14th Int'l Conference on Frontiers of Handwriting Recognition,
Crete island, Greece, 1-4 September 2014

A Tibetan Component Representation Learning Method for Online Handwritten Tibetan Character Recognition

Long-Long Ma, Jian Wu

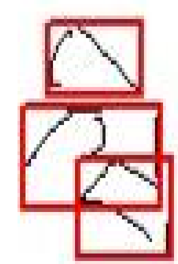
Institute of Software, Chinese Academy of Science

Outline

- Background
- Tibetan Character Structures
- Component-Based Recognition Framework
- Tibetan Component Representation Learning
- Experimental Results
- Conclusion

Background

- Online Handwritten Tibetan Character Recognition
  - Wide applications: tablets, digital pens and mobile devices
  - Statistical approach: high accuracy but large storage, large training sample needed
  - Structural approach: plausible to human perception but high computation
- Component-Based Recognition
  - Take advantage of vertical structures（only one structure)
    - Components: simpler structures, fewer classes
    - Less training samples needed



- Previous Works of Structural Recognition
  - Stroke segmentation and graph matching
    - Stroke-order free but computationally demanding
  - Dynamic programming (DP), HMM
    - Stroke-order dependent

  - component-based sequential matching
    - Stroke-based: stroke segmentation and component segmentation remains difficult
    - HMM-based: model-based component segmentation by level building, dependent on stroke and component order

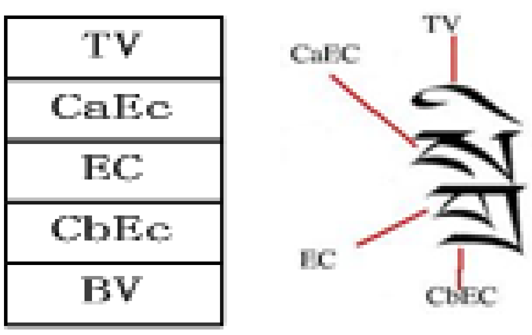
- Our Component-based Structural Approach
  - Statistical vector models for components
  - Over-segmentation of components
  - CRF-based integrated component segmentation-recognition

    - L.-L. Ma, J. Wu, A Component-based On-line Handwritten Tibetan Character Recognition Method Using Conditional Random Field, *Proc of 13th ICFHR* ,Bari, Italy, 2012

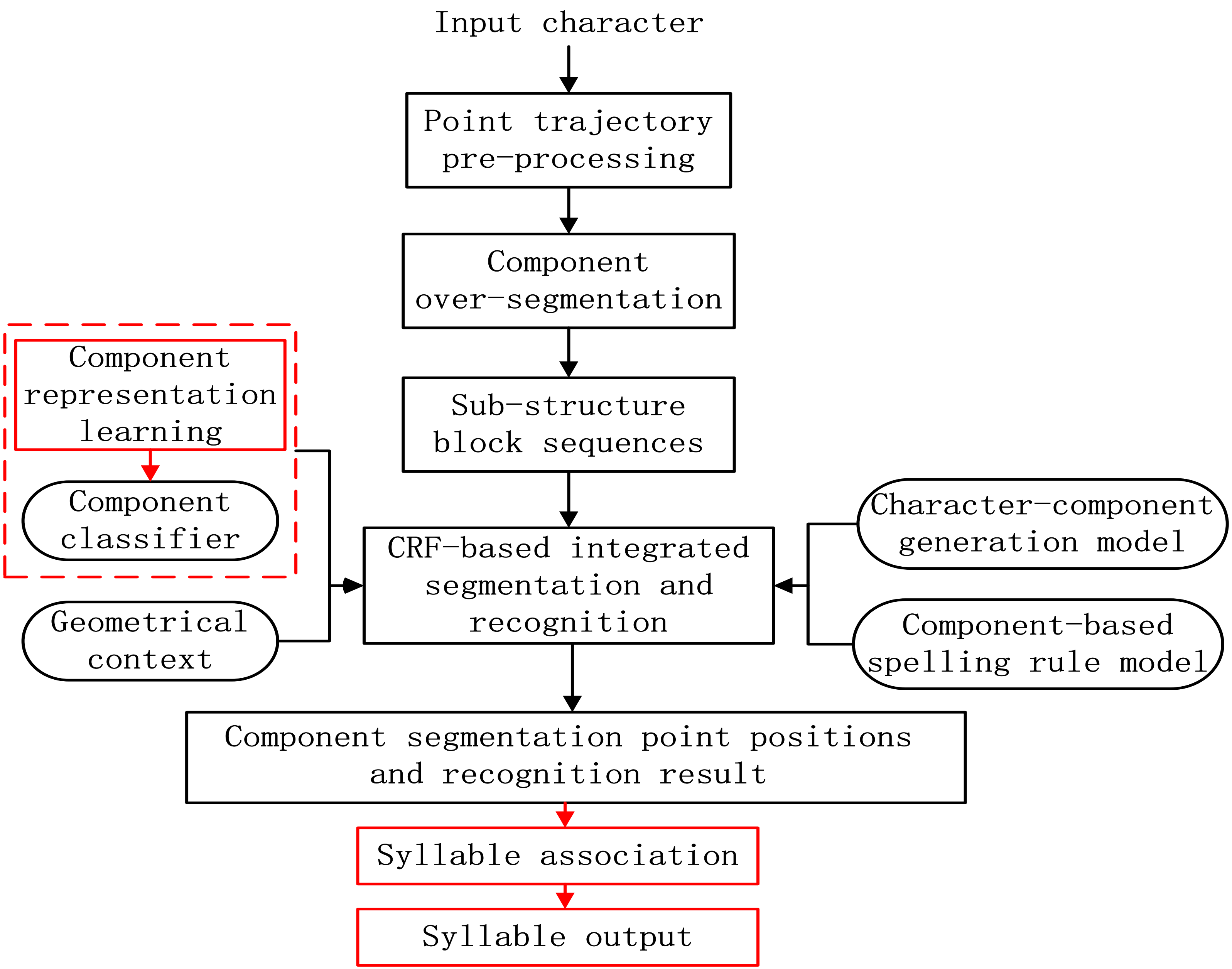
  - Remaining problem: component accuracy affects the whole Tibetan character recognition result
  - Proposed solution: to learn features automatically for Tibetan components

Tibetan Character Structures

- Vertical combination of consonant and vowel
  - Consonant: 30
  - Vowel: 4
- Vertical Structures
  - EC (essential consonant)
  - TV (the top vowel)
  - CaEc (the consonant above the EC)
  - CbEc (the consonant below the EC)
  - BV (the bottom vowel)

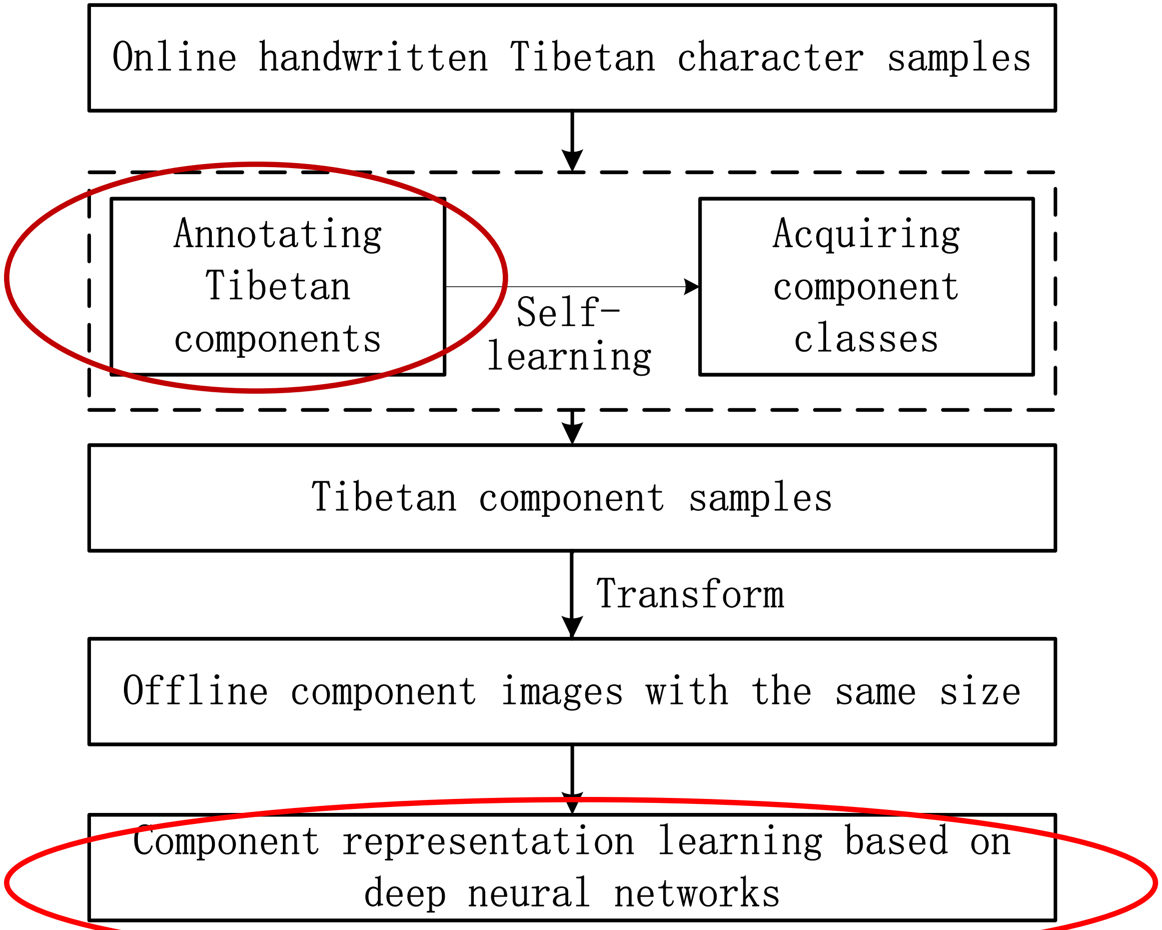


Component-Based Recognition Framework

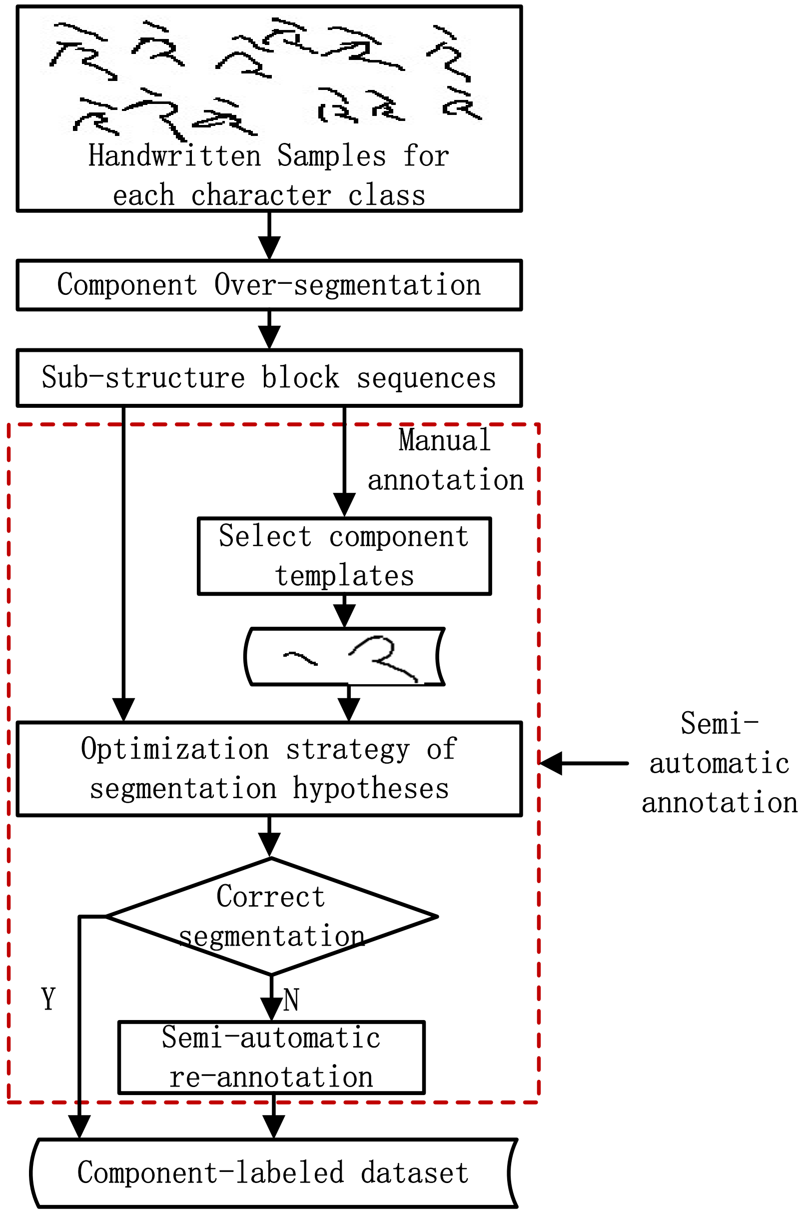


Component Representation Learning

- Rationale
  - Automatic component feature learning using DNN



- Semi-automatic component annotation
  - Optimizing segmentation hypotheses strategy
  - Semi-supervised learning idea

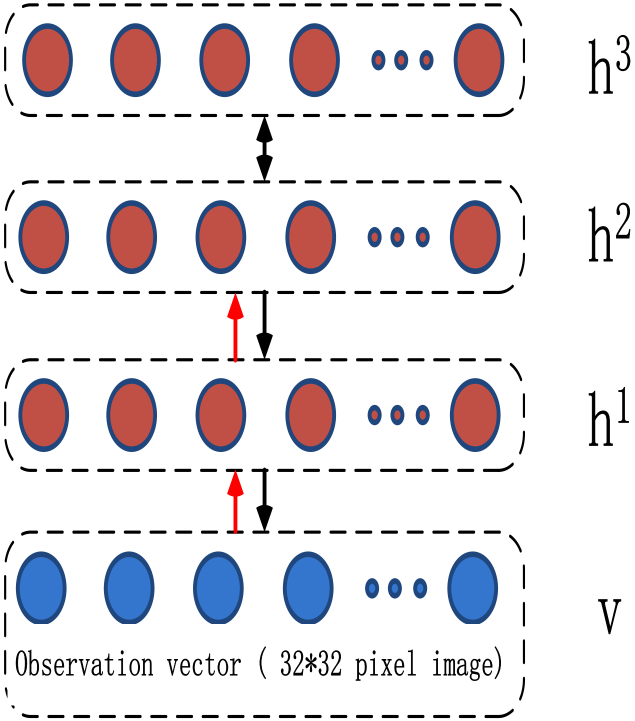


- Component Representation Learning
  - To use DBN to learn component structure
  - joint distribution between observe vector *v* and hidden layers *h*

$$P(v,h)=P(v,h^1,h^2,\cdots,h^l)=(\prod_{k=0}^{l-2}P(h^k\mid h^{k+1}))P(h^{l-1},h^l)$$

  - $P(h^{k-1}\mid h^k)$: for the visible units conditioned on the hidden units of RBM at level
  - $P(h^{l-1},h^l)$: the joint distribution in the top level RBM
  - Parameter optimization

by contrastive divergence(CD) algorithm



Whole Tibetan Character Recognition

- CRF-based integration segmentation and recognition
  - Objective
    - to find the optimal segmentation S from C

$$S^* = \arg\max_{S} P(S \mid C)$$

Sub-structure block sequence: $C = (c_1, c_2, \cdots, c_n)$
Candidate segmentation sequence: $S = (s_1, s_2, \cdots, s_m)$

  - Strategy
    - use CRF integrate multiple models to label the component segmentation points

$$P(S \mid C, \lambda) = \frac{1}{Z(C)} \exp\left(\sum_j \lambda_j F_j(S, C)\right)$$

Experiments

- MRG-OHTC Database
  - 562 character classes
  - 150 samples per class, 120 for training, 30 for testing
  - Annotated component dataset

Data	#Class	#Samples	#Training	#Test
Tibetan character	562	84,300	67,440	16,860
Tibetan component	120	173,250	138,600	34,650

  - Statistics of different component numbers

#component	#character	Percent (%)
1	110	19.57
2	311	55.34
3	141	25.09

Experiments

- Feature Extraction for Component
  - Hand-crafted features
    - Local stroke direction histogram on moment based trajectory normalization
    - 8-direction, 512D
  - Automatic learning features
- Classifier
  - Character/Component classification: MQDF, dimensionality reduction to 160D by FLDA

Experimental results

- Component recognition accuracy for different feature methods
  - DBN+pixel image 1024-600-400-160 network
  - DBN+ hand-crafted feature 512-600-400-160 network
  - Hand-crafted feature+LDA 512->160

Feature method	Component recognition accuracy (%)
DBN+ pixel image	**94.78**
DBN+ hand-crafted feature	89.05
Hand-crafted feature+LDA	91.62

- Whole Character Recognition

Method		#Class	Accuracy (%)
Component-based	CRF+DBN+pixel image	120	**94.09**
	CRF+hand-crafted feature		92.67
	Normalized path evaluation +DBN+pixel image		90.81
	Normalized path evaluation +hand-crafted feature		90.13
Holistic character		562	89.12

Conclusion

- Proposed Work
  - Representation learning method for obtaining automatically Tibetan component features
  - Whole character recognition by integrating many models
- Future Works
  - Aim: improve whole-character recognition accuracy
  - Discriminative learning of component models
  - Extension to Tibetan syllable recognition

Thank you