A Sparse Reconstruction Based Algorithm for Image and Video Classification
Tanaya Guha and Rabab Ward
Electrical and Computer Engineering, The University of British Columbia, Vancouver, Canada

Conclusion first:
- The proposed classification algorithm learns one dictionary for each class.
- Each sample is represented in terms of the reconstruction errors it produces w.r.t. each dictionary.
- This error based representation is highly discriminative and can be used as input to any traditional classifier.
- Since each dictionary is trained independently, the training process does not have to be repeated when a new class of data is added.
- Applicable to a wide variety of classification tasks involving both images and videos.

Introduction
Sparse representation has emerged as a key to successful pattern analysis and recognition. Its success largely depends on the choice of overcomplete dictionary.

How to build overcomplete dictionaries?
- use/combine predefined dictionaries
- use the training samples as dictionary elements — easy, large number of training samples needed, not easily extendable to video signals
- learn dictionary from data — not discriminative enough
- learn discriminative dictionary — adds up to the computation

Objective:
- Build a general framework for images and videos
- Enhance the discriminating power of learnt dictionaries

Proposed Approach

- Feature extraction
  - raw patches for images
  - spatio-temporal descriptors for videos [1]
- Random projection
- Learn N class-specific dictionaries using K-SVD algorithm [2]

\[
\min_{\Phi, Y} \| X - \Phi Y \|_F^2 \quad \text{s.t.} \quad \| Y \|_0 \leq \tau
\]

Each training sample is represented by a vector

\[
E = [e_1, e_2, \ldots, e_N]^T
\]

where \( e_i \) is the reconstruction error produced by the \( i \)th dictionary.

Nearest neighbor classification:
The distance between query (Q) and \( i \)th training sample is computed as

\[
d(Q, E_i) = \sqrt{(Q - E_i)^T L (Q - E_i)}
\]

where L is a class of Mahalanobis distance metric learned from the data fed to the classifier [3].

Acknowledgments
This research was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) and the Canada Research Chair in Multimedia Pattern Recognition.

References

Related work

Contact
Tanaya Guha, PhD Candidate
Image and Signal Processing Lab, ECE, UBC
email: tanaya@ece.ubc.ca
web: http://ece.ubc.ca/~tanaya

Results

- Face recognition
  - AT&T Face dataset: Benchmark, 400 frontal faces, 15 classes, variations in illumination, expression and facial details
  - Sample images from the AT&T Face dataset
  - Results on the AT&T Face dataset

- Biological species identification
  - Nematodes dataset: wormlike animal of high commercial and medical importance, diverse species, difficult to classify 50 images, 5 classes
  - Sample images from the Nematodes dataset
  - Results on the Nematodes dataset

- Action recognition in videos
  - UCF sports dataset: Challenging real videos, TV broadcast, occlusion, cluttered background, viewpoint variation
  - Sample frames from the UCF Sports dataset
  - Results on the UCF dataset

Contact
Tanaya Guha, PhD Candidate
Image and Signal Processing Lab, ECE, UBC
email: tanaya@ece.ubc.ca
web: http://ece.ubc.ca/~tanaya