Undersampled Face Recognition Via RADL

- → Summary
- → Benefits
- → Related work
- → Flow chart
- → Database.
- → Classification Formula → The Proposed Approach
- → Graph Comparision →
  - → Residual Function
  - → Error Function
- → RADL → Steps for Auxiliary Dictionary Learning.

Undersampled Face Recognition Via Robust Auxiliary Dictionary Learning

Summary

→ This approach → recognize face with few training images available per subject.

→ Intra & Interclass Variations Can be Successfully handled.

→ It handles → Corrupted regions in test image by learning Robust Auxiliary Dictionary from subjects not of interest.

Benefits of this approach

→ Only one or few training images per subjects are required.

→ It provides new tool, for - recognizing occluded face images by means of robust sparse coding and auxiliary dictionary learning.

→ It allows one to model intra-class variations including illumination and expression changes from external data.

SRC and Extended SRC

Assumption in SRC

* Large amount of training Data as over-complete Dictionary.

→ Thus, It will not work for Undersampled Face Recognition.

Extended SRC (ESRC)

$$\min_{X} \left\| y - [D, A] \begin{bmatrix} x_d \\ x_a \end{bmatrix} \right\|_2^2 + \lambda \| x$$

→ each Subject in D has one or few images

A → intra class Variation Dictionar

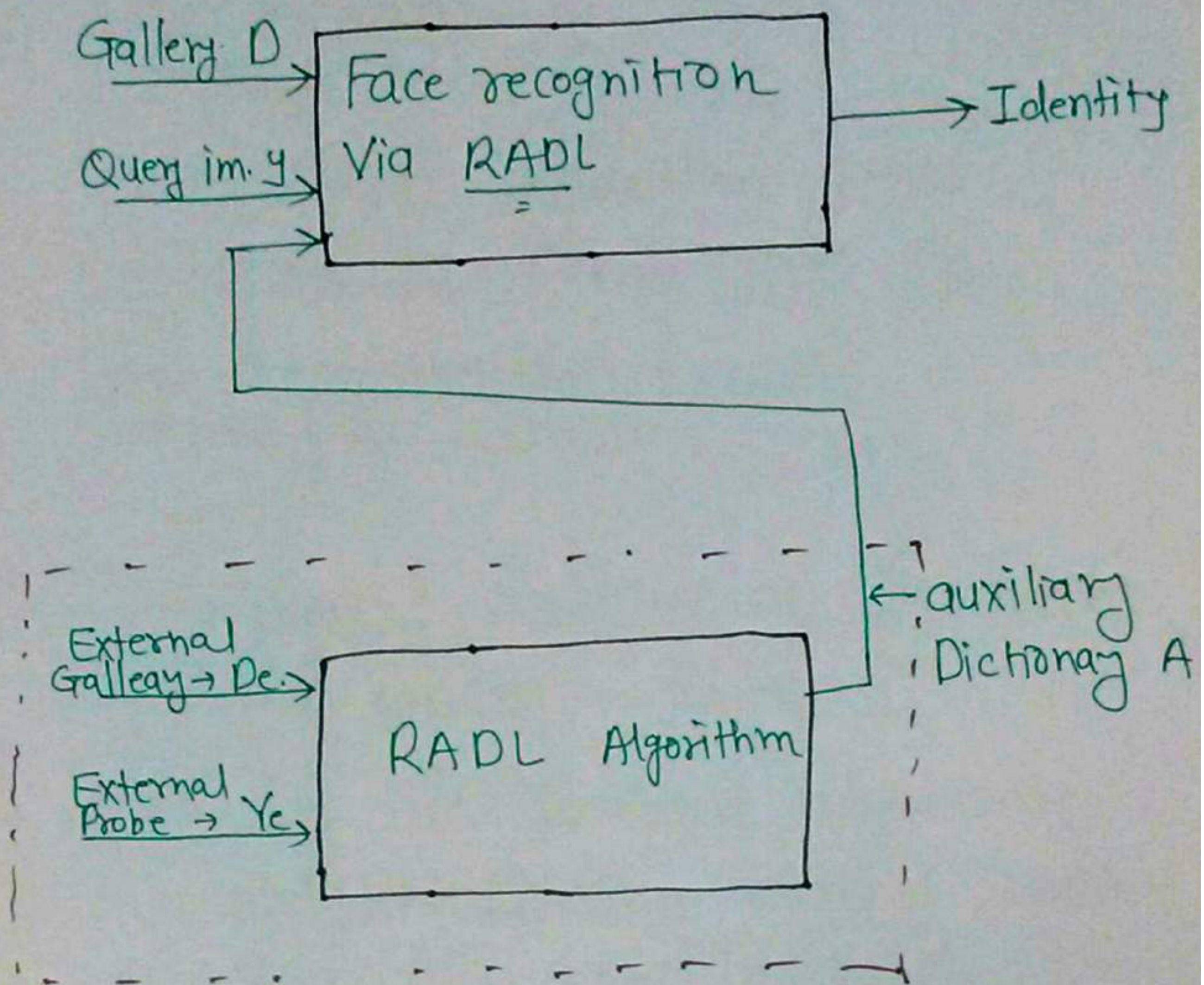
A → image data collected from external dataset. (Subject not of interest)

Problems with ESRC

1) → It directy apply External Data as A.
↳ might be noisy

2) → Computation would be Very expensive → due to large size of A. Because A should have all intra-class variations of inte-es

3) It assumes the type of occlusion to be Known when collecting External Data.


Gallery D
Query im. y
Face recognition Via RADL
Identity
auxiliary Dictionary A
External Galleay → De
External Probe → Ye
RADL Algorithm

Flowchart

Database

38 subjects → 64 images of each.
frontal face

→ images were downsampled before experiment → 34 X 30

→ ~~38~~ 32 subjects → ~~for~~ from database for recognition

→ 6 subjects for → external data. → for robust auxiliary dictionary learning

---

→ 3 images from each 32 subjects were selected to form → gallery → D

↳ remaining ~~64~~ 61 for testing.

→ correspond to 3 illumination conditions

Same condition 3 images were ~~ch~~ selected to form De. → $D_e \to \underbrace{6 \times 3}_{18 \text{ images}}$

→ Ye → random selection of 29 images from remaing images of 6 subjects

Face Recognition Via RADL

Classification Formulation

$y \in R^{d} \rightarrow$ query image

$D \in R^{d \times n} \rightarrow$ gallery matrix

$D = [D_1, D_2, \ldots D_L]$

↳ data matrixces from L Classes

$A \rightarrow \in R^{d \times m} \rightarrow$ auxiliary dictionary learned from external data.

the minimization Problem

$$\min_{x} \rho\left(y - [D, A]\begin{bmatrix} x_d \\ x_a \end{bmatrix}\right) + \lambda \|x\|_1$$

$\rho(\cdot) : R^d \to R$ is residual function

$$\rho(e) = \sum_{k=1}^{d} \rho(e_k)$$

$$\rho(e_k) = \frac{-1}{2\mu}\left(\ln\left(1+\exp(-\mu e_k^2 + \mu\delta)\right] - \ln(1+\exp\mu\delta)\right)$$

$e_k \to$ kth entry of

$e = y - [D, A]x$,

This type of Residual function have shown Promising Results in recent literatures of robust face recognition.

Optimization

:

$$\min_{X} \rho\left(y - [D, A]\begin{bmatrix} X_d \\ X_a \end{bmatrix}\right) + \lambda \|X\|_1$$

↳ Solution of this equation can be obtained by taking derivative of above function.

$$\frac{d}{dx}\left(\rho(e) + \lambda \|X\|_1\right)$$

$$= \sum_{k=1}^{d} \frac{d}{dx}\left(\rho(e_k)\right) + \lambda\, \partial \|X\|_1$$

$\partial \|X\|_1 \rightarrow$ derivation of $\|X\|_1$

→ the solution is

$$\frac{1}{2} \sum_{k=1}^{d} \omega(e_k) \frac{d e_k^2}{dx} + \lambda\, \partial \|X\|_1$$

$$\omega(e_k) = \frac{d\rho(e_k)}{de_k} \frac{1}{e_k} = \frac{\exp(-\mu e_k^2 + \mu\delta)}{1 + \exp(-\mu e_k^2 + \mu\delta)}$$

If $\omega(e_k) \rightarrow$ fixed as constant then

$$\frac{1}{2}\sum_{k=1}^{d} \omega(e_k)\, e_k^2 + \lambda \|x\|_1$$

$$= \frac{1}{2}\|W e\|_2^2 + \lambda \|x\|_1$$

$y - [D, A]\, x$

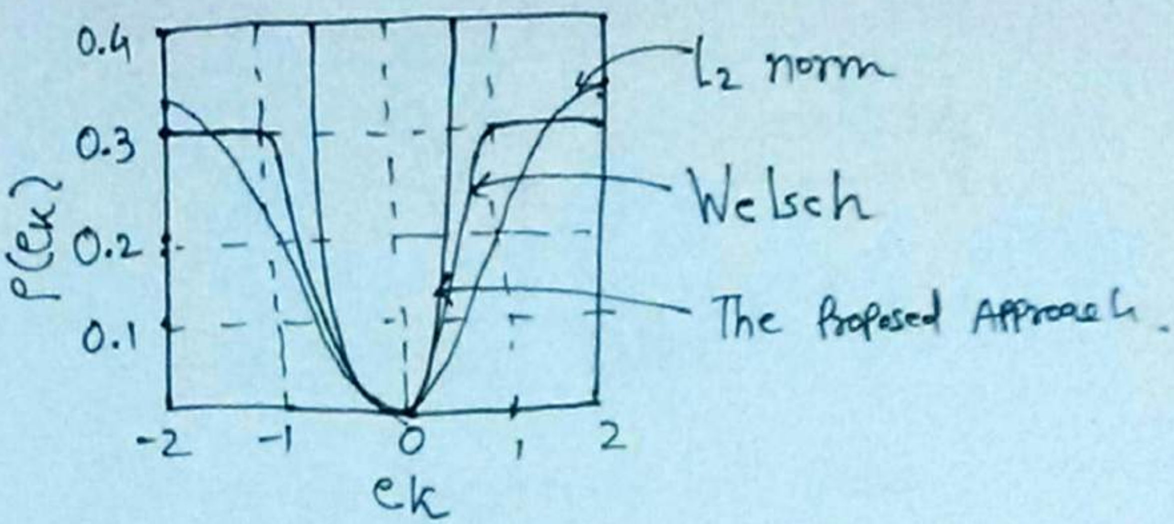
diag $(\omega(e_1), \omega(e_2) \ldots \omega(e_d))^{1/2}$

$$\min_{x} \left\| W\left(y - [D, A]\begin{bmatrix} x_d \\ x_a \end{bmatrix}\right)\right\|_2^2 + \lambda\|x\|_1$$

fixed

$\Rightarrow$ ~~it can be~~ ~~solved~~ by ~~the~~ ~~Homotopy,~~

$\rightarrow$ The above e.q. is solved by Homotopy method.

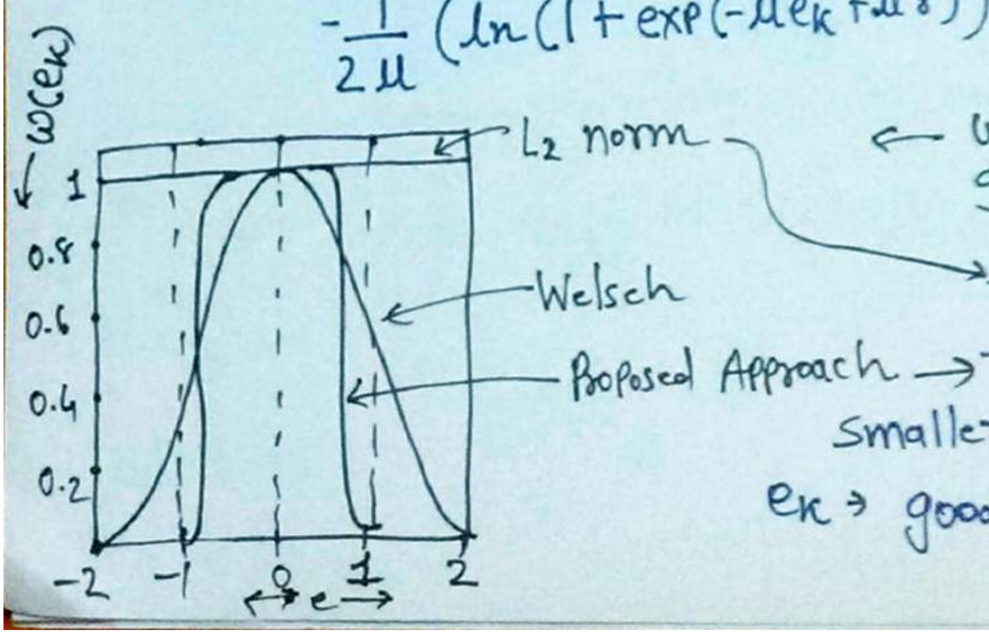


Residual function Graph

L2-norm → Sensitive to outliers.
↳ grows quadratically

The Proposed :- $\rho(e_k) =$

$$-\frac{1}{2\mu}\left(\ln(1+\exp(-\mu e_k^2+\mu\delta)) - \ln(1+\exp \mu\delta)\right)$$



← weight function graph.

→ constant.

→ output smaller for large e_k → good for occlusion

Occlusion

RADL algorithm treat Occlusion as pixels that have large reconstruction errors.

→ weight function of L2- norm is ~~a~~ constant function.

→ In the proposed algorithm weight function outputs smaller value for large |ek|

The residual function → minimize the influence of outliers.

$P \rightarrow$ no. of subjects in external dataset.

Y_e $\quad$ D_e

$\uparrow$ Probe set $\quad$ $\uparrow$ Gallery set.

$$Y_e = [y_e^1, y_e^2, \ldots y_e^N] \in R^{d \times N}$$

N images of diff. intraclass variations to be modeled.

$$D_e \in R^{d \times rP}$$

$\uparrow$ no. of images per subject.

if $r = 1$ then

$$D_e \in R^{d \times P}$$

the minimization problem is

$$\min_{A, X} \sum_{i=1}^{N} \rho\left(y_e^i - [D_e, A]\begin{bmatrix} x_d^i \\ x_a^i \end{bmatrix}\right) + \lambda \|x^i\|_1 + \eta \rho(y_e^i - D_e \delta_{i1}(x_d^i) - A x_a^i)$$

Classification Rule

*→ Assign test image to class with minimum reconstruction error.

$$\underbrace{\text{Sparse coding}}_{\text{Stage-1}} + \underbrace{\text{Dichonary }\cancel{\text{Upad}}\text{ Update}}_{\text{Stage-2}}$$

Stage-1: X

Stage-2: A

Sparse coding

↳ fix → A and optimize ~~with~~ e.g. with respect to X.

$$\min_{x^i} \rho(y_e^i - [D_e, A]x^i) + \lambda \|x^i\|_1 + \eta \rho(y_e^i - D_e \delta_{il}(x_d^i) - A x_d^i]$$

$$\min_{x_i} \|W_g(y_e^i - [D_e, A]x^i)\|_2^2 + \lambda \|x^i\|_1 + \eta \|W_c(y_e^i - D_e \delta_{il}(x_d^i) - A x_d^i\|_2^2$$

$A \in R^{d \times m} \leftarrow$ no. of Dictionary atoms.

↳ Auxiliary Dictionary.

$x^i = [x_d^i ; x_a^i]$

Sparse coefficients of y_e^i

$x_d^i \leftarrow$ coefficients associated with D_e.

$x_a^i \leftarrow$ coefficients associated with A

$X = [x^1, x^2, \dots x^N]$

$\in R^{(m+p) \times N}$

$\delta_{i_l}(x_d^i) \rightarrow$ Vector with non-zero entries in x_d^i that associated with class i_l ($i_l \rightarrow$ label of y_e^i in external Dataset)

λ & $\eta \rightarrow$ Control the weights of sparsity & class wise reconstruction error.

1st term → data representation

2nd term → Sparsity constraint

last term → Reconstruction error for classi

$W_g = \mathrm{diag}\,(\omega(g_1), \omega(g_2) \dots \omega(g_d))^{1/2}$

$W_c = \mathrm{diag}\,(\omega(c_1), \omega(c_2) \dots \omega(c_d))^{1/2}$

B → Dictionary Update A

fix X and optimize e.g. with respect to A.

$$\min_{\alpha^j} \sum_{i=1}^{N} \rho(y_e^i - [D_e, A] x^i) + \eta \rho(y_e^i - D_e f_{ie}(X_d^i) - A X_d^i)$$

for $j = 1, 2 \dots m$

α^j → jth column of A.

W_g

W_c.

update each column of j of A by αj → one after other.