Camera Source Identification with Limited Labeled Training Set

Yue Tan[1], Bo Wang[1], Ming Li[1],

Yanqing Guo[1], Xiangwei Kong[1],Yunqing Shi[2]

[1] School of Information and Communication Engineering, Dalian University of Technology

[2] New Jersey Institute of Technology

bowang@dlut.edu.cn

Outline

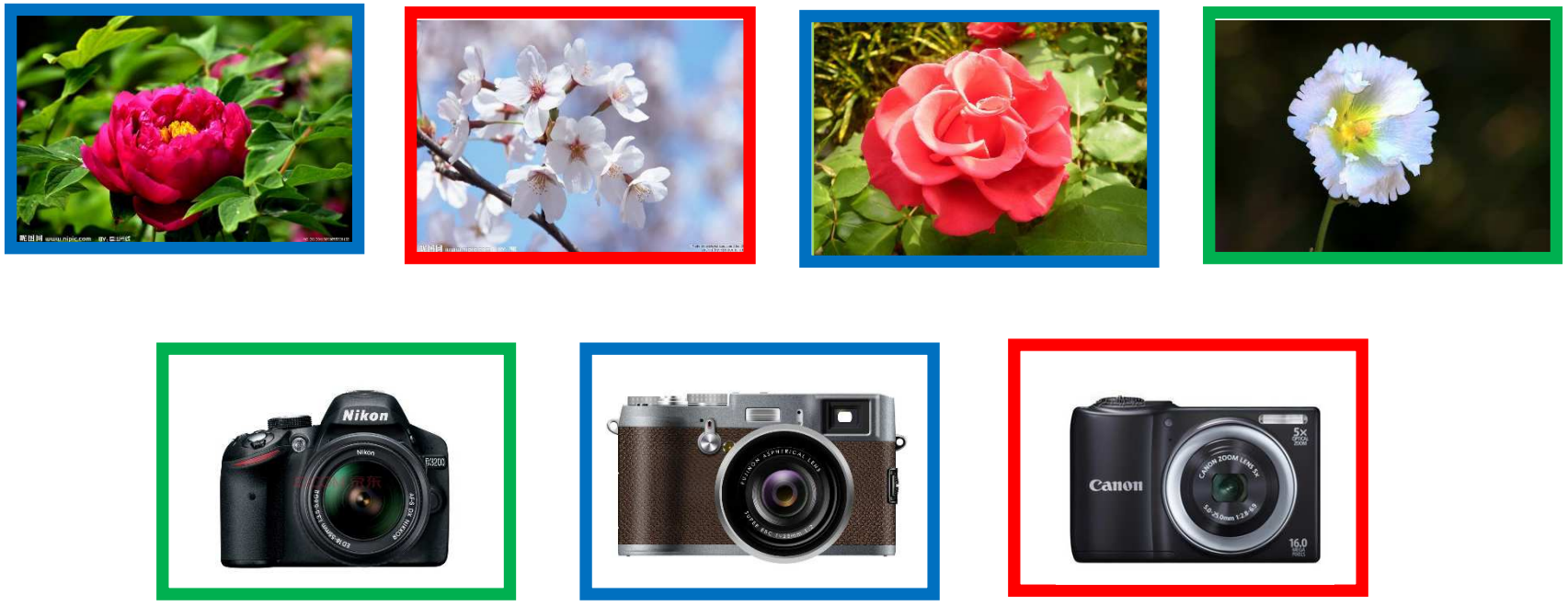
- Introduction
- Motivation of Our Work
- Proposed Method
- Experiments and Results
- Conclusions

Outline

- Introduction
- Motivation of Our Work
- Proposed Method
- Experiments and Results
- Conclusions

Camera Source Identification

- Tracing a unique intrinsic fingerprint && Using statistical characteristics



Outline

- Introduction
- Motivation of Our Work
- Proposed Method
- Experiments and Results
- Conclusions

Motivation of Our Work

Approaches	Accuracy	Samples
Estimation of CFA pattern and interpolation kernel. [8]	90%	200
Color features, image quality measurement (IQM), high order wavelet characteristics (HOWS).[9]	88.02%	150
Uniform gray-scale invariant local binary patterns (LBP)[10]	98%	150~300

Sufficient Labeled Training Samples → High Accuracy

Limited Labeled Training Samples → Accuracy Drop **Below 10%**

Outline

- Introduction
- Motivation of Our Work
- Proposed Method
- Experiments and Results
- Conclusions

The Proposed Method

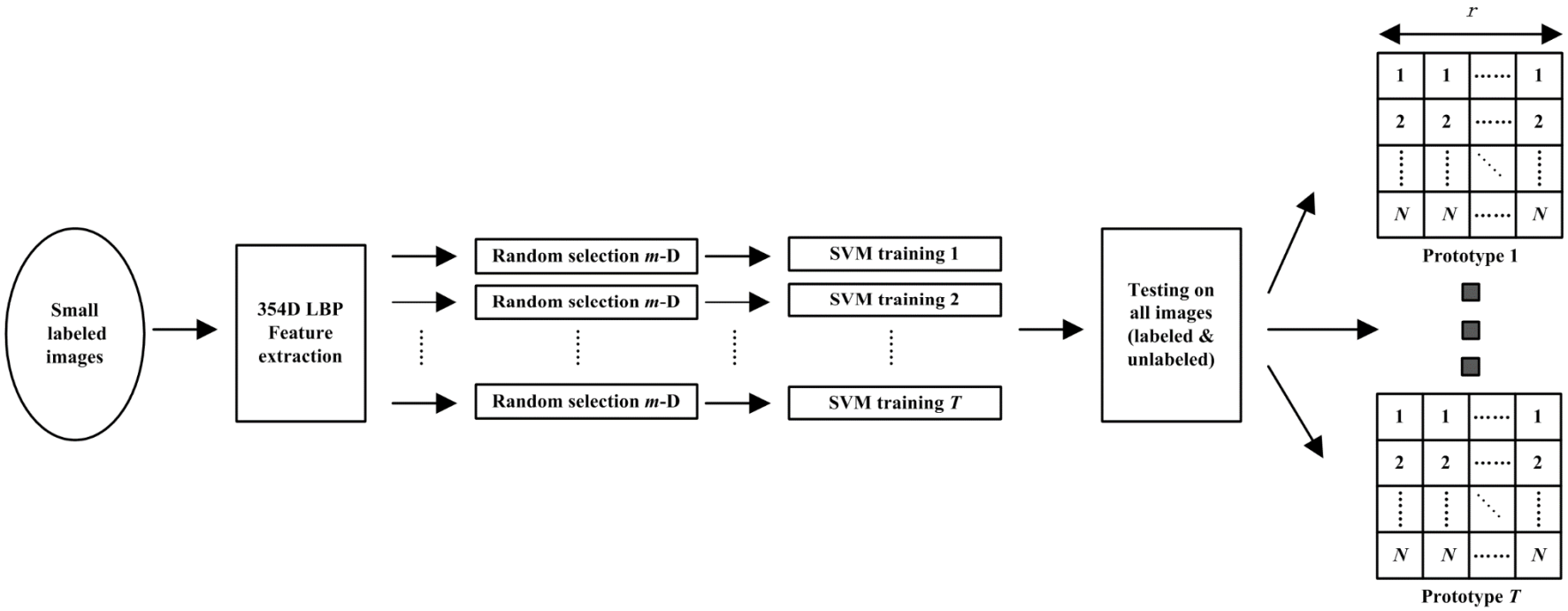
- The Framework of Proposed Method:

➢ Constructing Prototype Set: Using the information of unlabeled samples.

➢ Ensemble Projection: Making combination of the information from all prototypes.

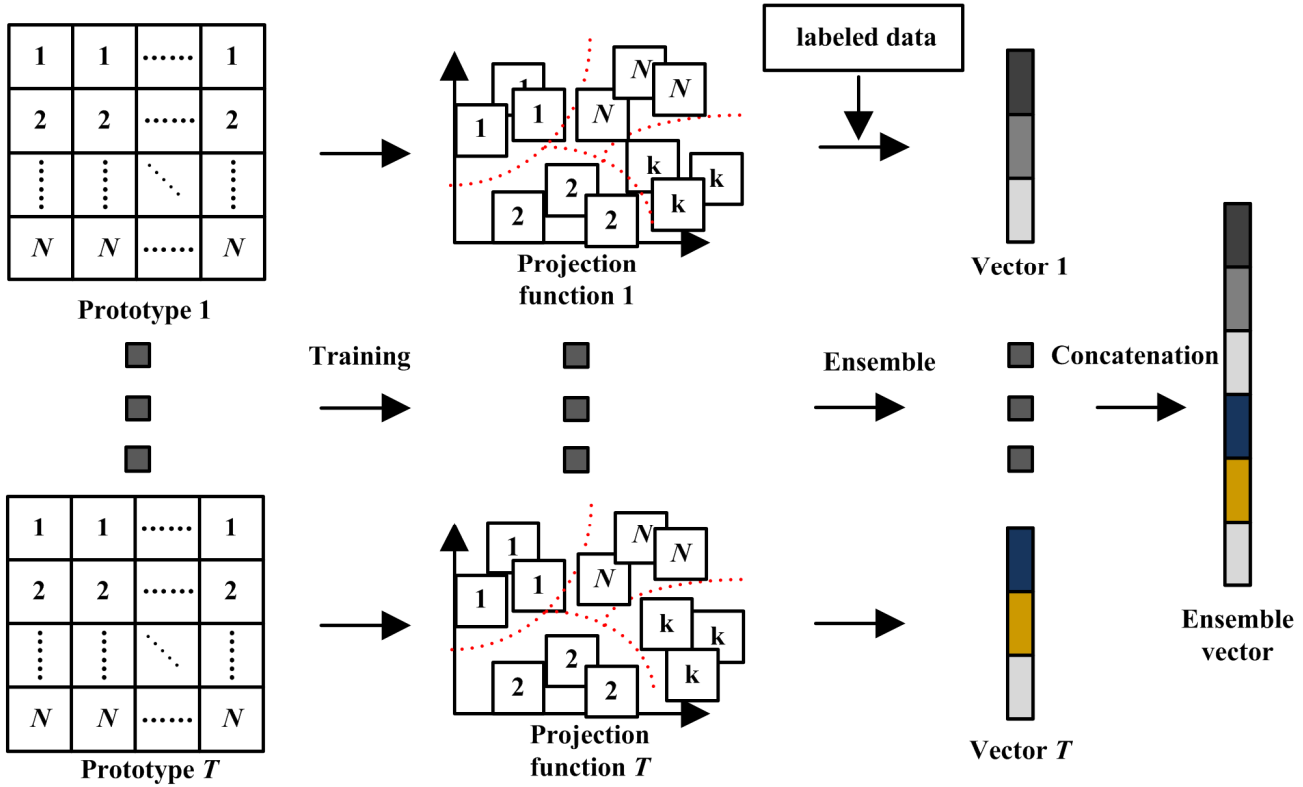
The Proposed Method

- Constructing Prototype Set:



The Proposed Method

- Ensemble Projection:



Outline

- Introduction
- Motivation of Our Work
- Proposed Method
- Experiments and Results
- Conclusions

Dataset & Features

- Experimental Dataset

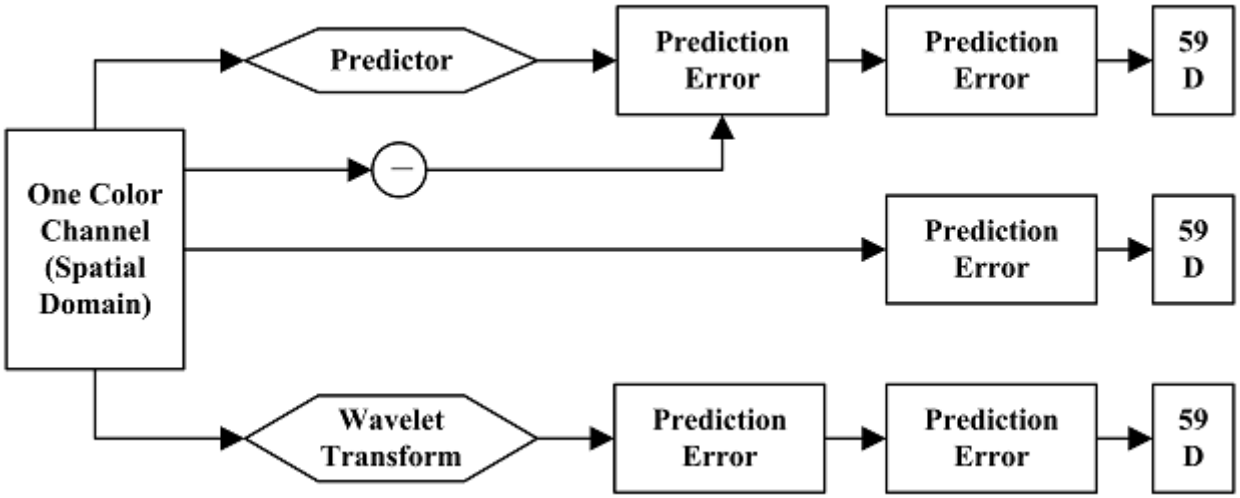
Camera model	Resolution	Abbr.
Casio_EX_Z150	3264 × 2448	CEZ
Kodak_M1063	3664 × 2748	KM1
Nikon_CoolPixS710	4352 × 3264	NCP
Olympus_mju	3648 × 2736	OMJ
Panasonic_DMC	3264 × 2736	PDM
Praktica_DCZ5.9	2560 × 1920	PDC
Nikon_D200	3872 × 2592	ND1
Ricoh_GX100	3648 × 2736	RGX
FujiFilm_FinePixJ50	3264 × 2448	FFP
Pentax_OptioA40	4000 × 3000	POA
Rollei_RCP_7325X	3072 × 2304	RRC
Samsung_L74wide	3072 × 2304	SLW
Samsung_NV15	3648 × 2736	SNV
Sony_DSC_H50	3456 × 2592	SD1
Sony_DSC_T77	3648 × 2736	SD2
Agfa_Sensor530s	2560 × 1920	AFS
Canon_Ixus70	3072 × 2304	CI7
Nikon_D70	3008 × 2000	ND2

Dresden Image Dataset:

- *18 camera models*
- *350 JPEG images for each model*

Dataset & Features

- Features: LBP (Local Binary Pattern)



Camera model identication using local binary patterns. In Proc. IEEE Int. Conf. Multimedia & Expo (ICME), Melbourne, Australia, pp. 392-397(2012).

Experiments and Results

- Average confusion matrix

Average TP=90.2		Predicted																	
		CEZ	KM1	NCP	OMJ	PDM	PDC	ND1	RGX	FFP	POA	RRC	SLM	SNY	SD1	SD2	AFS	CI7	ND2
Actual	CEZ	87.7	*	1.3	*	2.3	*	*	*	*	3.7	*	*	3.0	*	*	*	*	*
	KM1	*	90.0	*	*	1.3	*	*	*	*	1.3	1.0	*	*	4.0	*	*	*	*
	NCP	*	*	92.7	*	*	2.7	*	*	*	1.7	*	*	*	*	*	*	1.0	*
	OMJ	*	*	*	92.0	*	*	*	*	*	*	*	4.0	2.7	*	*	*	*	*
	PDM	*	1.3	1.0	1.0	90.0	2.3	*	*	*	1.0	*	*	*	1.3	*	*	*	*
	PDC	*	*	*	*	1.7	95.3	*	*	*	*	*	*	2.3	*	*	*	*	*
	ND1	*	*	*	*	1.7	2.0	90.0	*	*	*	*	1.0	3.7	*	*	*	1.0	*
	RGX	*	5.0	*	*	*	*	*	85.0	*	*	*	*	1.0	4.3	2.7	*	*	1.0
	FFP	*	*	*	*	*	1.0	*	*	90.0	*	2.3	*	2.0	*	*	*	*	2.7
	POA	*	*	*	1.7	*	*	*	*	*	89.3	*	1.0	*	*	*	1.0	*	4.3
	RRC	*	*	*	*	*	*	2.7	*	*	*	92.3	*	2.0	*	*	*	*	*
	SLM	*	*	1.3	*	*	*	*	*	*	*	1.3	93.7	3.0	*	*	*	*	*
	SNY	*	*	*	*	*	4.0	1.0	*	1.0	*	*	*	89.0	1.0	*	*	*	*
	SD1	*	1.3	*	*	*	*	*	6.7	*	*	*	*	*	88.7	2.3	*	*	*
	SD2	*	*	*	*	*	*	*	2.0	*	*	*	*	*	9.0	88.0	*	*	*
	AFS	*	*	*	*	*	*	*	*	*	1.3	1.0	*	*	*	1.0	91.0	*	5.0
	CI7	1.0	*	*	*	1.0	*	*	*	*	*	*	*	1.3	*	*	*	94.3	*
	ND2	*	*	*	*	*	*	1.0	3.7	*	3.7	*	*	*	1.0	*	5.3	*	83.7

The number of labeled samples is L = 50.
The number of prototype sets is T=200.
The number of the samples of each class in the prototype sets is r = 50.

Experiments and Results

- Average accuracy of camera source identification with different number of labeled image samples *L*.

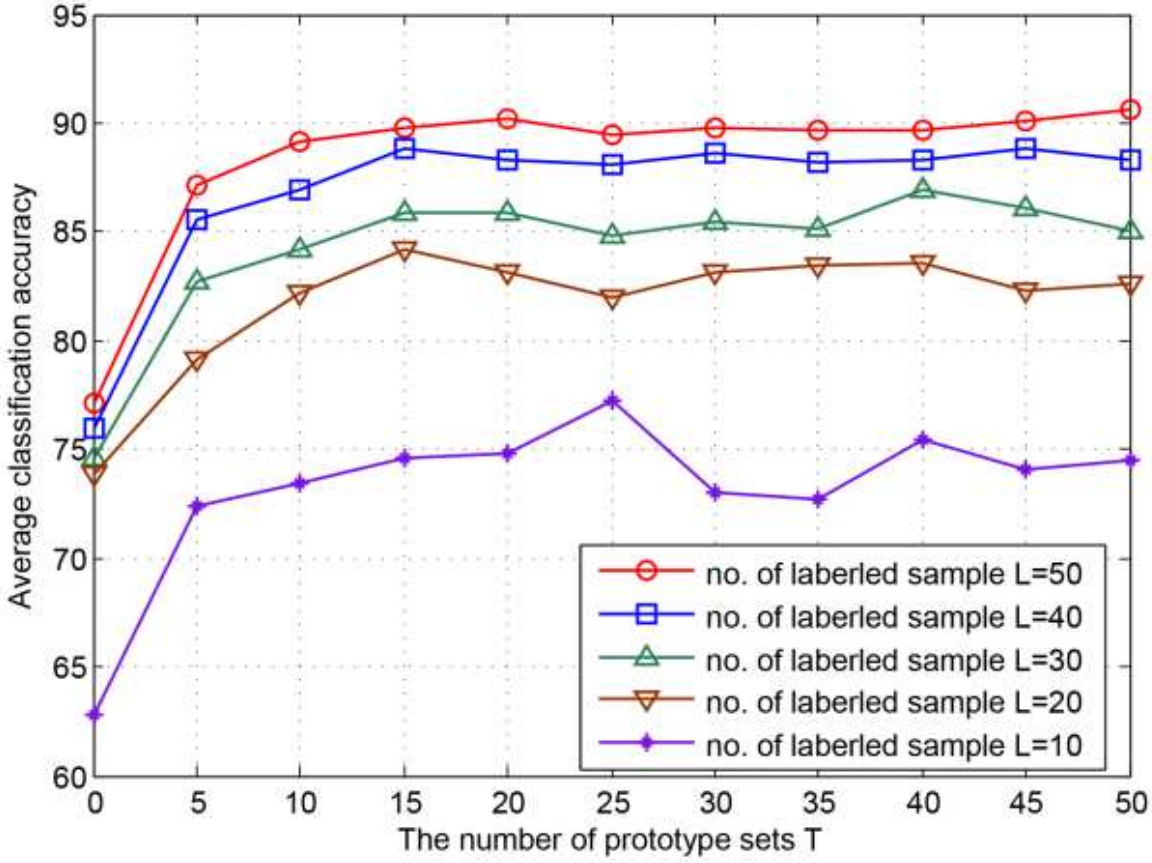
Algorithm	$L = 50$	$L = 40$	$L = 30$	$L = 20$	$L = 10$
LBP	36.0%	26.7%	29.3%	20.9%	8.4%
EP	**90.2%**	**88.3%**	**85.0%**	**82.6%**	**74.5%**

Sufficient Labeled Samples → High Accuracy

Limited Labeled Samples → ~~Accuracy Drop Below 10%~~ Maintain 74.5%

Experiments and Results

- Accuracy rate versus the number of prototype sets T.



T>15, *L*=50:

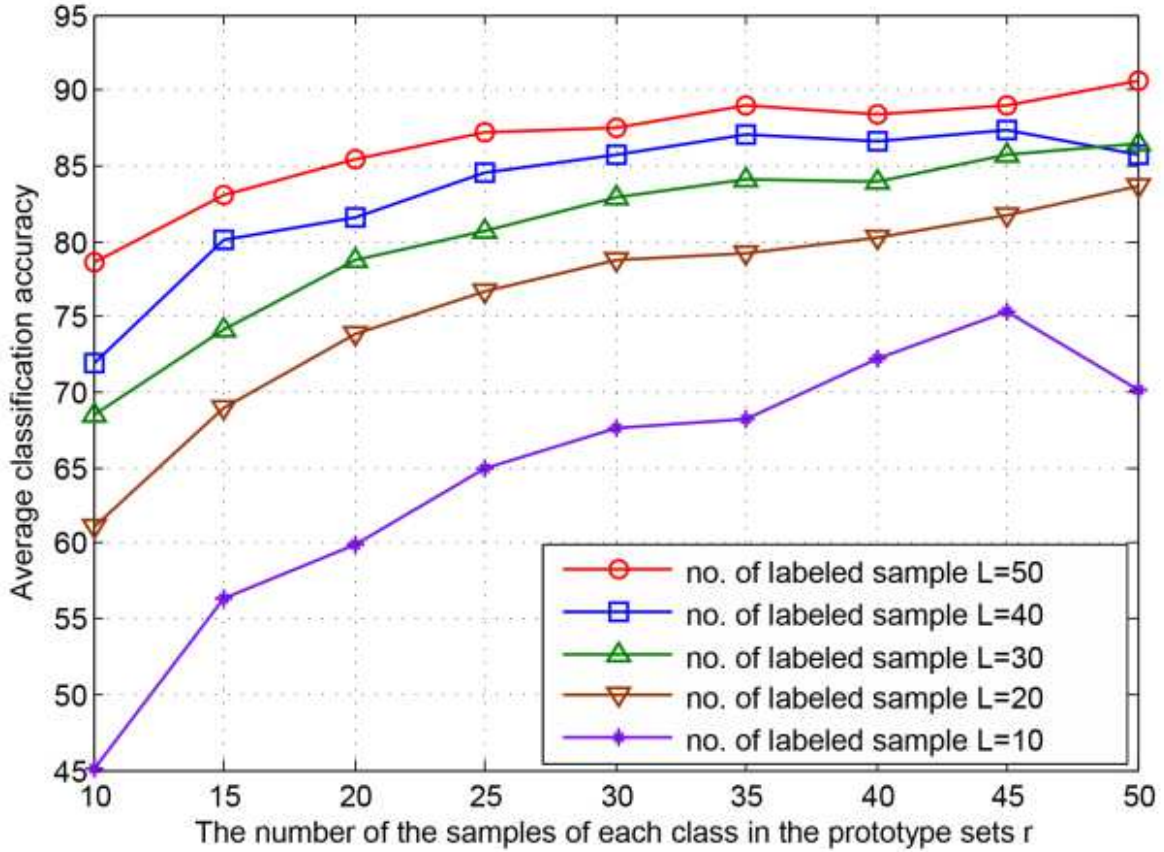
Accuracy maintained at a high level.

L<10:

Result presents a lot of volatility.

Experiments and Results

- Accuracy rate versus the number of the samples of each class in the prototype sets r.



Outline

- Introduction
- Motivation of Our Work
- Proposed Method
- Experiments and Results
- Conclusions

Conclusions

- Main Contribution
  - Proposed EP method achieves a notable higher average accuracy than previous algorithms when labeled training samples is limited.
- Future Work
  - Generalize our work to other features.
  - Generalize our work to improve the identification accuracy rate.

Thank you!

Questions please?

Email: bowang@dlut.edu.cn