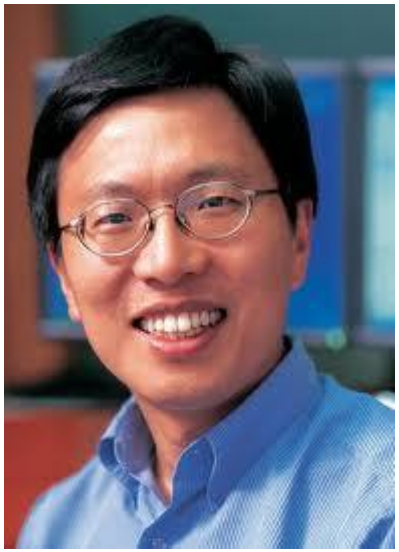
OCR and HWR Applications

杜俊
jundu@ustc.edu.cn

Development Driven Research (DDR)



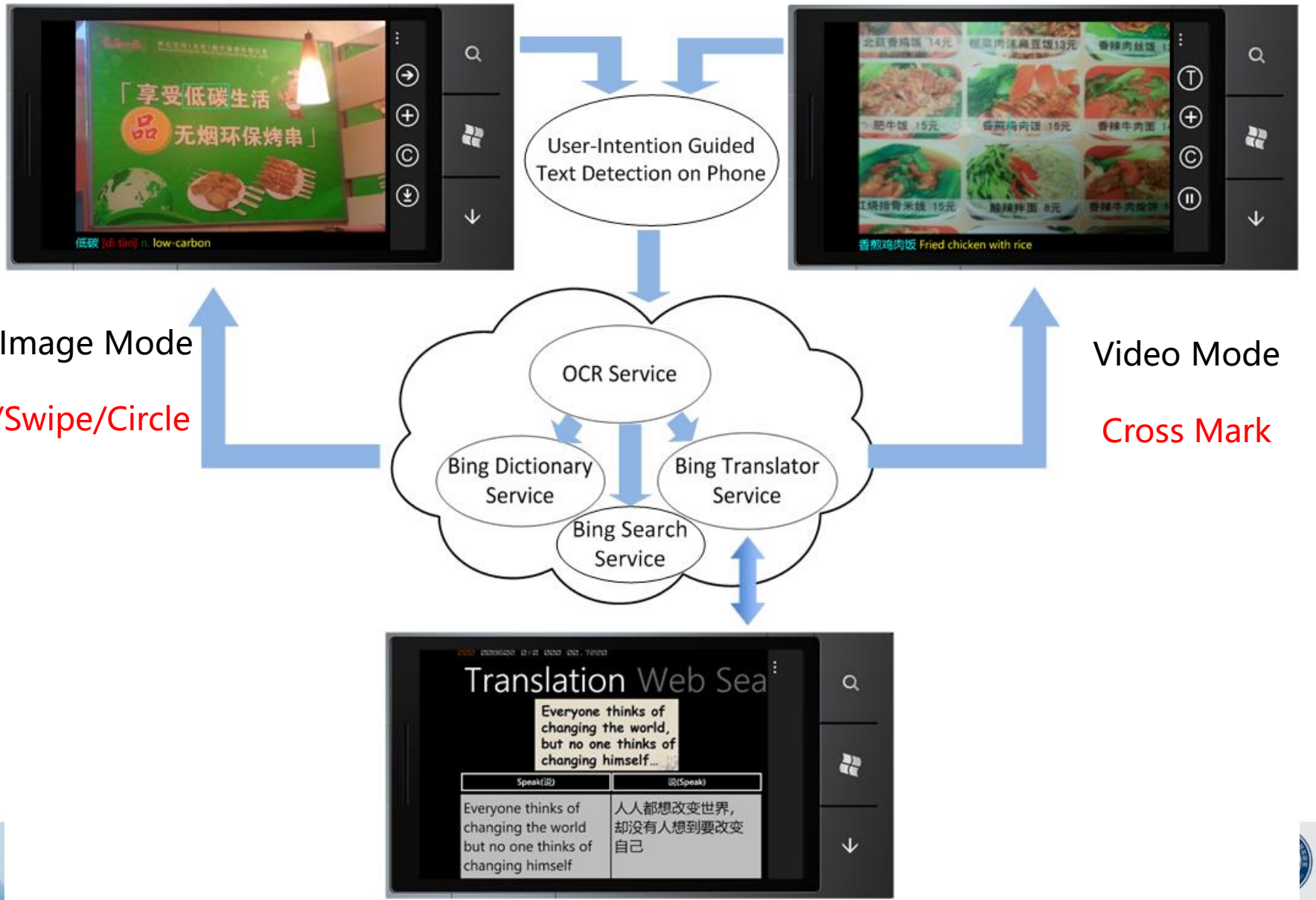
Project 1:
OCR Translation and Search Using Windows Phone

Jun Du, Qiang Huo, Lei Sun, Jian Sun,
"Snap and Translate Using Windows Phone," ICDAR 2011


开封公交进站口
Kaifeng Bus Stop Jin mouth
文明乘车 请自觉排队进站
CivilizatinRidership Please consciously queue
mop.com

Our Demo

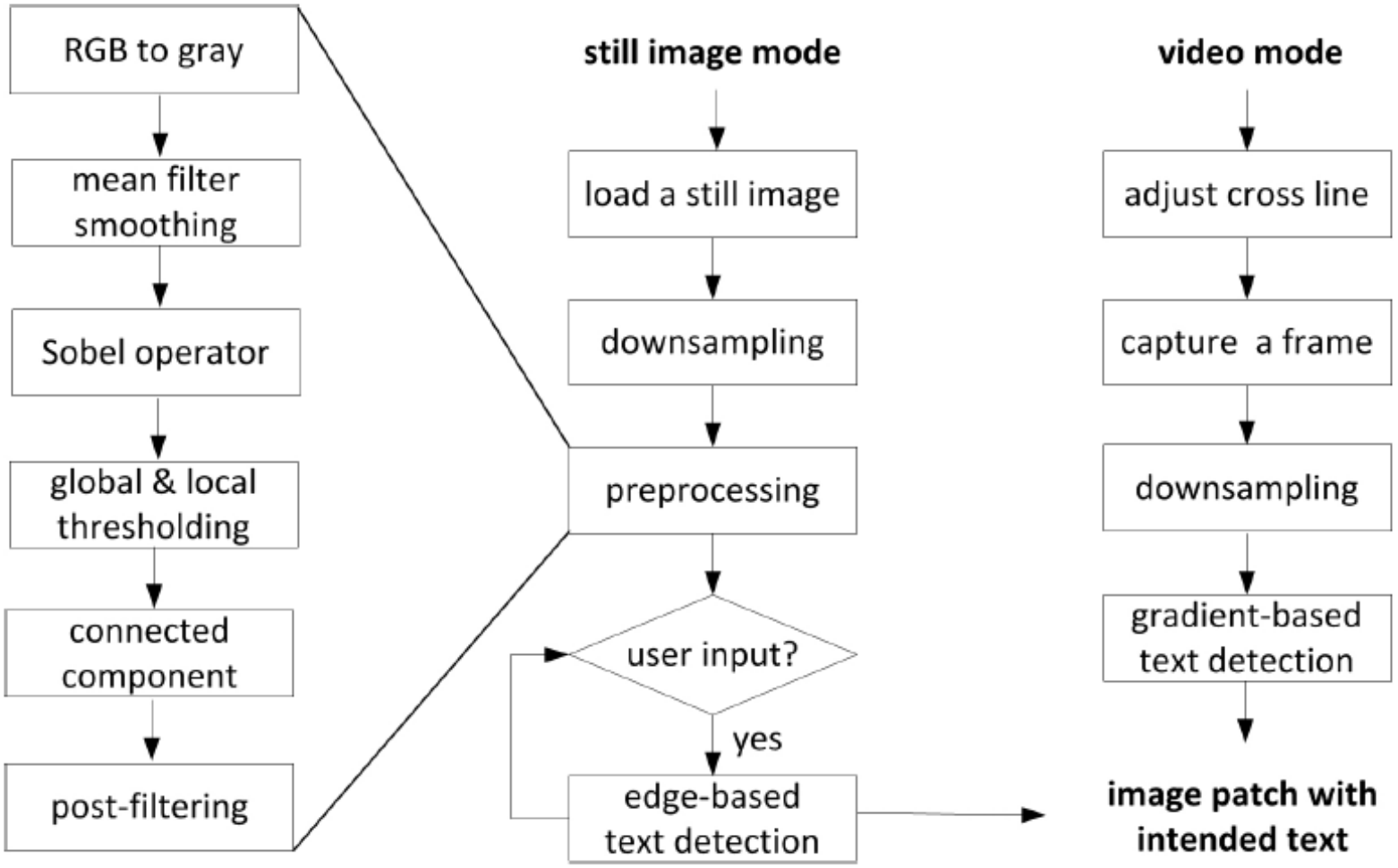
System Architecture



User-Intention Guided Text Detection (I)

- Fast
- Accurate
- Interactive
- Controllable

User-Intention Guided Text Detection (II)



Related Tech-Transfer

- TechFest 2011
  - Most of people want to get this app
- Transfer to OCR Product Team
  - Impact on the UI design of Microsoft Translator app
- Transfer to Bing Dictionary Team (Shipped)
  - OCR-based mouse hovering translation feature
- Transfer to China Incubation Group
  - OCR-based search feature in MyHealth365 Project

My First Ship-it Award

Word Lens

- iPhone App
- Augmented Reality
- Limited Application Scenarios
  - Simple background
  - Word look-up

Microsoft Translator WP App

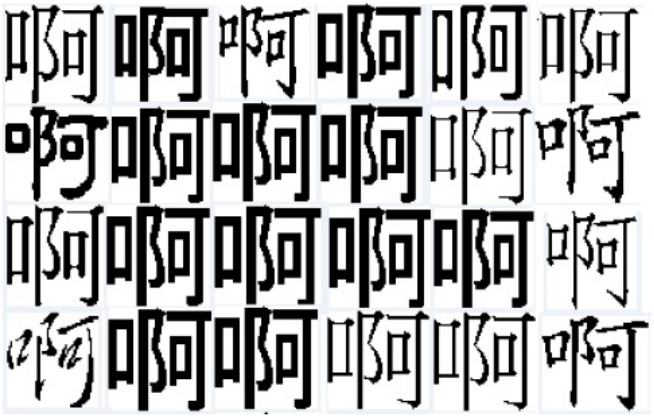
- Three Modes
  - Text
  - Image (OCR: local engine)
  - Speech (ASR: cloud service)

- Translation
  - Local dictionary for most common phases
  - Cloud service

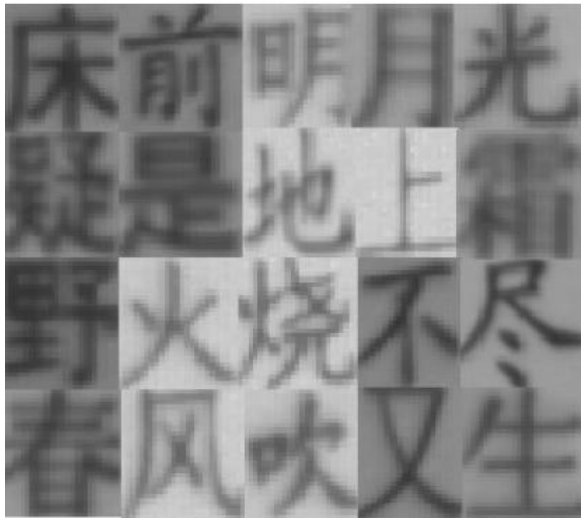
Project 2: Discriminative Linear Regression for OCR Adaptation

Jun Du, Qiang Huo,
"A discriminative linear regression approach to adaptation of multi-prototype based classifiers and its applications for Chinese OCR,"
Pattern Recognition, 2013

Font-Style Adaptation



Low-Quality Text Adaptation



Prototype-based Classifier

- Prototypes for each character class
- Discriminant functions

$$g_i(\boldsymbol{x};\boldsymbol{\lambda}_i) = -\min_k \| \boldsymbol{x} - \boldsymbol{m}_{ik} \|^2$$

- Decision

$$r(\boldsymbol{x};\boldsymbol{\Lambda}) = \arg\max_i g_i(\boldsymbol{x};\boldsymbol{\lambda}_i) \ .$$

SSM-MCE Classifier Training

- MCE Objective Function

$$l(\boldsymbol{X};\boldsymbol{\Lambda}) = \frac{1}{R_1}\sum_{r=1}^{R_1}\frac{1}{1+\exp[-\alpha d(\boldsymbol{x}_r;\boldsymbol{\Lambda})+\beta]}$$

- Sample Separation Margin

$$d(\boldsymbol{x}_r;\boldsymbol{\Lambda}) = \frac{-g_p(\boldsymbol{x}_r;\boldsymbol{\lambda}_p)+g_q(\boldsymbol{x}_r;\boldsymbol{\lambda}_q)}{2 \parallel \boldsymbol{m}_{p\hat{k}} - \boldsymbol{m}_{q\bar{k}} \parallel}$$

$$\begin{aligned}
\hat{k} &= \arg\min_k \parallel \boldsymbol{x}_r - \boldsymbol{m}_{pk} \parallel^2 \\
q &= \arg\max_{i\in\mathcal{M}_r} g_i(\boldsymbol{x}_r;\boldsymbol{\lambda}_i) \qquad g_i(\boldsymbol{x};\boldsymbol{\lambda}_i) = -\min_k \parallel \boldsymbol{x} - \boldsymbol{m}_{ik} \parallel^2 \\
\bar{k} &= \arg\min_k \parallel \boldsymbol{x}_r - \boldsymbol{m}_{qk} \parallel^2
\end{aligned}$$

Model-Space Adaptation

- Linear Transformation

$$\hat{\boldsymbol{m}}_{ik} = \mathcal{F}(\boldsymbol{m}_{ik};\boldsymbol{\Theta}) = \boldsymbol{A}_{e_i}\boldsymbol{m}_{ik} + \boldsymbol{b}_{e_i}$$

- MCE Objective Function

$$l(\boldsymbol{Y};\boldsymbol{\Lambda},\boldsymbol{\Theta}) = \frac{1}{R_2}\sum_{r=1}^{R_2}\frac{1}{1+\exp[-\alpha d(\boldsymbol{y}_r;\boldsymbol{\Lambda},\boldsymbol{\Theta})+\beta]}$$

- Sample Separation Margin

$$d(\boldsymbol{y}_r;\boldsymbol{\Lambda},\boldsymbol{\Theta}) = \frac{-g_p(\boldsymbol{y}_r;\hat{\boldsymbol{\lambda}}_p)+g_q(\boldsymbol{y}_r;\hat{\boldsymbol{\lambda}}_q)}{2\parallel \hat{\boldsymbol{m}}_{p\hat{k}} - \hat{\boldsymbol{m}}_{q\bar{k}} \parallel}$$

Feature-Space Adaptation

- Linear Transformation

$$\boldsymbol{x}_r = \mathcal{F}(\boldsymbol{y}_r;\boldsymbol{\Theta}) = \boldsymbol{A}\boldsymbol{y}_r + \boldsymbol{b}$$

- MCE Objective Function

$$l(\boldsymbol{Y};\boldsymbol{\Lambda},\boldsymbol{\Theta}) = \frac{1}{R_2}\sum_{r=1}^{R_2}\frac{1}{1+\exp[-\alpha d(\boldsymbol{y}_r;\boldsymbol{\Lambda},\boldsymbol{\Theta})+\beta]}$$

- Sample Separation Margin

$$d(\boldsymbol{y}_r;\boldsymbol{\Lambda},\boldsymbol{\Theta}) = \frac{-g_p(\boldsymbol{x}_r;\boldsymbol{\lambda}_p)+g_q(\boldsymbol{x}_r;\boldsymbol{\lambda}_q)}{2\parallel \boldsymbol{m}_{p\hat{k}}-\boldsymbol{m}_{q\bar{k}}\parallel}$$

Rprop Optimization

- Update the step size adaptively

$$S = \frac{\partial l(\mathcal{X}, \mathcal{L}; \mathbf{\Lambda}^{(t-1)})}{\partial m_{ikd}} \cdot \frac{\partial l(\mathcal{X}, \mathcal{L}; \mathbf{\Lambda}^{(t)})}{\partial m_{ikd}}$$

$$\Delta_{ikd}^{(t)} = \begin{cases} \min\left(\eta^{+}\Delta_{ikd}^{(t-1)}, \Delta_{\max}\right) & \text{if } S > 0 \\ \max\left(\eta^{-}\Delta_{ikd}^{(t-1)}, \Delta_{\min}\right) & \text{if } S < 0 \\ \Delta_{ikd}^{(t-1)} & \text{else} \end{cases}$$

$$0 < \eta^{-} < 1 < \eta^{+}$$

Font-Style Adaptation

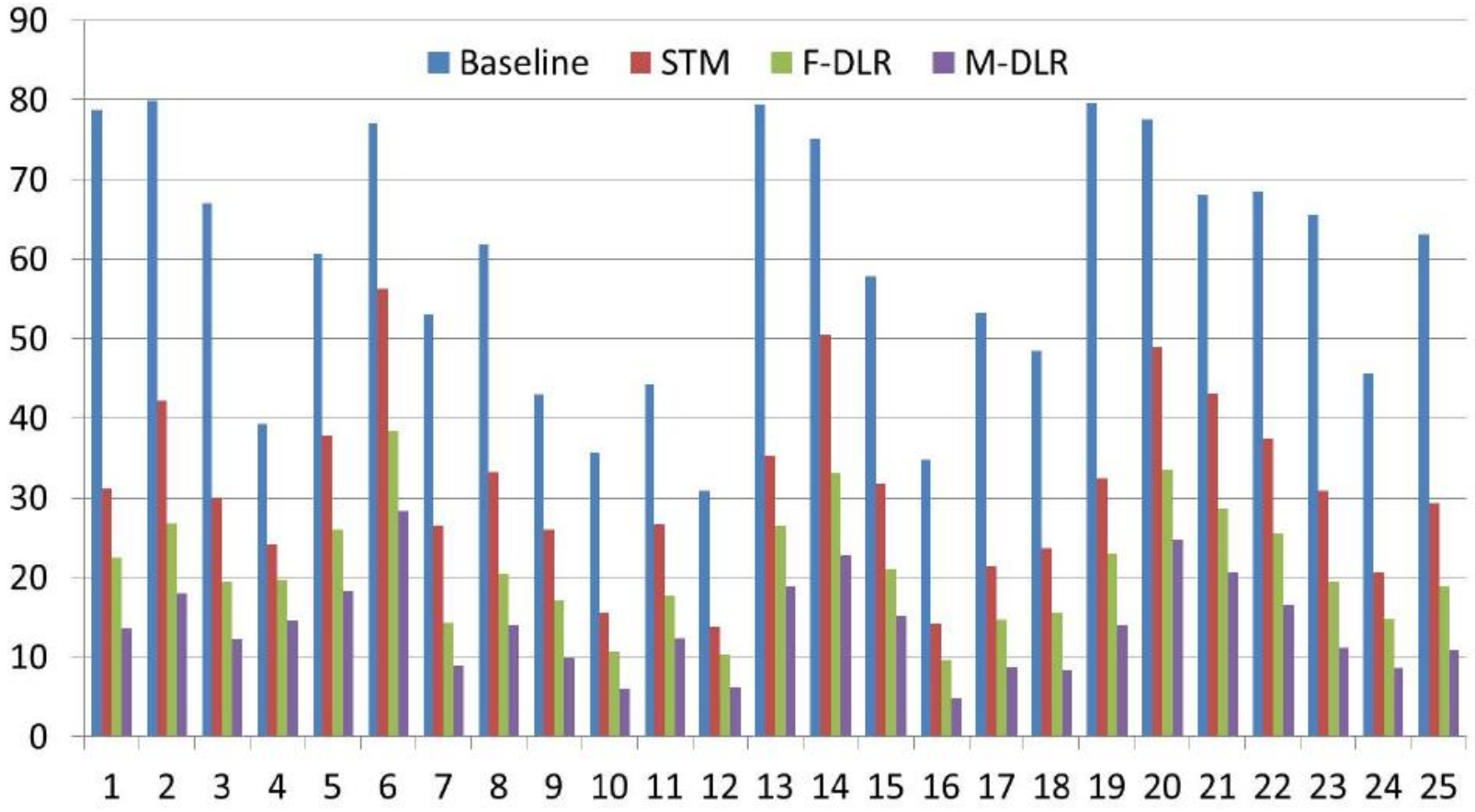


Fig. 6. Performance (character recognition error rate in % on each testing set) comparison of the baseline classifier and different approaches for supervised adaptation to *each of 25 new font styles*.

Low-Quality Text Adaptation

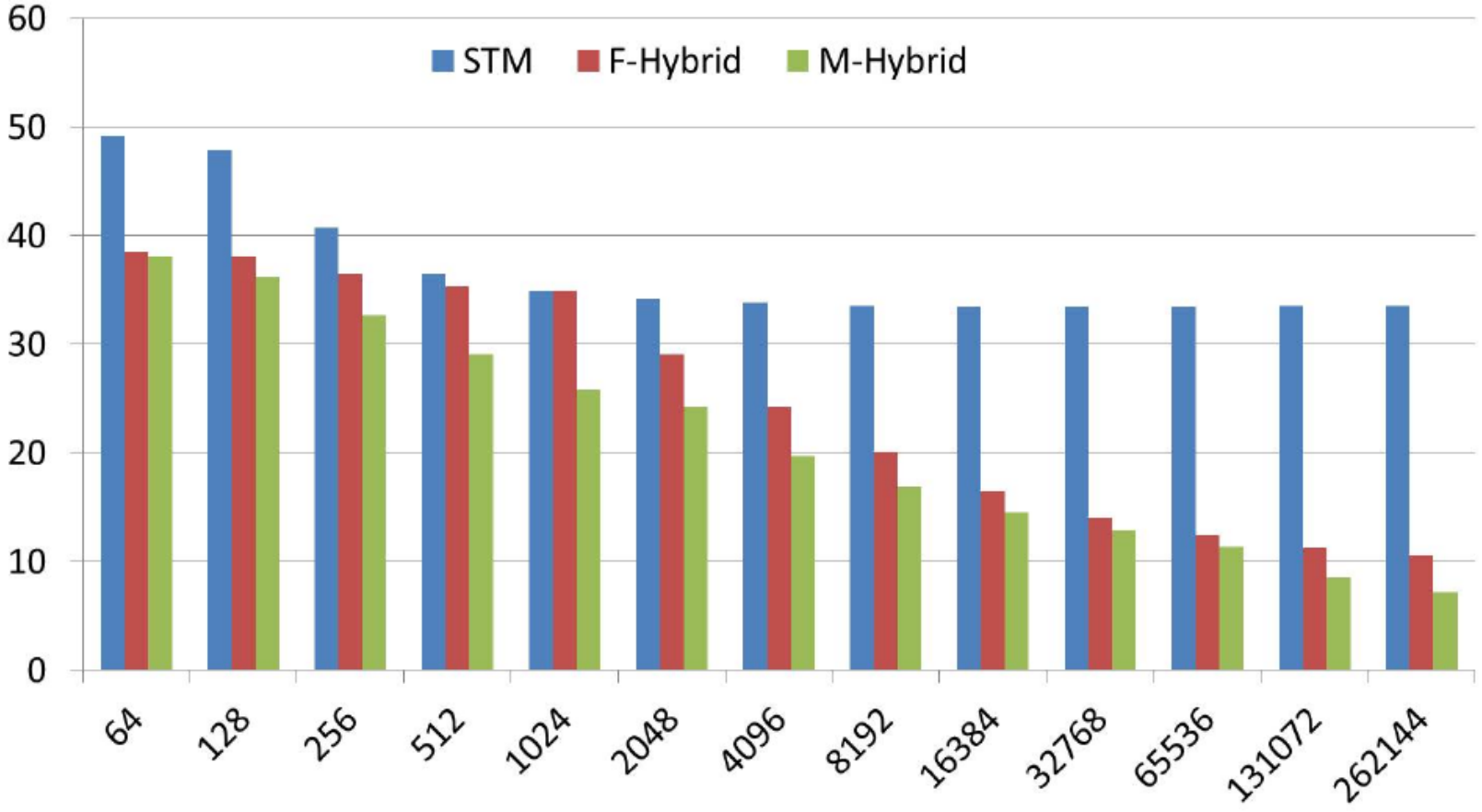
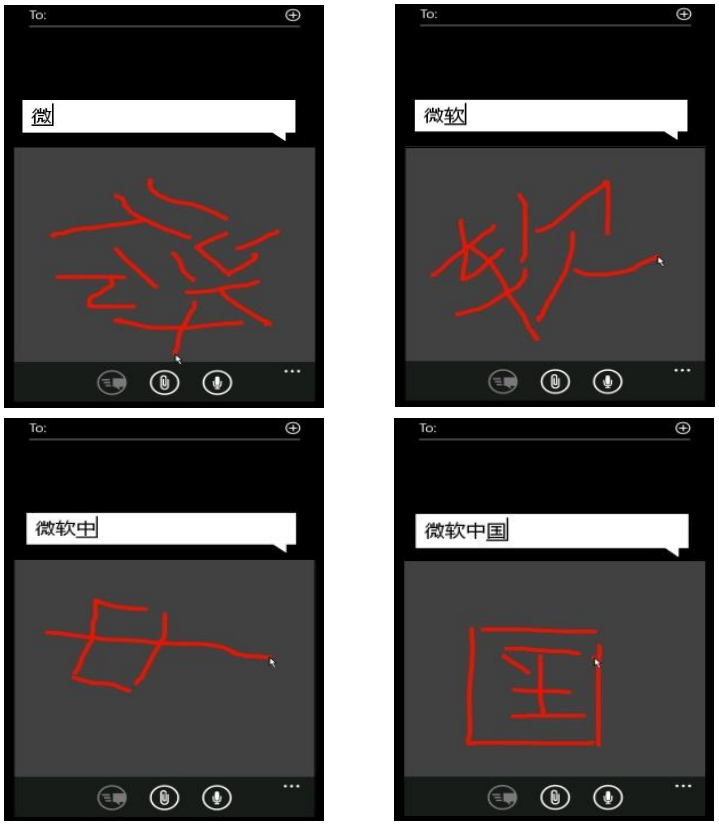


Fig. 9. Performance (character recognition error rate in % on testing set) comparison of different approaches for supervised adaptation with different number of adaptation samples of low-quality text (Baseline recognition error rate is 46.98%).

Project 3:
Snake: a WP8 HWR engine

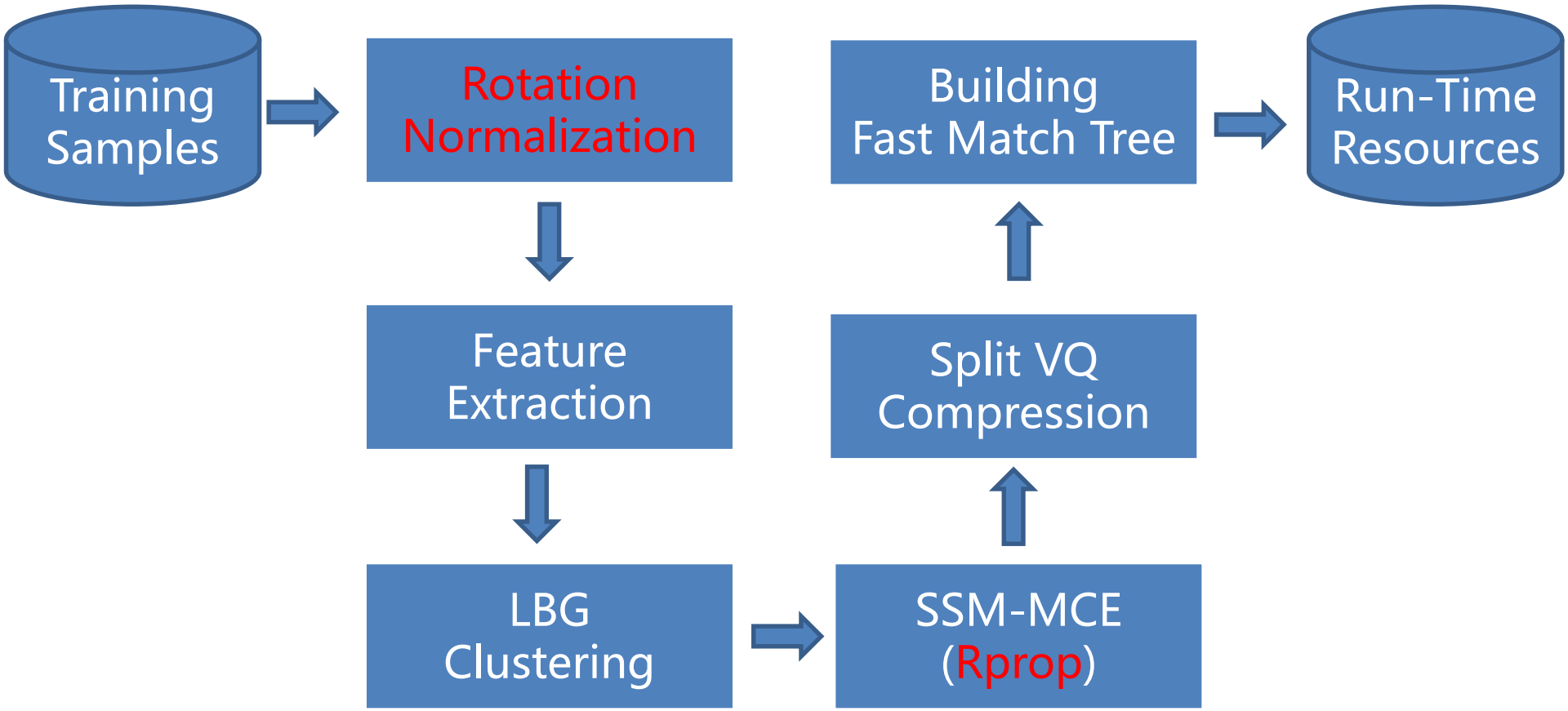


Jun Du, Qiang Huo, Kai Chen
"Designing compact classifiers for rotation-free recognition of large vocabulary online handwritten Chinese characters," ICASSP, 2012

Motivation

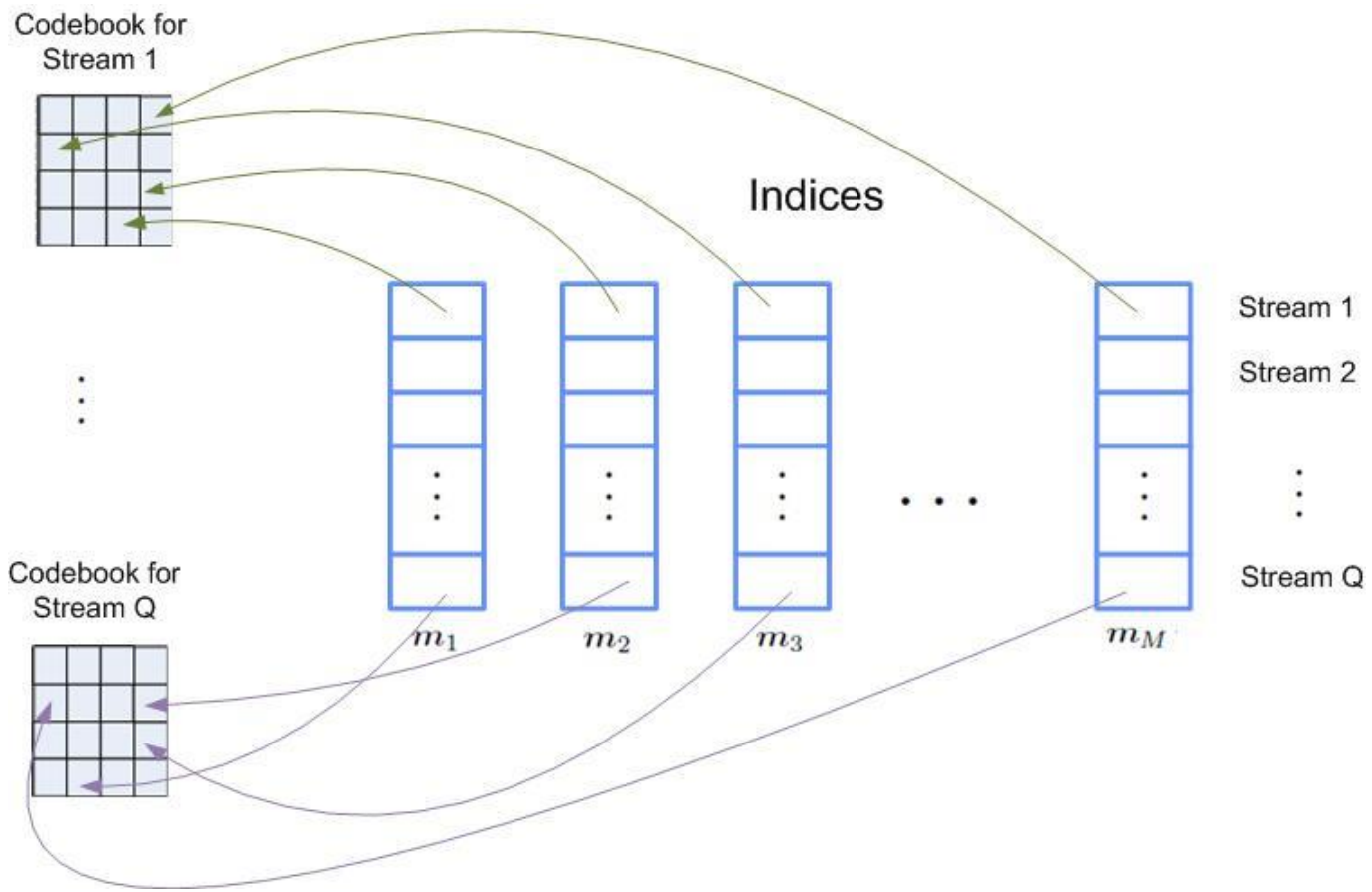
- Requirement of WP8 HWR in China Markets
  - Support GB18030 (27533 China Characters)
  - Robust to rotational distortion
- The Status of HWR Product Team
  - The original team is reorganized to STC
  - The new team have no capability to build the engine

Training Stage

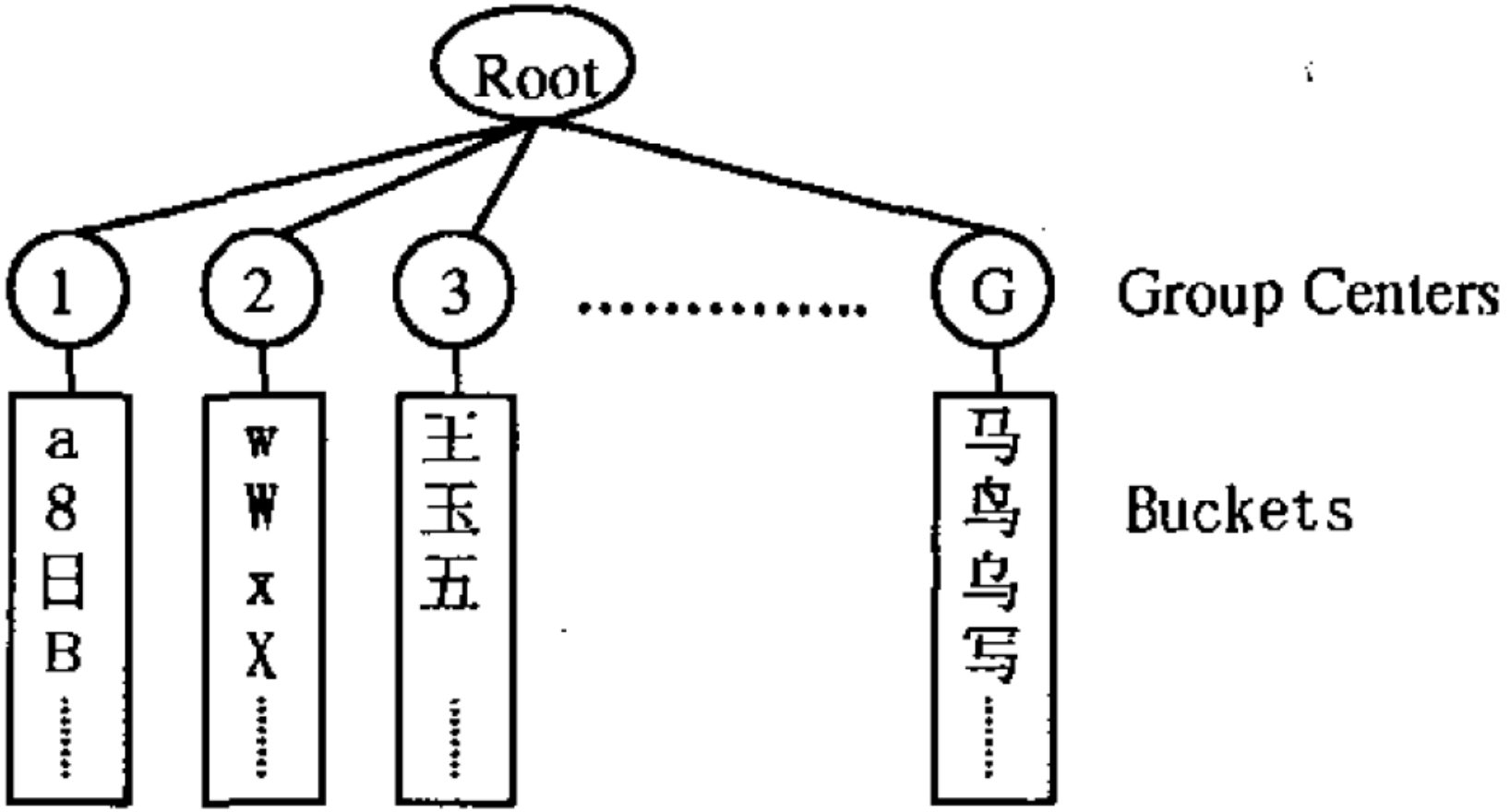


Split VQ Compression

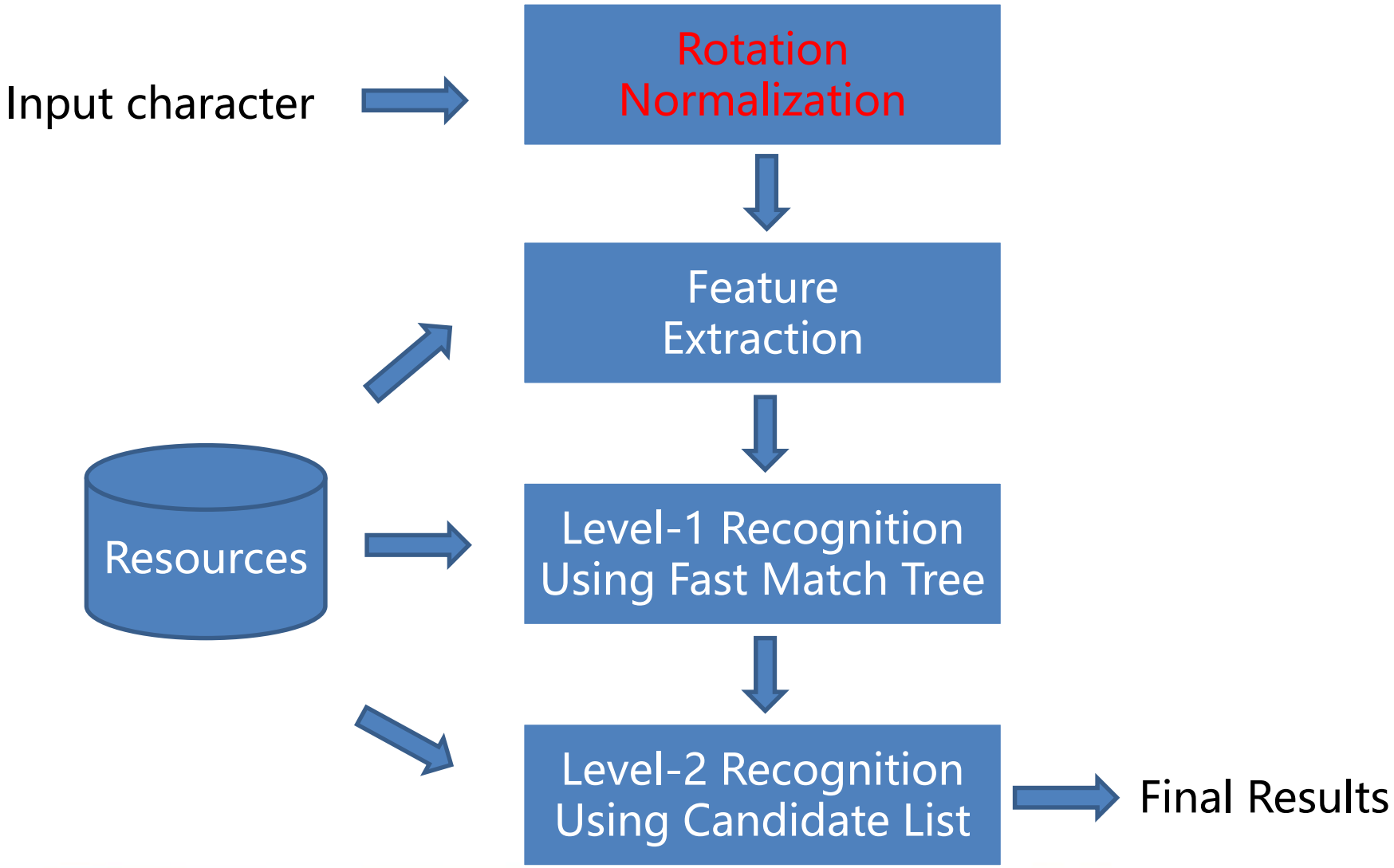
- Split D-dimensional prototype vectors into Q streams
- Sub-vectors in different streams are quantized by VQ with different codebooks



Fast-Match Tree



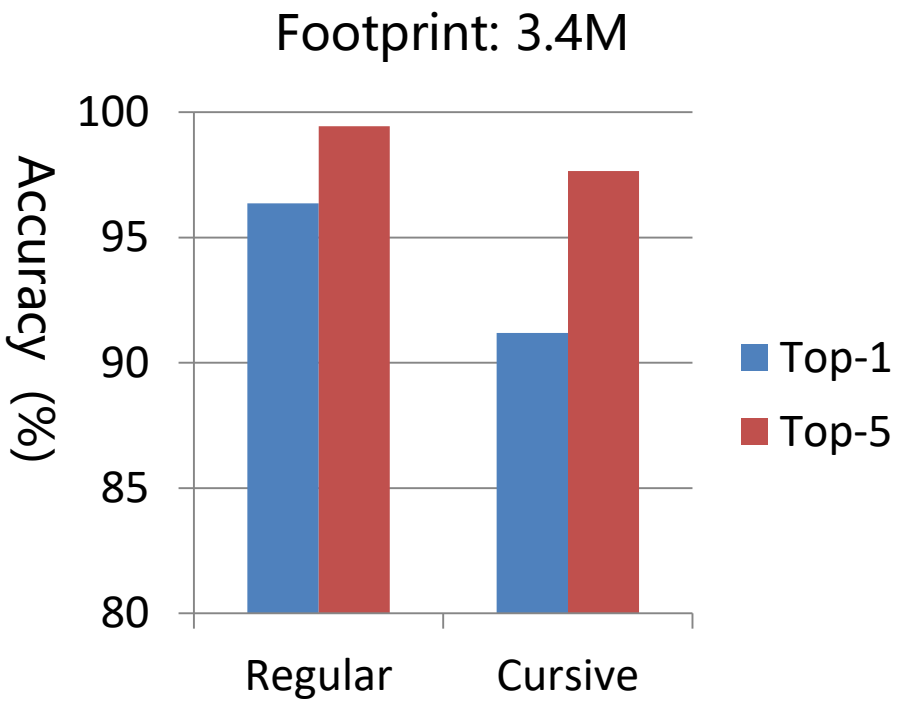
Recognition Stage



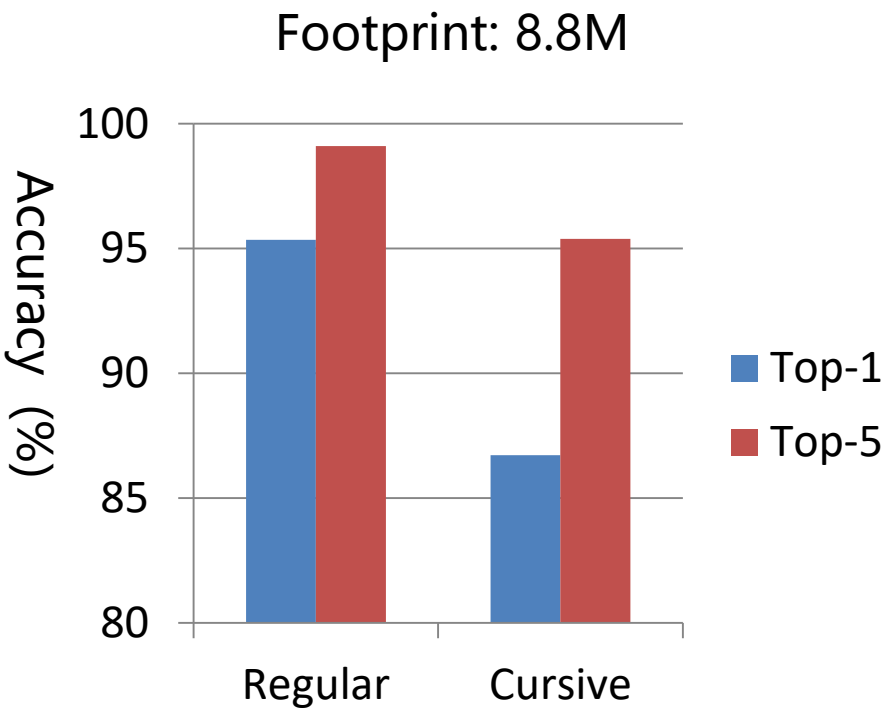
A Magic for Rotation Normalization

- Starting point of each stroke
- Ending point of each stroke
- Averaged starting point over all strokes
- Averaged ending point over all strokes

Rotation-Sensitive vs. Rotation-Free



Recognizer #1: rotation-sensitive



Recognizer #2: rotation-free

Comparison with Mango (WP7.5) System

	Current System	Mango System
Vocabulary	27533 Characters	6763 Characters
Footprint	6MB+	2MB
Accuracy	94.5%	94.2%
Recognition Time Per Character	7ms	11ms

Summary

- A Chinese HWR solution for WP8 Apollo release
  - Rotation-free
  - Fast
  - Compact
  - Accurate
  - Large vocabulary