Joint and Discriminative Dictionary Learning for Facial Expression Recognition

Sriram Kumar, Behnaz Ghoraani, Andreas Savakis
Facial Expression Recognition

• **Applications** :
  Human-computer interaction driver monitoring, health and wellness, entertainment, surveillance and others.

• **Challenges** :
  similarities in facial appearance may interfere with the recognition of facial expressions.

Sample facial expression dataset illustrating: anger, disgust, fear, happy, sad, surprise. (top to bottom, left to right)
Sparse Representation

- Let a sample $x \in \mathbb{R}^d$ be represented on a dictionary of samples $D \in \mathbb{R}^{d \times p}$ via the sparse coefficients $a$, as follows:

$$x = D a$$
Dictionary Learning

• Motivation:
  ▪ SR process can become time consuming and even unstable.
  ▪ Dictionary is not optimized to represent the data efficiently.
  ▪ Better SR can be harnessed when the dictionary is learned from the data [1].

Dictionary Learning

• Fisher Discrimination Dictionary Learning [2]

• Dictionary Learning Separating Commonality and Particularity [3]

K-SVD

• Learns an over-complete dictionary from the training data

• K-SVD is an iterative technique wherein,
  1. training samples are first sparsely coded using the current dictionary estimate
  2. then dictionary elements are updated one at a time while keeping the remaining atoms fixed

• Can be efficiently implemented using Batch Orthogonal Matching Pursuit [4].

The objective function tries to minimize the reconstruction error as follows:

\[ \{D, a\} = \arg \min_{D, a} \| x - Da \|_2^2 \quad \text{s.t.} \quad \| a \|_0 \leq \delta \]

- \( D \) is the learnt dictionary and \( a \) are the sparse coefficients.
- \( \| a \|_0 \) norm which counts the number of non-zero elements in the coefficient vector
- \( \delta \) controls the amount of sparsity in the coefficient vector \( a \).
K-SVD

• **Limitation:**
  • The atoms present in the dictionary aren’t necessarily discriminative
  • This might cause the sparse coefficients to be less discriminative

• **Solution:**
  • Discriminative dictionary learning framework
Fisher Discrimination Dictionary Learning

• This framework jointly learns a dictionary and discriminative sparse codes based on Fisher's Discrimination criterion.
  
  • Let $\mathbf{D} = [\mathbf{D}_1, \mathbf{D}_2, ..., \mathbf{D}_c]$ be the class specific dictionaries that is learnt and $c$ is the total number of classes.

• Let $\mathbf{A}$ be the sparse codes.
Fisher Discrimination Dictionary Learning

• The objective function $J_{(D,A)}$ is given by

$$J_{(D,A)} = \arg\min_{(D,A)} \left\{ \sum_{i=1}^{c} r(X_i, D, A_i) + \lambda_1 \|A\|_1 + \lambda_2 f(A) \right\}$$

  - $r(X_i, D, A_i)$ is the dictionary learning function
  - $\|A\|_1$ norm promoting sparsity
  - $f(A)$ is a function promoting discriminative sparse codes.
  - $\lambda_1, \lambda_2$ parameters controlling the amount of sparsity & discrimination
Fisher Discrimination Dictionary Learning

- The dictionary learning function \( r(X_i, D, A_i) \) is defined as,

\[
r(X_i, D, A_i) = \|X_i - DA_i\|_F^2 + \|X_i - D_iA_i^i\|_F^2 + \sum_{j=1}^{c} \|D_jA^j_i\|_F^2
\]

- \( \|X_i - DA_i\|_F^2 \) corresponds to the total reconstruction error.

- \( \|X_i - D_iA_i^i\|_F^2 \) corresponds to the class specific reconstruction error.

- \( \|D_jA^j_i\|_F^2 \) enforces the sparse codes to be representative of \( c^{th} \) class alone.
Fisher Discrimination Dictionary Learning

- The function promoting discriminative sparse codes is defined as,

\[ f(A) = tr(S_W(A) - S_B(A)) + \eta \|A\|_F^2 \]

- \( S_W(A) \) represents the within class scatter matrices of the sparse codes
- \( S_B(A) \) represents the between class scatter matrices of the sparse codes
- \( \eta \) is a regularization parameter
- \( \|A\|_F^2 \) enforces convexity
Fisher Discrimination Dictionary Learning

• **Limitation:**
  • Two atoms present in the sub-dictionaries might be correlated

• **Solution:**
  • Dictionary learning framework that learns a shared dictionary
Dictionary Learning Separating Commonality and Particularity

- The objective function of the DL-COPAR framework is given by,

\[
J = \sum_{c=1}^{C} \left\{ \|X_c - DA_c\|_F^2 + \|\widetilde{Q}^T_c A_c\|_F^2 \right. \\
\left. + \|X_c - D\tilde{Q} \tilde{Q}^T A_c\|_F^2 + \lambda \phi(A_c) \right\} + \eta \sum_{c=1}^{C+1} \sum_{j=1}^{C+1} Q(D_c, D_j)
\]

Where,

\[
Q_c = [q^c_1, ..., q^c_j, ..., q^c_{K_c}] \in \mathbb{R}^{K \times K_c}
\]

\[
\widetilde{Q}/c = [Q_1, ..., Q_{c-1}, Q_{c+1}, ..., Q_C]
\]

Selection operator
Dictionary Learning Separating Commonality and Particularity

\[ J = \sum_{c=1}^{C} \left\{ \| X_c - DA_c \|_F^2 + \| \tilde{Q}_c^T A_c \|_F^2 \right\} + \| X_c - \tilde{D}\tilde{Q} \tilde{Q}_c^T A_c \|_F^2 + \lambda \phi(A_c) \left\{ + \eta \sum_{c=1}^{C+1} \sum_{j=1}^{C+1} Q(D_c, D_j) \right\} \]

- \( \| X_c - DA_c \|_F^2 \) corresponds to the total reconstruction error
- \( \| \tilde{Q}_c^T A_c \|_F^2 \) enforces the sparse codes to be representative of \( c^{th} \) class alone
- \( \| X_c - \tilde{D}\tilde{Q} \tilde{Q}_c^T A_c \|_F^2 \) corresponds to class specific reconstruction error
Dictionary Learning Separating Commonality and Particularity

$$J = \sum_{c=1}^{C} \left\{ \left\| X_c - DA_c \right\|_F^2 + \left\| \tilde{Q}_{/c} A_c \right\|_F^2 \right\} + \eta \sum_{c=1}^{C+1} \sum_{j=1 \atop j \neq c}^{C+1} Q(D_c, D_j)$$

- $\emptyset(A_c)$ is the regularization function to enforce sparse solution
- $Q(D_c, D_j) = \left\| D_c^T D_j \right\|_F^2$ measures if two atoms are similar
- $\lambda$ and $\eta$ are the parameters that control the amount of regularization
Dictionary Learning Separating Commonality and Particularity

\[ J = \sum_{c=1}^{C} \left\{ \|X_c - DA_c\|_F^2 + \|\tilde{Q}^T/A_c\|_F^2 \right\} + \eta \sum_{c=1}^{C+1} \sum_{j \neq c}^{C+1} Q(D_c, D_j) \]

- \( Q_c = [q_1^c, q_j^c, \ldots, q_{K_c}^c] \epsilon \mathbb{R}^{K_c \times K_c} \), where the \( j^{th} \) column of \( Q_c \) is a vector of zeros except for a value of 1 at the \( j^{th} \) location such that \( Q_c^T Q_c = I \)

- \( Q/c = [Q_1, Q_{c-1}, \ldots, Q_c] \)

- \( \tilde{Q}_c = [Q_c, Q_{c+1}] \), where \( Q_c, Q_{c+1} \) correspond to the selection operator for the class specific dictionary and the common dictionary respectively.
Classification

**K-SVD**

Reconstruction error for each sub-dictionary is computed to determine the class label.

Ridge Regression based classification: \[ C = (AA^T)^{-1}A^TH \]

\( H \in \mathbb{R}^{C \times N} \) is a sparse ground truth matrix and \( N \) is the total no. of samples. Each column of corresponds to a training sample, where the element is set to belongs to that class, or 0 otherwise.

**FDDL**

Reconstruction error for each sub-dictionary is computed to determine the class label.

**COPAR**

Reconstruction error for each sub-dictionary with shared dictionary atoms is computed to determine the class label.

**Local Sparse Coding**
Visualizations of atoms learnt by K-SVD

Dictionary with 150 atoms
Visualizations of atoms learnt by FDDL

Dictionary with 30 atoms/class
Visualizations of atoms learnt by COPAR

Dictionary with 30 atoms/class + 10 shared atoms

HAPPY

SAD

ANGRY

FEAR

SURPRISE

SHARED DICTIONARY
Dataset

• The extended CK+ [5] expression dataset contains 118 subjects in 327 video sequences exhibiting the expressions of anger, disgust, fear, happiness, sadness, surprise, and contempt.
• For each expression sequence, the last six frames were extracted which contained the onset of the expression to the peak expression.
• The images were resized to $24 \times 21$, and normalized such that they had unit $L_2$ norm.

## Results

<table>
<thead>
<tr>
<th>Method</th>
<th>Number of Expressions</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>K-SVD</td>
<td></td>
<td>93.0</td>
</tr>
<tr>
<td>FDDL</td>
<td>Five Expressions</td>
<td>95.0</td>
</tr>
<tr>
<td>DL-COPAR</td>
<td></td>
<td>99.0</td>
</tr>
<tr>
<td>DL-COPAR</td>
<td>Six Expressions</td>
<td>98.1</td>
</tr>
<tr>
<td>MSR [4]</td>
<td>Six Expressions</td>
<td>94.6</td>
</tr>
</tbody>
</table>

---

Thank You