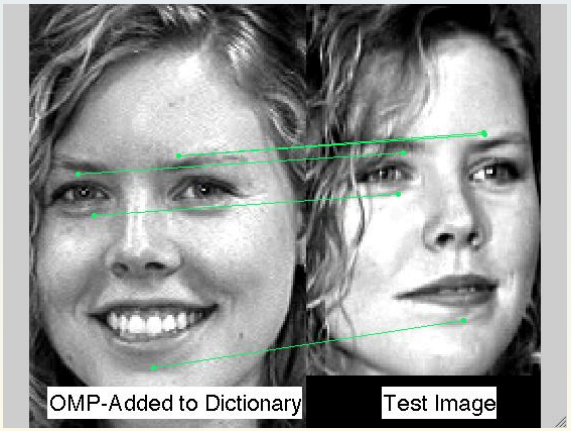


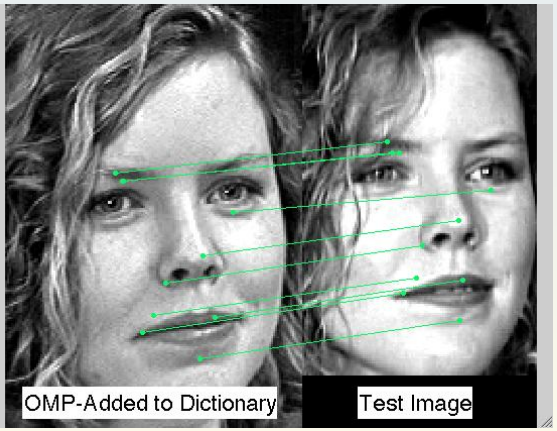
KEY RESULTS

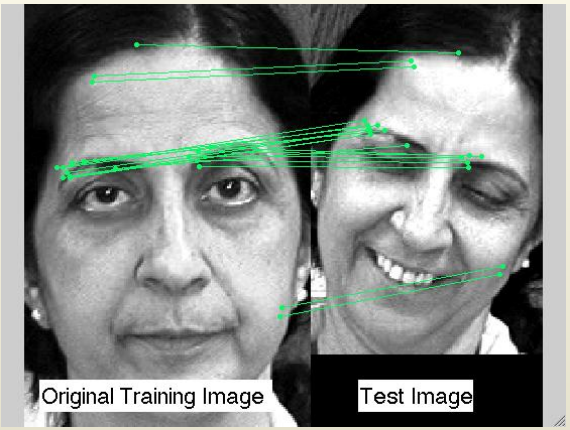
RECOGNITION INVARIANCE

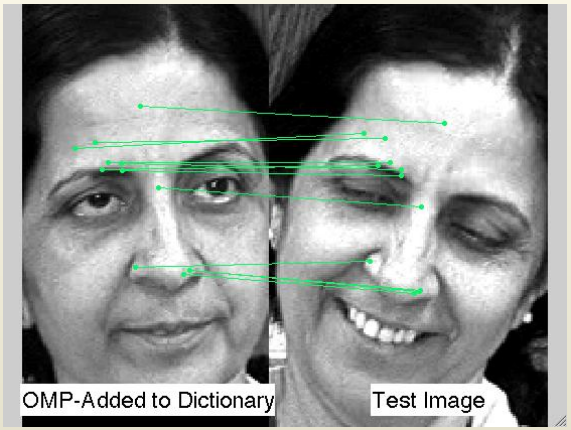
Joint Dictionary Learning and Classification

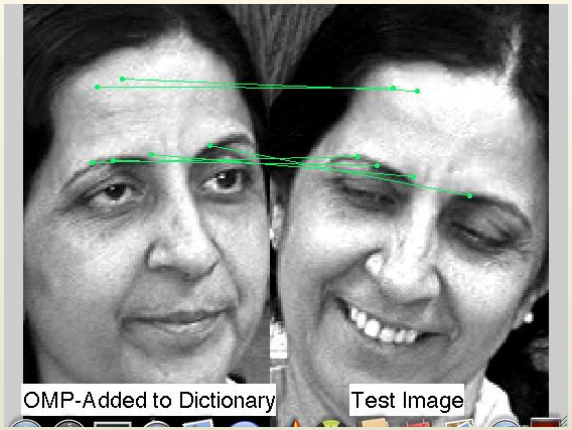












SUMMARY

- **New methodology to object recognition and learning.**
  - **Scale, pose, illumination and appearance invariance**
  - **Assumes minimum labeled data is available while it exploits unlabeled data to expand current knowledge**
- **Results are superior to those published in literature.**

Multi-PIE Database	Session2 (Frontal)	Session1 (30° pose change)	Training Samples
SIFT-OMP	95.8%	95%	1
[24]	95.2%	N/A	7
[4] S-SC	95%	N/A	7
[4] U-SC	94.6%	N/A	7
NS	77.6%	N/A	
NN	67.3%	N/A	
LDA	49.4%	N/A	

	GT Database	Training Samples
[23]	96.6%	5
SIFT-OMP	96.5 %	1

[1] Wagner, Yi Ma
[2] Yang, Yu, Huang
[NS], [NN], Near Sub-space, Lee
[LDA], Belhuemer
[3] Nefian