Face Recognition in the wild

Tiecheng Su
Department of Electrical and Computer Engineering
University of Rochester
tsu5@ur.rochester.edu

Machine vision, 2016

Dataset

Labeled Faces in the Wild(LFW)dataset

- large-scale:13K images, 5.7K people
- high variability in appearance
- data pruning:
  only choose person has more than 10 images, ends up with 127 people
- collected using Viola-Jones face detector
- split into training set and test with ratio 8:2

Objective

Face descriptor for recognition

- dense sampling
- relevant face parts learnt automatically
- compact and discriminative

pipeline: dense SIFT $\rightarrow$ Fisher vector encoding$\rightarrow$ linear SVM

Dense Features

face image $\rightarrow$ set of local features

Dense SIFT

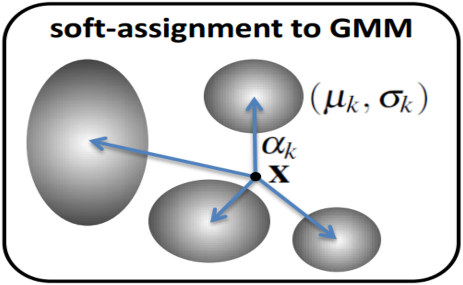
- dense scale-space grid:
  2 pix step, 5 scales
- 24×24 patch size
- SIFT
- 64-D PCA-SIFT
- augmented with (x,y):66-D

Face Fisher Vector

set of local features $\rightarrow$ high-dim Fisher vector

Fisher Vector(FV)encoding

- describes a set of local features in a single vector
- diagonal-covariance GMM as a codebook
  - appearance:SIFT
  - location:(x,y)
- Initializing a GMM model with K-means before running EM



Face Fisher Vector

set of local features $\rightarrow$ high-dim Fisher vector

- Feature FV-feature space location statistics
  1^{st} order stats(k-th Gaussian):$\Phi_k^1 \sim \alpha_k(\frac{x_p - \mu_k}{\delta_k})$
  2^{nd} order stats(k-th Gaussian):$\Phi_k^2 \sim \alpha_k(\frac{x_p - \mu_k}{\delta_k^2} - 1)$

$$\Downarrow$$

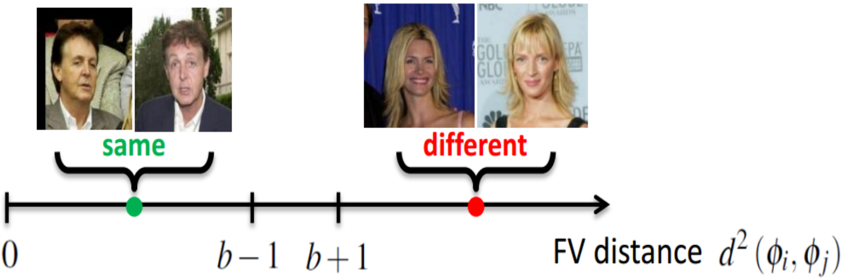
$$\phi = [\underbrace{\Phi_1^{(1)}}_{\text{66-D}}, \underbrace{\Phi_1^{(2)}}_{\text{66-D}}, ..., \Phi_K^{(1)}, \Phi_K^{(2)}]$$

FV dimensionality:$66 \times 2 \times 256 = 33792$
(for a mixture of 256 Gaussians)

Distance Learning

high-dim FV $\rightarrow$ low-dim face descriptor

- Large-margin distance constraints: $y_{ij}(b - d^2_W(\Phi_i, \Phi_j)) > 1$
  $y_{ij} = 1$ if (i,j) is the same person, Φ_i, Φ_j-FV



Karen Simonyan,Omkar M. Parkhi,Andrea Vedaldi,Andrew Zisserman. Fisher Vector Faces in the Wild

Projection Learning

- Low-rank Mahalanobis distance:
  $d^2_W(\Phi_i, \Phi_j) = \|W\Phi_i - W\Phi_j\|_2^2 = (\Phi_i - \Phi_j)^T W^T W(\Phi_i - \Phi_j)$
- Large-margin objective:
  $\mathrm{argmin}_{W,b} \sum_{i,j} \max[1 - y_{ij}(b - \Phi_i - \Phi_j)^T W^T W(\Phi_i - \Phi_j), 0]$
  - regularization by $W \in R^{p\times d}$, $p \ll d$
  - stochastic gradient descent
  - initialized by PCA-whitening

pros

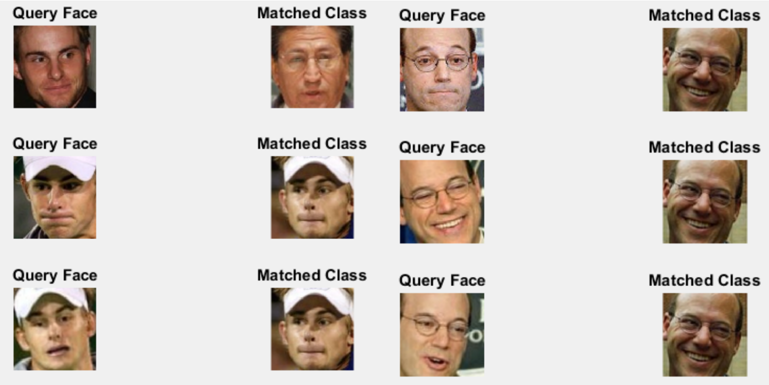
- Models dependencies between FV elements
- Explicit dimensionality reduction

cons

- Non-convex

Results

SIFT density	GMM Size	Desc. Dim.	Distance Function	accuracy,%
2 pix	256	33792	diag. metric	79.1
2 pix	256	128	low-rank Mah. metric	83.7



Summary

- Fisher Vector Face(FVF) representation
  - dense SIFT(no need for sophisticated landmark detectors)
  - Fisher vector
  - discriminative dimensionality reduction
- Future works
  - multi-feature image representations