IEEE 2017 Conference on Computer Vision and Pattern Recognition

EFFICIENT IMAGE SET CLASSIFICATION USING LINEAR REGRESSION BASED IMAGE RECONSTRUCTION

Syed Afaq Ali Shah*[1], Uzair Nadeem*[1], Mohammed Bennamoun[1], Ferdous Sohel[2] and Roberto Togneri[1]

1. The University of Western Australia
2. Murdoch University

Presented By: Qiuhong Ke

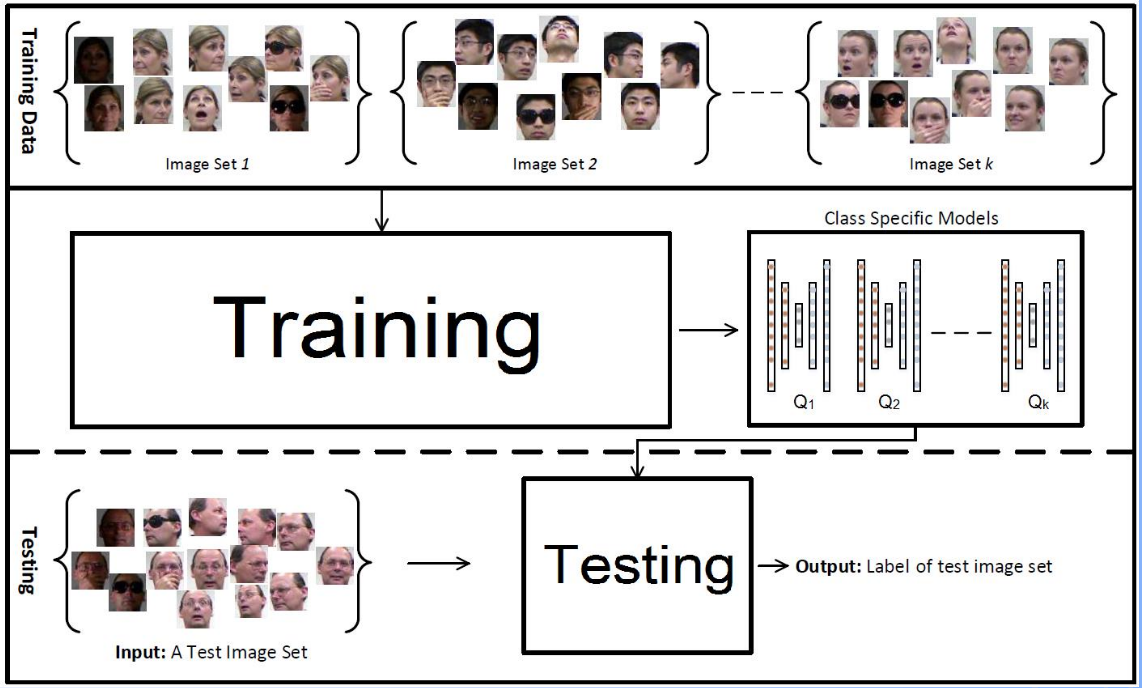
* The first two authors contributed equally to this work

OUTLINE

- Introduction
- Proposed Technique
- Experiments and Analysis
  - Datasets
  - Preprocessing
  - Results
- Conclusion

INTRODUCTION

- **"The problem of recognition from multiple images[1]"**
- Gallery or training set consists of image sets for each class
- Each image-set contains multiple images of same class
- Test set also contains multiple images of same class



General Block Diagram of image set classification

[1] T.-K. KIM, J. KITTLER, AND R. CIPOLLA, "DISCRIMINATIVE LEARNING AND RECOGNITION OF IMAGE SET CLASSES USING CANONICAL CORRELATIONS," IEEE TPAMI, VOL. 29, NO. 6, PP. 1005–1018, 2007.

INTRODUCTION: ADVANTAGES

- Can effectively handle appearance variations:
  - Viewpoint changes
  - Occlusions
  - Non-rigid deformation
  - Variations in illumination
- Applications in biometrics including surveillance, video based face recognition and person re-identification in a network of security cameras[1]

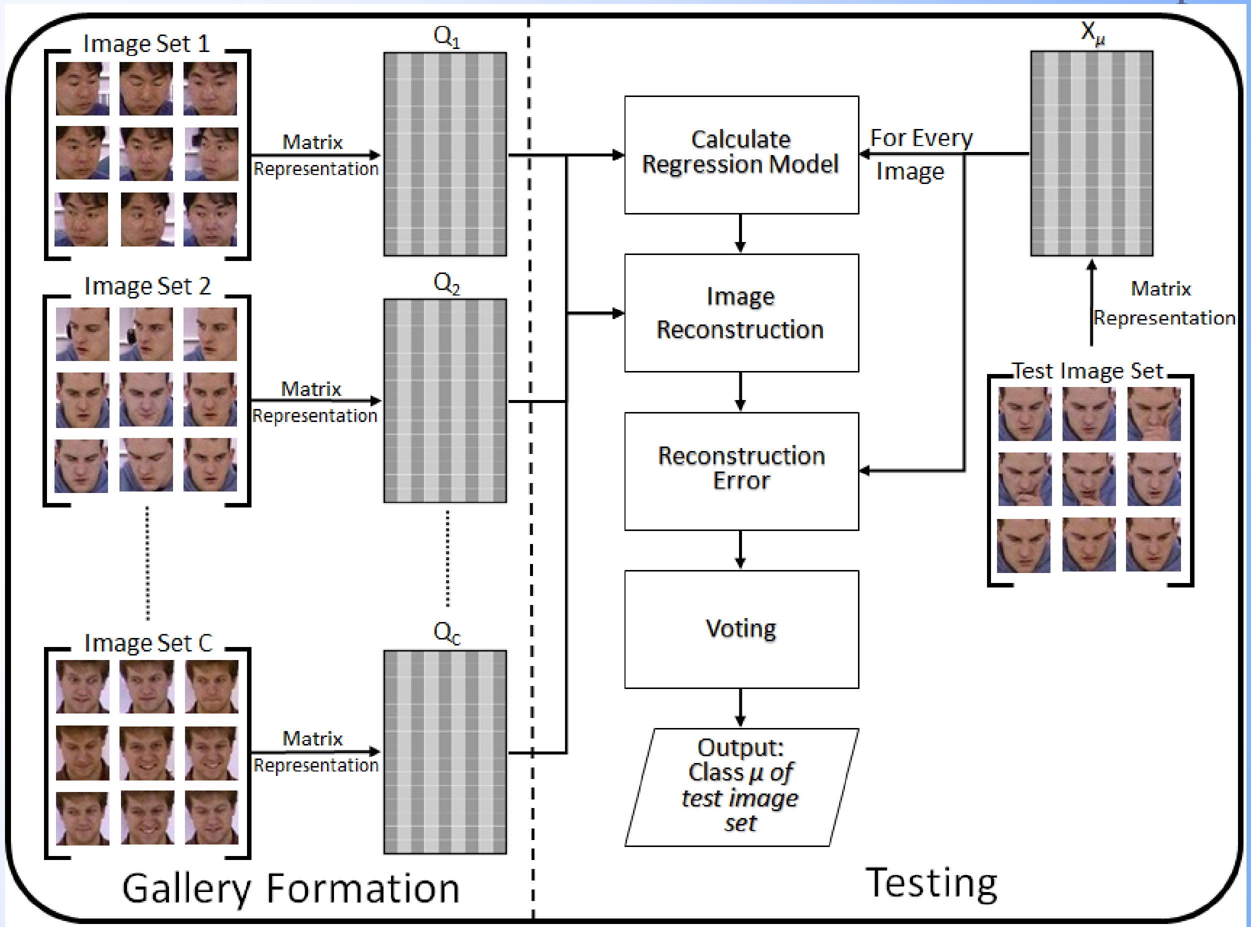
[1] M. HAYAT, M. BENNAMOUN, AND S. AN, "DEEP RECONSTRUCTION MODELS FOR IMAGE SET CLASSIFICATION," IEEE TPAMI, VOL. 37, NO. 4, PP. 713–727, 2015.

INTRODUCTION: CHALLENGES

- High data requirement
- Low resolution
- Parameter tuning
- Hand crafted features
- Computational time
- Inclusion of new classes

PROPOSED TECHNIQUE

- A novel non-parametric approach
- Based on image reconstruction using Linear Regression Classification (LRC)[1]



Block Diagram of the proposed technique

[1] I. NASEEM, R. TOGNERI, AND M. BENNAMOUN, "LINEAR REGRESSION FOR FACE RECOGNITION," IEEE TPAMI, VOL. 32, NO. 11, PP. 2106–2112, 2010.

PROPOSED TECHNIQUE: MATRIX REPRESENTATION

- Gallery Sets or Regressors

$$Q_c = [q_c^1 q_c^2 q_c^3 \ldots q_c^N] \in \mathbb{R}^{T \times N}$$

$$c = 1, 2, 3, \ldots, C$$

- Test Set

$$X_\mu = [x_\mu^1 x_\mu^2 x_\mu^3 \ldots x_\mu^M] \in \mathbb{R}^{T \times M}$$

$\boldsymbol{\mu}$ = Unknown class of test set

N = No. of images in gallery set

C = No. of unique classes

M = No. of images in test set

T = No. of pixels in downsampled images

PROPOSED TECHNIQUE: TWO IMPLEMENTATIONS

Vector Implementation	Matrix Implementation
Estimation of regression model parameters using Least squares based solution	
$x_\mu^m = Q_c\gamma_c^m$ $\gamma_c^m = (Q_c'Q_c)^{-1}Q_c'x_\mu^m$	$X_\mu = Q_c\Gamma_c$ $\Gamma_c = (Q_c'Q_c)^{-1}Q_c'X_\mu$
The regression model is used to reconstruct the test image	
$\widehat{x}_c^m = Q_c\gamma_c^m$	$\hat{X}_c = Q_c\Gamma_c$

PROPOSED TECHNIQUE: DECISION MAKING

- Reconstruction error as distance metric

$$d_c^m = \left\| x_\mu^m - \widehat{x}_c^m \right\|_2, \ c = 1, 2, ..., C, \ m = 1, 2, ..., M$$

- Weighted Voting

$$\theta_c^m = e^{-\alpha d_c^m}, \ c = 1, 2, ..., C, \ m = 1, 2, ..., M$$

$$\Theta_c = \sum_{m=1}^{M} \theta_c^m, \ c = 1, 2, ..., C$$

$$\mu = arg \ \max_c(\Theta_c) \ c = 1, 2, ..., C$$

PROPOSED TECHNIQUE:
FAST LINEAR IMAGE RECONSTRUCTION

- Moore-Penrose pseudoinverse[1] to calculate the inverse matrix of the regressor
- Two Matrix operations for testing

Let $\widetilde{Q}_c$ be the pseudoinverse of Q_c

$$\Gamma_c = \widetilde{Q}_c X_\mu$$
$$\widehat{X}_c = Q_c(\widetilde{Q}_c X_\mu)$$

- Two times faster on ETH-80 dataset
- Gain in computational efficiency is proportional to dataset size

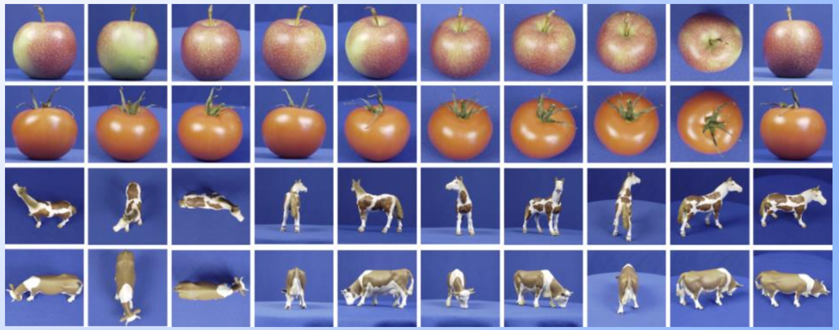
[1] J. STOER AND R. BULIRSCH, INTRODUCTION TO NUMERICAL ANALYSIS. SPRINGER SCIENCE & BUSINESS MEDIA, 2013, VOL. 12.

EXPERIMENTS AND ANALYSIS: DATASETS

- **CMU Motion of Body Dataset (CMU MoBo)**
  - 96 videos of 24 individuals
- **UCSD/ Honda Dataset**
  - 59 videos of 20 individuals
  - Significant head rotations and pose variations
  - Partial occlusions in some frames
- **YouTube Celebrity Dataset (YTC)**
  - 1910 videos of 47 celebrities and politicians
  - Videos are noisy, low resolution and highly compressed
- **ETH-80 Object Dataset**
  - Eight object categories consisting of ten image sets each

Random images of four individuals from CMU/MoBo Dataset

Histogram Equalized, grayscale random images of four celebrities from YTC Dataset.



Random images of four classes from ETH-80 dataset

EXPERIMENTS AND ANALYSIS: PREPROCESSING

- Used considerably less gallery data compared to other techniques
- Viola and Jones face detection algorithm for MoBo and Honda datasets
- Incremental Learning Tracker[1] to track faces in YTC dataset
- Histogram equalized
- No feature extraction: Used downsampled grayscale raw images

[1] D. A. ROSS, J. LIM, R.-S. LIN, AND M.-H. YANG, "INCREMENTAL LEARNING FOR ROBUST VISUAL TRACKING," IJCV, VOL. 77, NO. 1-3, PP. 125–141, 2008.

RESULTS: AVERAGE ACCURACIES and STANDARD DEVIATIONS

Methods↓ / Datasets→	MoBo	YTC	Honda	ETH-80
TIS	96.81 ± 1.97	50.21 ± 3.59	88.21 ± 3.86	75.50 ± 4.83
DCC	88.89 ± 2.45	51.42 ± 4.95	92.56 ± 2.25	91.75 ± 3.74
MMD	92.50 ± 2.87	54.04 ± 3.69	92.05 ± 2.25	77.50 ± 5.00
MDA	80.97 ± 12.28	55.11 ± 4.55	94.36 ± 3.38	77.25 ± 5.46
AHISD	92.92 ± 2.12	61.49 ± 5.63	91.28 ± 1.79	78.75 ± 5.30
CHISD	96.52 ± 1.18	60.42 ± 5.95	93.62 ± 1.63	79.53 ± 5.32
GEDA	84.86 ± 3.24	52.48 ± 4.45	91.28 ± 5.82	79.50 ± 5.24
SANP	97.64 ± 0.94	65.60 ± 5.57	95.13 ± 3.07	77.75 ± 7.31
CDL	90.00 ± 4.38	56.38 ± 5.31	98.97 ± 1.32	77.75 ± 4.16
RNP	96.11 ± 1.43	65.82 ± 5.39	95.90 ± 2.16	81.00 ± 3.16
MSSRC	97.50 ± 0.88	59.36 ± 5.70	97.95 ± 2.65	90.50 ± 3.07
SSDML	95.14 ± 2.20	66.24 ± 5.21	86.41 ± 3.64	81.00 ± 6.58
DLRC	91.60 ± 2.78	65.55 ± 5.16	92.31	86.5 ± 6.03
MMDML	97.8 ± 1.0	—	**100.00 ± 0.0**	94.5 ± 3.5
ADNT	**97.92 ± 0.73**	**71.35 ± 4.83**	**100.00 ± 0.0**	**98.12 ± 1.69**
PLRC	93.74 ± 4.3	61.28 ± 6.37	89.74	87.72 ± 5.67
SFSR	96.0	—	96.8	—
Ours	**98.33 ± 1.27**	**66.45 ± 5.07**	**100.00 ± 0.0**	**94.75 ± 4.32**

RESULTS: EXPERIMENTS AT LOW RESOLUTIONS

Datasets ↓	Methods ↓	20 × 20 Resolution	15 × 15 Resolution
MoBo	ADNT [1]	91.81 ± 2.40	90.56 ± 3.13
	Ours	**98.75 ± 1.38**	**99.31 ± 1.18**
YTC	ADNT [1]	61.06 ± 5.67	57.66 ± 4.85
	Ours	**64.40 ± 5.22**	**65.25 ± 5.05**
Honda	ADNT [1]	100.00 ± 0.00	99.74 ± 0.81
	Ours	100.00 ± 0.00	**100.00 ± 0.00**
ETH-80	ADNT [1]	88.75 ± 6.26	90.25 ± 4.63
	Ours	**95.50 ± 4.04**	**92.75 ± 6.39**

[1] M. HAYAT, M. BENNAMOUN, AND S. AN, "DEEP RECONSTRUCTION MODELS FOR IMAGE SET CLASSIFICATION," IEEE TPAMI, VOL. 37, NO. 4, PP. 713–727, 2015.

RESULTS: COMPUTATIONAL TIME ANALYSIS ON ETH-80 DATASET

Methods ↓	Total Training Time (seconds)	Test Time per Image Set (seconds)
TIS	NR	0.045
DCC	13.36	0.311
MMD	NR	8.43
MDA	1.22	0.005
AHISD	NR	0.095
CHISD	NR	0.213
ADNT	278.8	0.026
GEDA	2.7	0.068
SANP	NR	105.7
CDL	76.21	1.40
RNP	NR	0.027
MSSRC	NR	4.78
SSDML	21.92	0.577
Ours	**NR**	**0.0046**
Ours (Fast)	**NR**	**0.0028**

CONCLUSION

- A novel technique for image set classification
- Competitive results while using considerably less gallery data
- Superior classification accuracies at low resolutions
- No training, feature extraction or parameter tuning
- Achieves fastest computational time.
- Easy to add new classes

THANK YOU

SUPPORTED BY:

THE UNIVERSITY OF WESTERN AUSTRALIA

SEEK WISDOM

Australian Government

Australian Research Council