Discriminative Common Vectors with Kernels

Review

Miguel A. Veganzones

Grupo de Inteligencia Computacional
Universidad del País Vasco

Outline

1. Introduction
2. FLD's modifications
3. Discriminative Common Vector
   - Introduction
   - DCV by using the Range Space of S_W
   - DCV by using Difference Subspaces and the Gram-Schmidt Orthogonalization Procedure
4. Kernel DCV

Context

- Face recognition problem.
- Each image is represented by a vector in a wh-dimensional space.
- This space is called the sample space or the image space, and its dimension is typically very high.
- There is redundant information.

Objective

- Find a subspaces based features extraction method that could success under the small sample size problem.
- Small sample size problem: when sample space dimensionality is larger than the number of samples in the training set.
- Subspaces based common methods:
  - Principal Component Analysis (PCA) -> unsupervised
  - Fisher's Linear Discriminant (FLD) -> supervised

Principal Component Analysis (PCA)

- Projections that maximize the total scatter matrix (covariances), S_T.

$$J_{PCA}(W_{opt}) = \arg\max_{W} \left|W^T S_T W\right|$$

- The maximum is given by the eigenvectors of S_T.
- The projection directions are also called eigenfaces. Any face image in the sample space can be approximated by a linear combination of the significant eigenfaces.
- Tends to model unwanted within-class variations (lighting, expressions, occlusions,...) and the resulting classes tend to have more overlapping than other approaches.

Fisher's Linear Discriminant (FLD)

- Overcomes the limitations of the Eigenfaces method.
- Projections that maximize the between class scatter matrix, S_B, and minimize the within class scatter matrix, S_W.

$$J_{FLD}(W_{opt}) = \arg\max_{W} \frac{|W^T S_B W|}{|W^T S_W W|}$$

- The maximum is given by the eigenvectors of $S_W^{-1} S_B$.
- Not applicable within "small sample size problem" because S_W is singular in this case.

Scatter matrices

- Between class scatter matrix:

$$S_B = \sum_{i=1}^{C} N_i(\mu_i - \mu)(\mu_i - \mu)^T$$

- Within class scatter matrix:

$$S_W = \sum_{i=1}^{C} \sum_{m=1}^{N_i} (x_m^i - \mu_i)(x_m^i - \mu_i)^T$$

- Total scatter matrix:

$$S_T = S_B + S_W = \sum_{i=1}^{C} \sum_{m=1}^{N_i} (x_m^i - \mu)(x_m^i - \mu)^T$$

Road map

- FLD Modifications
- Discriminative Common Vector Method (DCV)
- Rough Common Vector Method (RCV)
- Discriminative Common Vector with Kernels (KDCV)

Some FLD-based methods

- Pseudoinverse method: replacing S_W^{-1} by its pseudoinverse.
- Perturbation method: adding a small perturbation matrix Δ to S_W in order to make it non-singular.
- Rank Decomposition method: making successive eigen-decompositions of the total scatter matrix S_T and the between class scatter matrix S_B.
- They are computationally expensive since the scatter matrices are very large.

Fisherface method

- Two stage method: PCA + Linear Discriminant Analysis.
- PCA is used to reduce data dimensionality so as to make S_W non-singular.
- By PCA use some directions corresponding to the small eigenvalues of S_T are thrown away, removing dimensions with potential discriminative information.

Null Space Method

- Based on the modified FLD criterion:

$$J_{MFLD}(W_{opt}) = \arg\max_{W} \frac{|W^T S_B W|}{|W^T S_T W|}$$

- This method has been proposed to be used when the dimension of the sample space is larger than the rank of S_W.
- The MFLD criterion attains its maximum when all image samples are projected onto the null space of S_W, and then PCA is applied to the projected samples to obtain the optimal projection vectors.
- The performance of the Null Space method improves if the null space of S_W is large.
- There is not an efficient algorithm for applying this method in the original sample space.

PCA + Null Space Method

- PCA is applied to remove the null space of S_T, which contains the intersection of the null spaces of S_W and S_B.
- Then the optimal projection vectors are found in the remaining lower dimensional space by Null Space method.
- The difference with the Fisherface method is that, here S_W is typically singular in the reduced space because all eigenvectors corresponding to the non-zero eigenvalues of S_T are used for dimension reduction.

Direct-LDA method

- Uses the simultaneous diagonalization method.
- First, the null space of S_B is removed and then, the projection vectors that minimize S_W in the transformed space are selected from the range space of S_B.
- Removing the null space of S_B by dimensionality reduction will also remove part of the null space of S_W removing important discriminant information.
- Futhermore, the whitening proccess over S_B is redundant.

Comparisons

Table: Comparisons of performance across methods for $n > C - 1$

Rank	Accuracy	Training Time	Testing Time	Storage Requirements
1	DCV, PCA + Null Space	Direct-LDA	DCV, PCA + Null Space	DCV, PCA + Null Space
2	Fisherface	DCV	Fisherface, Direct-LDA	Fisherface, Direct-LDA
3	Direct-LDA	Eigenface	Eigenface	Eigenface
4	Eigenface	Fisherface		
5		PCA + Null Space		

Outline

Introduction

- DCV addresses the limitations of previous methods that use the null space of S_W to find the optimal projection vectors.
- It can be only used when the dimension of the sample space is larger than the rank of S_W.
- This approach extracts the common properties of classes in the training set by eliminating the differences of the samples in each class.

Algorithms

- Previous works in word recognition obtain a common vector for each class by removing all the features in the direction of the eigenvectors corresponding to the non-zero eigenvalues of the scatter matrix of its own class.
- Cevikalp's work describes two algorithms to obtain DCV for face recognition:
  - Instead of using a given class's own scatter matrix, he uses the within-classes scatter matrix of all classes to obtain the common vector.
  - He gives an alternative algorithm based on the subspace methods and the Gram-Schmidt orthogonalization procedure.

Outline

S_W Null Space based criterion

- In the special case where $w^T S_W w = 0$ and $w^T S_B w \neq 0$ for all $w \in R^d \setminus \{0\}$, the modified FLD criterion attains a maximum.
- A projection vector w satisfying the above conditions does not necessarily maximizes the between-class scatter matrix. In this case, a better criterion is given by:

$$J(W_{opt}) = \arg \max_{|W^T S_W W = 0|} \left|W^T S_B W\right| = \arg \max_{|W^T S_W W = 0|} \left|W^T S_T W\right|$$

Direct algorithm

- To find the optimal projection vectors w in the null space of S_W, the face samples are projected onto the null space of S_W and then, the projection vectors are obtained by PCA.
- However, this task is computationally intractable since the dimension of the null space can be very large.
- A more efficient way of doing it is by using the orthogonal complement of the null space of S_W, which typically is significantly lower-dimensional space.

Feasible algorithm

Description

- Let R^d be the original sample space, V be the range space of S_W, and $V^\perp$ be the null space of S_W:

$$V = span\{\alpha_k | S_W \alpha_k \neq 0, \qquad k = 1, \ldots, r\}$$

$$V = span\{\alpha_k | S_W \alpha_k = 0, \qquad k = r+1, \ldots, d\}$$

- Where:
  - $r < d$ is the rank of S_W
  - $\{\alpha_1, \ldots, \alpha_d\}$ is an orthonormal set, and $\{\alpha_1, \ldots, \alpha_r\}$ is the set of orthonormal eigenvectors corresponding to the non-zero eigenvalues of S_W.

Feasible algorithm
Goal

- Considering the matrices $Q = [\alpha_1 \ldots \alpha_r]$ and $\tilde{Q} = [\alpha_{r+1} \ldots \alpha_d]$.
- Since $R^d = V \oplus V^{\perp}$, every face image $x_m^i \in R^d$ has a unique decomposition of the form

$$x_m^i = y_m^i + z_m^i$$

  - where $y_m^i = Px_m^i = QQ^T x_m^i \in V$, $z_m^i = \tilde{P}x_m^i = \tilde{Q}\tilde{Q}^T x_m^i \in V^{\perp}$ and P and $\tilde{P}$ are the projection operators onto V and $V^{\perp}$ respectively.
- The goal is to compute:

$$z_m^i = x_m^i - y_m^i = x_m^i - Px_m^i$$

Common vectors

- The eigenvectors can be obtained from the M by M matrix, $A^T A$ where A is a d by M matrix of the form

$$A = \left[x_1^1 - \mu_1 \ldots x_N^1 - \mu_1 x_1^2 - \mu_2 \ldots x_N^C - \mu_C\right]$$

- Let λ_k and v_k be the kth non-zero eigenvalue and the corresponding eigenvector of $A^T A$. Then, $\alpha_k = Av_k$ will be the eigenvector that corresponds to the kth non-zero eigenvalue of S_W.
- It turns out that we obtain the same unique vector for all samples of the same class, which are defined as the common vectors:

$$x_{com}^i = x_m^i - QQ^T x_m^i = \tilde{Q}\tilde{Q}^T x_m^i, \qquad m = 1, \ldots, N; i = 1, \ldots, C$$

Maximizing criterion

- After obtaining the common vectors x^i_{com}, optimal projection vectors will be those that maximize the scattering of the common vectors:

$$J(W_{opt}) = \arg \max_{|W^T S_W W = 0|} \left| W^T S_B W \right| = \arg \max_{|W^T S_W W = 0|} \left| W^T S_T W \right| = \arg \max_{W} \left| W^T S \right.$$

- W is a matrix whose columns are the orthonormal optimal projection vectors w_k, and S_{com} is the scatter matrix of the common vectors

$$S_{com} = \sum_{i=1}^{C} (x^i_{com} - \mu_{com})(x^i_{com} - \mu_{com})^T, \qquad i = 1, \ldots, C$$

Obtaining optimal projections

- All eigenvectors corresponding to the non-zero eigenvalues of S_{com} will be the optimal projection vectors.
- Instead of using S_{com} that is typically a large d by d matrix, the smaller matrix $A^T_{com}A_{com}$ of size C by C can be used, where

$$A_{com} = \left[x^1_{com} - \mu_{com} \ldots x^C_{com} - \mu_{com}\right]$$

- Each class is discriminated by a discriminative common vector:

$$\Omega_i = W^T x^i_m, \quad m = 1, \ldots, N; i = 1, \ldots, C$$

- To recognize a test image x_{test}, the feature vector of this image is found by $\Omega_{test} = W^T x_{test}$, and the Euclidean distance to each class's discriminative common vector gives the classification.

Algorithm

- Step 1: compute the non-zero eigenvalues and corresponding eigenvectors of S_W by using the matrix $A^T A$. Set $Q = [\alpha_1 \dots \alpha_r]$ where r is the rank of S_W.
- Step 2: choose any sample from each class and project it onto the null space of S_W to obtain the common vectors.
- Step 3: compute the eigenvectors w_k with non-zero eigenvalues of the matrix $A_{com}^T A_{com}$. Use these eigenvectors to form the projection matrix $W = [w_1 \dots w_{C-1}]$.

Outline

Summary

Questions?

Thank you very much for your attention.

- Contact:
  - Miguel Angel Veganzones
  - Grupo Inteligencia Computacional
  - Universidad del País Vasco - UPV/EHU (Spain)
  - E-mail: miguelangel.veganzones@ehu.es
  - Web page: http://www.ehu.es/computationalintelligence