A Benchmark Study of Large-scale Unconstrained Face Recognition

Shengcai Liao, Zhen Lei, Dong Yi, and Stan Z. Li

Center for Biometrics and Security Research

08/04/2014

Labeled Faces in the Wild (LFW)

❑ **Successful database for unconstrained face recognition research**

- 13,233 face images of 5,749 subjects collected from the Internet
- Widely used by researchers for benchmark evaluation

G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Technical Report 07-49, University of Massachusetts, Amherst, October 2007.

LFW Benchmark Protocols

☐ **10-fold cross-validation**

☐ **Training:**

- Image restricted: use only the defined 300 match/non-match pairs for each fold
- Image unrestricted: all possible match/non-match pairs within each fold can be used
- Unsupervised: use images with no class labels
- Outside data: additional data outside LFW for training

☐ **Test:**

- 300 match/not-match pairs of each fold for classification
- Report mean accuracy and standard deviation

Limitation of LFW Benchmark

□ **Not fully exploit the whole database for evaluation**

  • Only 3,000 matches and 3,000 non-matches

□ **Limited room for algorithm development**

  • Today 97% mean accuracy can be achieved

□ **Not able to evaluate verification rate (VR) at low false accept rate (FAR)**

  • Due to the limited number of non-matches

BLUFR: A New Benchmark Protocol

- **10 random trials designed with the LFW images**
- **Training set for each trial:**
  - 1,500 subjects
  - 3,524 images on average
  - 85,341 genuine matches and 6,122,185 impostor matches
- **Test set for each trial:**
  - 4,249 subjects
  - 9,708 images on average
  - 47,117,778 pairs of matching scores
- **Fused performance report: ($\mu - \sigma$)**
  - Force comparison of the standard deviation
  - Rank algorithms with their "lowest" performances

Benchmark Scenarios and Performance Measures

❑ **Verification**

- 156,915 genuine matches and 46,960,863 impostor matches
- Report VR at FAR=0.1%
- Plot ROC of VR vs. FAR

❑ **Open-set identification**

- Gallery set: 1,000 subjects, one image per subject
- Genuine probe set: 4,350 images of the 1,000 subjects
- Impostor probe set: 4,357 images of the other 3,249 subjects
- Report detection and identification rate (DIR) at rank 1 and FAR=1%
- Plot ROC of DIR at rank 1 vs. FAR

Summary of BLUFR on LFW

❑ **Average statistics of 10 trials**

| Image set			No. Classes	No. Images	No. Genuine matches	No. Impostor matches
Development	Train		100	521	3,925	131,535
	Test		100	673	29,992	196,136
Evaluation	Train		1,500	3,524	85,341	6,122,185
	Test	All	4,294	9,708	156,915	46,960,863
		Gallery	1,000	1,000	-	-
		Genuine probe	1,000	4,350	-	-
		Impostor probe	3,249	4,357	-	-

Baseline Algorithms

☐ **3 kinds of features**

- Hand-crafted feature: LBP
- Learning based descriptor: LE
- Well-aligned high dimensional feature: HighDimLBP

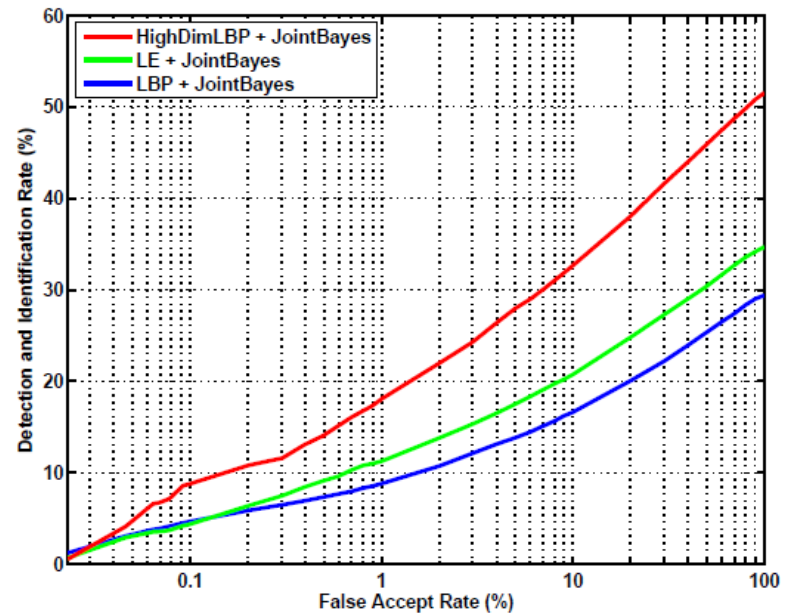☐ **7 kinds of learning algorithms**

- PCA
- LDA
- LMNN
- ITML
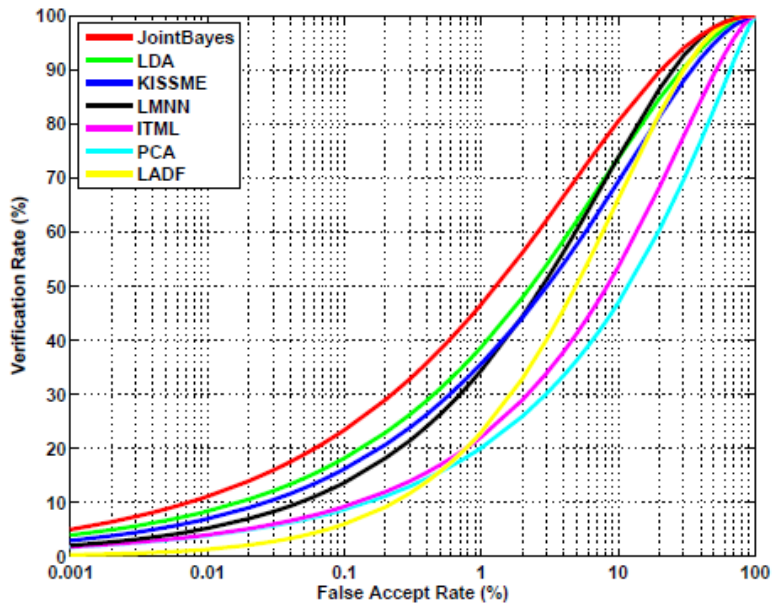- KISSME
- LADF
- JointBayes

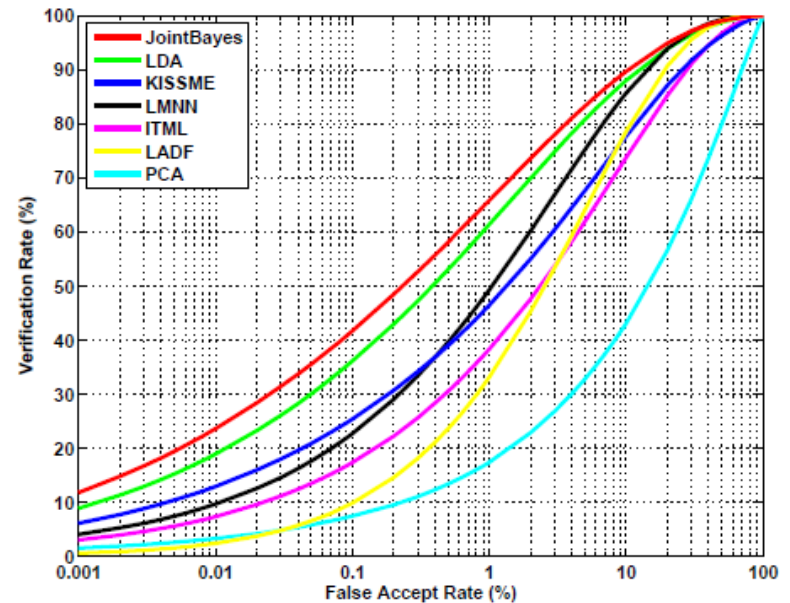# Comparison of Features
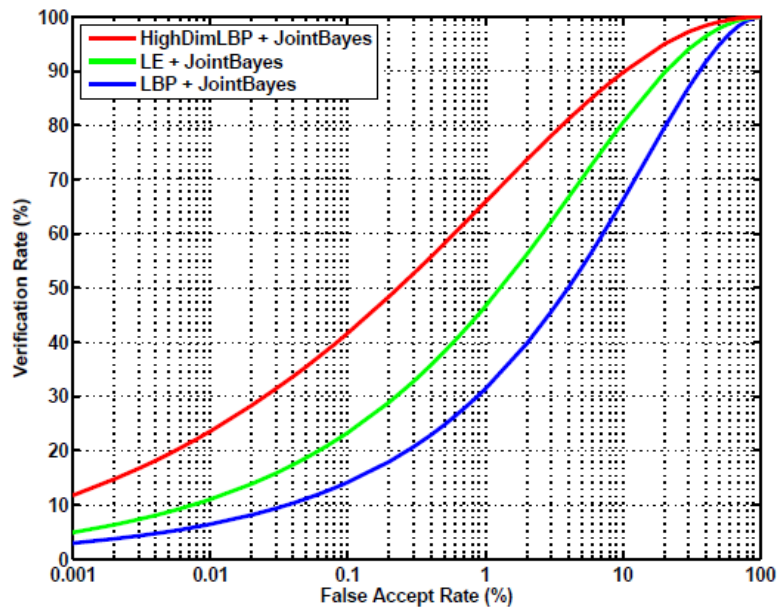
(a) Verification

(b) Open-set identification
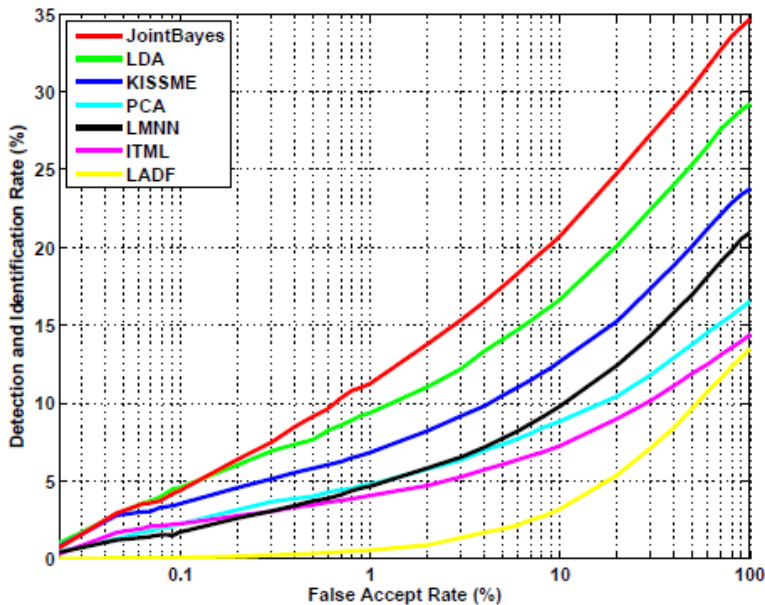
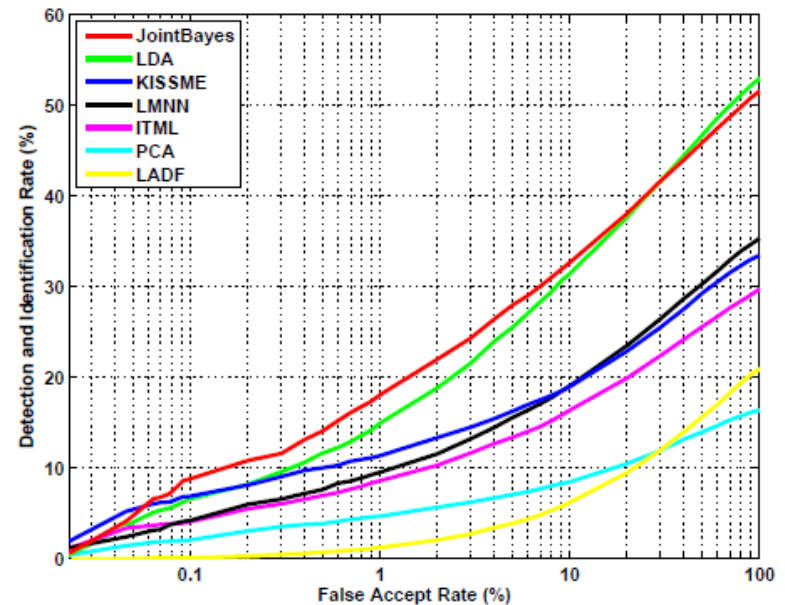# Comparison of Learning Algorithms

□ **Verification**

(b) LE

(c) HighDimLBP

Comparison of Learning Algorithms

□ **Open-set identification**

(b) LE

(c) HighDimLBP

Baseline Results for Verification

Method	FAR=0.1%			FAR=1%		
	LBP	LE	HighDimLBP	LBP	LE	HighDimLBP
JointBayes [7]	**14.18**	**23.31**	**41.66**	**31.39**	**46.60**	**65.84**
LDA [2]	9.80	18.12	36.12	22.56	38.68	61.39
KISSME [14]	11.48	16.12	25.35	27.84	35.59	46.45
LMNN [23]	9.46	13.57	22.68	25.55	34.36	49.29
ITML [9]	9.87	9.16	17.32	23.37	22.06	38.32
LADF [15]	4.77	5.92	9.82	18.32	22.93	33.15
PCA [22]	8.28	8.61	7.41	18.69	20.03	17.38

Baseline Results for Open-set Identification

Method	FAR=1%			FAR=10%		
	LBP	LE	HighDimLBP	LBP	LE	HighDimLBP
JointBayes [7]	**8.82**	**11.26**	**18.07**	**16.61**	**20.73**	**32.63**
LDA [2]	5.18	9.38	14.94	9.45	16.66	31.39
KISSME [14]	5.02	6.83	11.34	9.45	12.69	18.94
LMNN [23]	3.49	4.66	9.53	7.44	9.81	19.04
ITML [9]	4.42	4.07	8.59	8.11	7.25	16.36
PCA [22]	4.81	4.77	4.70	8.97	8.84	8.46
LADF [15]	0.33	0.57	1.24	2.11	3.20	6.20

Conclusions

- **We discussed the limitations of the standard LFW benchmark**

- **A new benchmark protocol, BLUFR, is proposed**

- **Performance for large-scale unconstrained face recognition is still poor:**

  - 41.66% VR at FAR=0.1%
  - 18.07% DIR at rank 1 and FAR=1%

- **A benchmark toolkit is released:**

  - http://www.cbsr.ia.ac.cn/users/scliao/projects/blufr/index.html