DOST Dataset

Downtown Osaka Scene Text Dataset

Masakazu Iwamura, Takahiro Matsuda
Naoyuki Morimoto, Hitomi Sato
Yuki Ikeda and Koichi Kise

Osaka Prefecture University

Agenda

1. Introduction
2. Unique Features of DOST Dataset
3. Construction of DOST Dataset
4. Known Issues
5. Evaluation
6. Conclusion

Agenda

Recent Improvement of Scene Text Recognition

Legend:
- IIIT5K 50
- IIIT5K 1k
- IIIT5K None
- SVT 50
- SVT None
- ICDAR2003 50
- ICDAR2003 Full
- ICDAR2003 50k

Recent results are 80+% or even 90+%

This does not mean these methods can read a wide variety of text in the real environment

Text in Real Environment

- ## We mean
  - Text captured without intention (as much as possible)
  - Text not screened so as to be easily read (with regard to resolution, capture angle and so on)

We present DOST Dataset

Agenda

Unique Features of DOST Dataset

1. **Aim: evaluation of methods in the real environment**
   - Not aiming at training classifiers like MJSynth and SynthText datasets

2. **Completely not intentionally captured**
   - The most similar is ICDAR2015 Challenge 4 "incidental scene text" dataset captured with Google Glass
   - DOST is even free from face direction

Unique Features of DOST Dataset

3. **Video dataset captured with omnidirectional camera**

- ICDAR2013 & 2015 Challenge 3: single direction
- YouTube Video (YTV) Dataset: YouTube Videos

4. **Contains multiple images of single word**

Unique Features of DOST Dataset

5. Large scale

- Contains largest number of word Images
  - Excluding synthesized datasets (MJSynth and SynthText)
  - Excluding dataset containing numbers only (Google Streetview House Number dataset)

No. of Images Contained in Existing Datasets

	0	20,000	40,000	60,000	80,000

Dataset	No. of Images
ICDAR2003	509
ICDAR2013 Chal. 2	462
ICDAR2015 Chal. 4	1,670
NEOCR	659
KAIST	3,000
SVT	349
IIIT5K	5,000
COCO-Text	63,686
ICDAR2013 Chal. 3	15,277
ICDAR2015 Chal. 3	27,824
YVT	11,791
DOST	32,147

Almost double
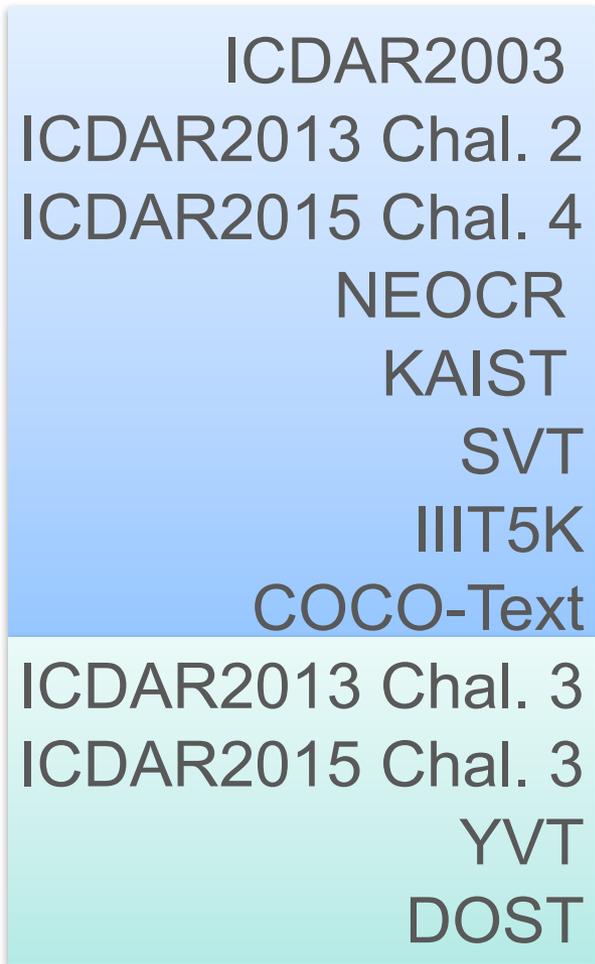
No. of Word Images Contained in Existing Datasets

Image DB

Video DB

Dataset	Count
	0 — 200,000 — 400,000 — 600,000 — 800,000
ICDAR2003	2,268
ICDAR2013 Chal. 2	2,524
ICDAR2015 Chal. 4	17,548
NEOCR	5,238
KAIST	3,000
SVT	904
IIIT5K	5,000
COCO-Text	173,589
ICDAR2013 Chal. 3	93,598
ICDAR2015 Chal. 3	125,141
YVT	16,620
DOST	797,919

Images were captured in shopping streets where a lot of texts exist

x4.6
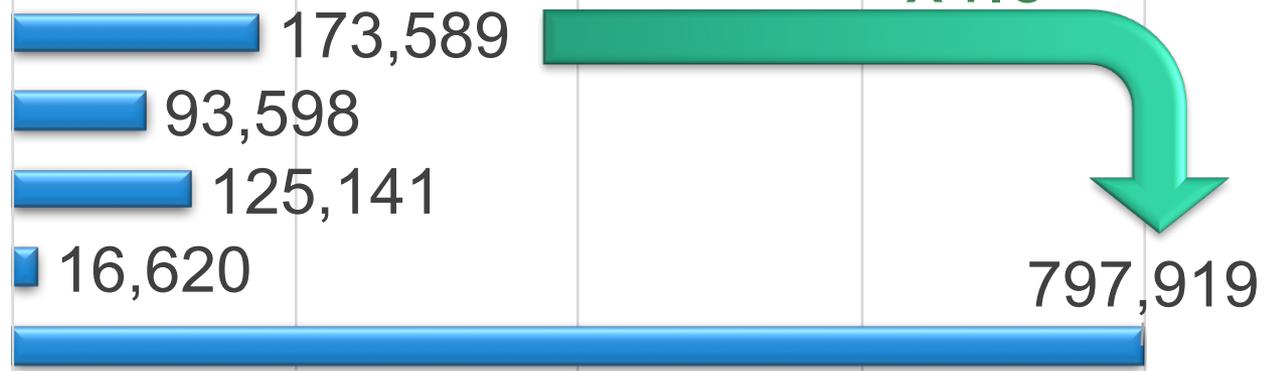
No. of Word Sequences in Existing Video Datasets

	0	5,000	10,000	15,000	20,000	25,000

ICDAR2013 Chal. 3 — 1,962

ICDAR2015 Chal. 3 — 3,562

x6.3

YVT — 245

DOST — 22,398
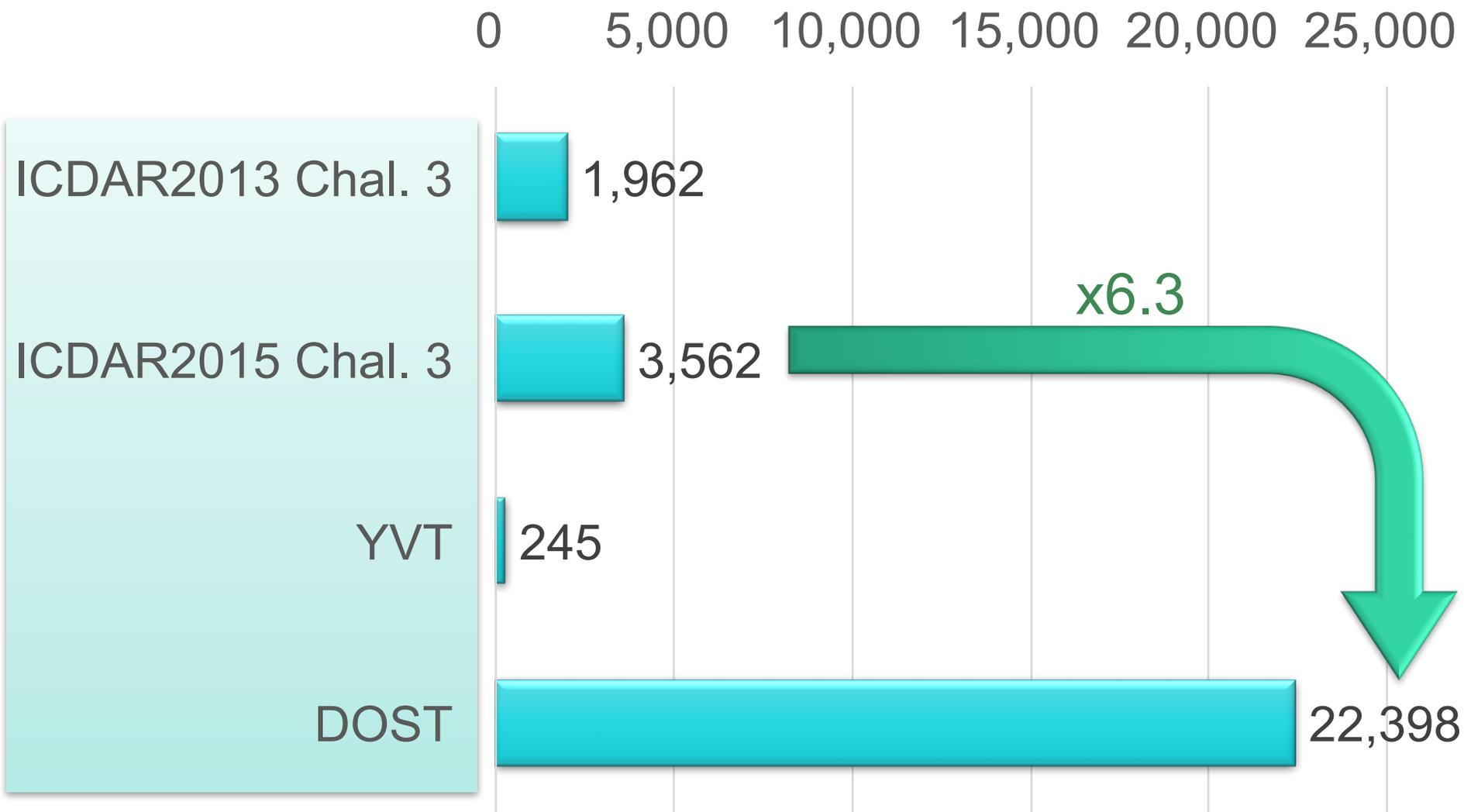
Unique Features of DOST Dataset

6. Contains Japanese characters

| 10円 | 桜子 | テナント募集 | 浪速区 | マッサージ | 印紙 | お弁当 |

- On the other hand, a lot of non-Japanese words are contained

| OPEN | _AMBA | SECOM | ROCKY | Mickey |

No. of Ground Truthed Characters per Category

Category	Count
Alphabet	837,489
Kanji	723,805
Katakana	696,697
Hiragana	355,158
Digit	324,742
Symbol	22,802

Japanese characters
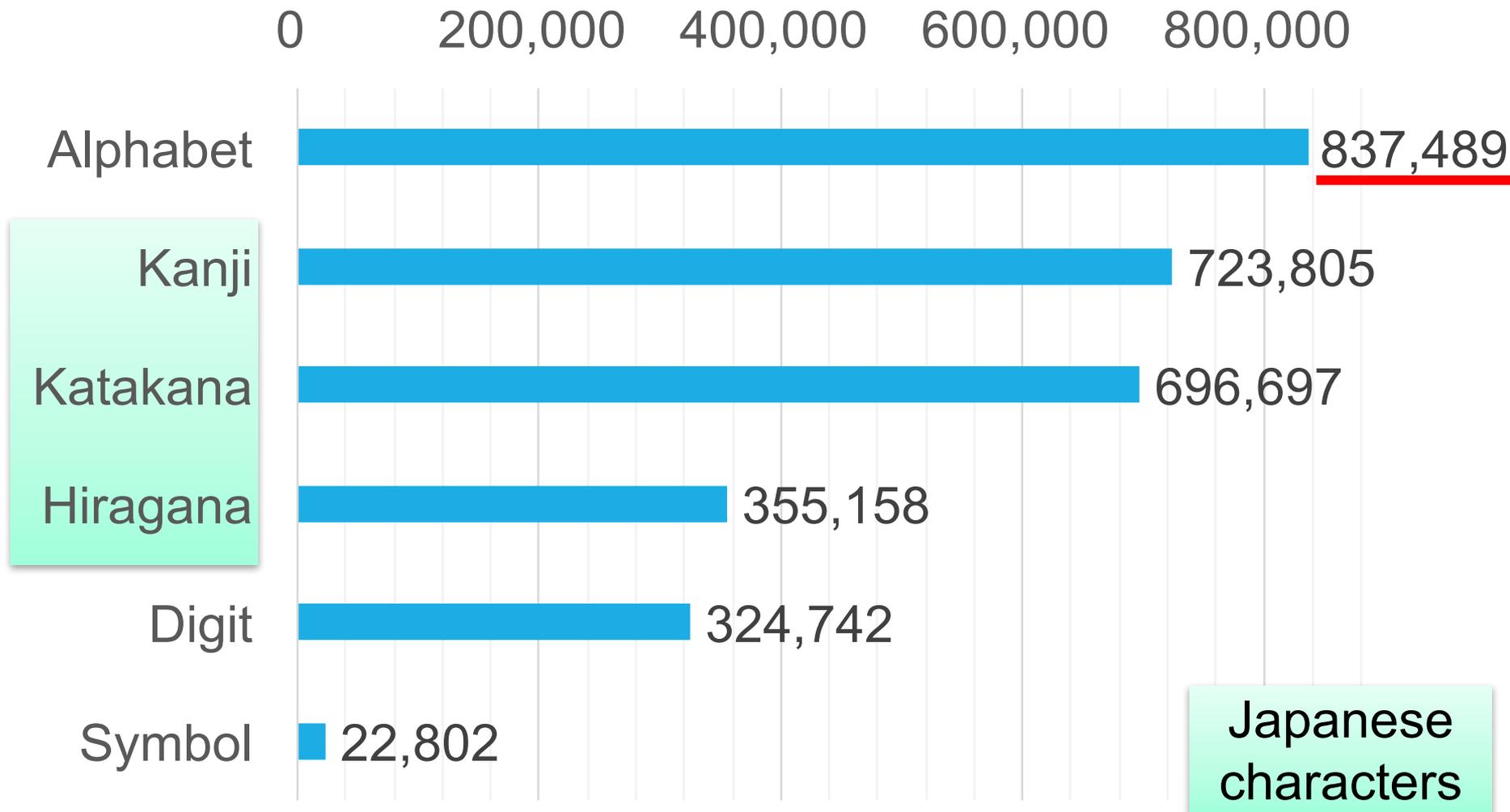
No. of Ground Truthed Characters per Category
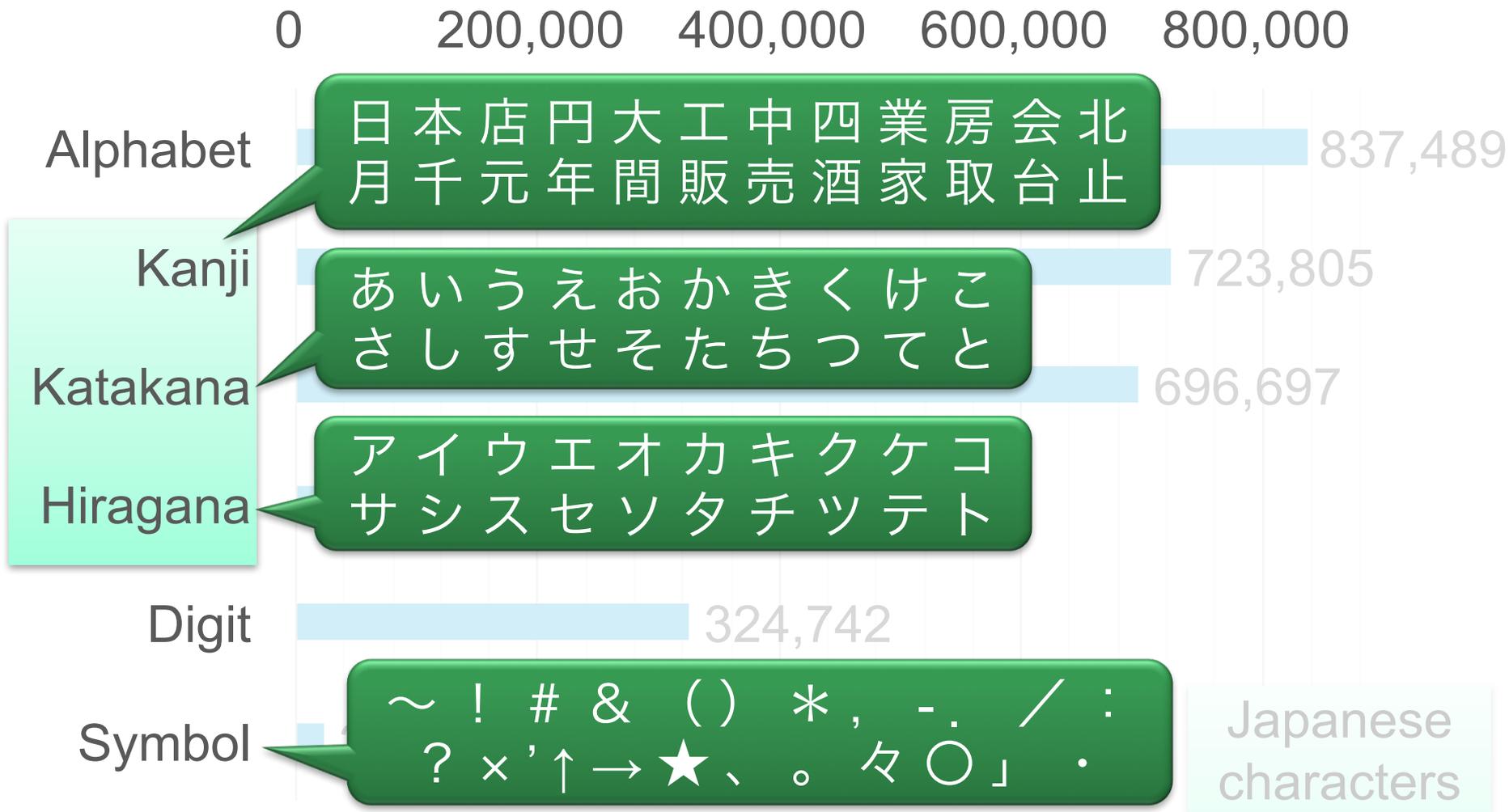
Unique Features of DOST Dataset

7. **Manually ground truthed**

- Amazon Mechanical Turk is not usable
- Hiring students costed a lot!

Agenda

Construction of DOST Dataset

1. Image capture Completed in 2012

- Point Grey Research LadyBug 3
- 1,200x1,600 pixels, 6.5 fps

Upward camera

Five horizontal cameras

Omnidirectional camera

Laptop PC

Battery

Cart
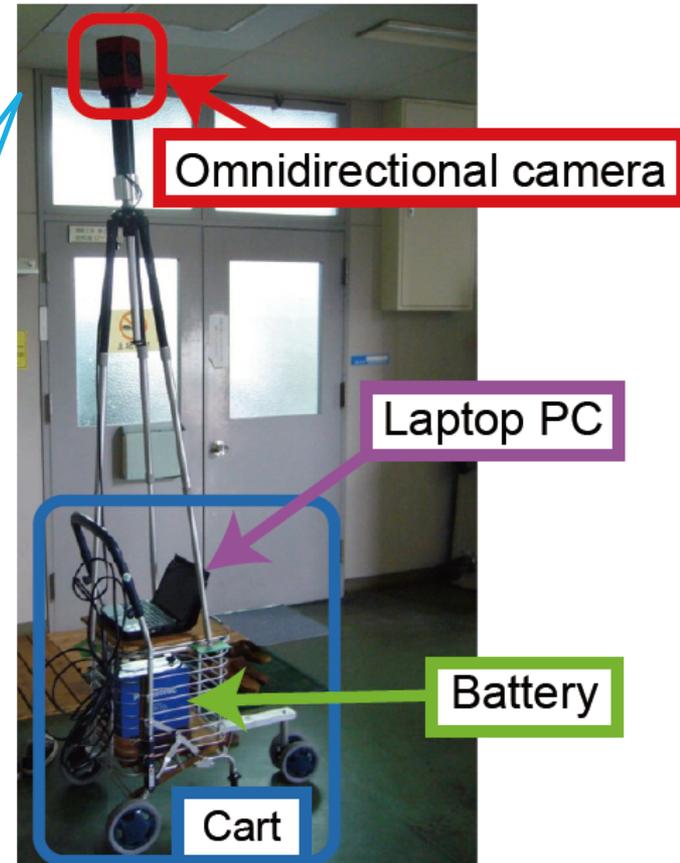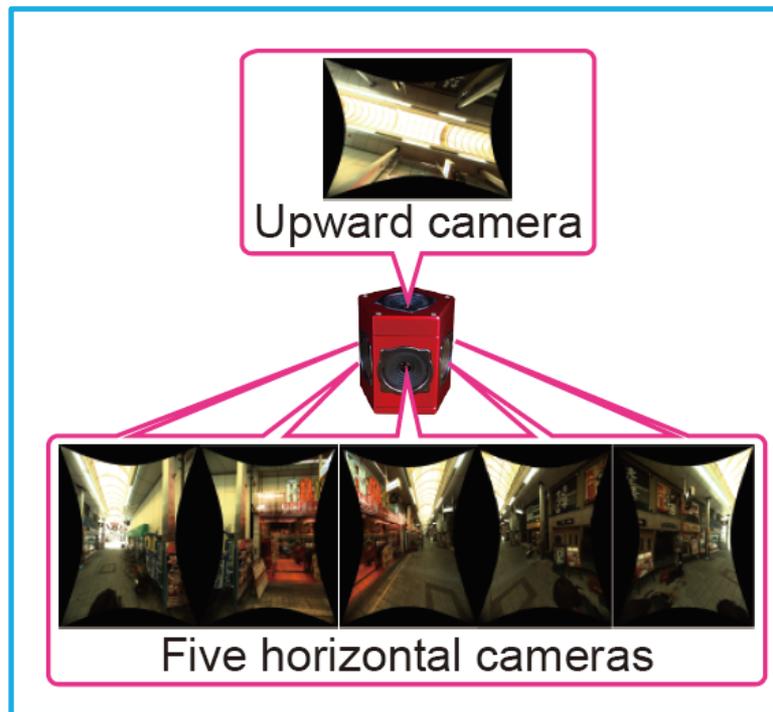
Place, time length, the number of images of capture

| Place	Length [h]	#Image
Sakai-Higashi	0.73	101,874
Namba	3.71	521,988
Shinsaibashi	0.25	35,100
Abiko	0.50	70,614
Tennoji	0.38	53,754
Total	5.57	783,150

Construction of DOST Dataset

2. Manual ground truthing

- Most of GT policies are shared with ICDAR2013 & 2015 Challenge 3 datasets
- GT software was developed
  - Reuse GT information in neighboring frames

3. Privacy preservation

- Faces were blurred
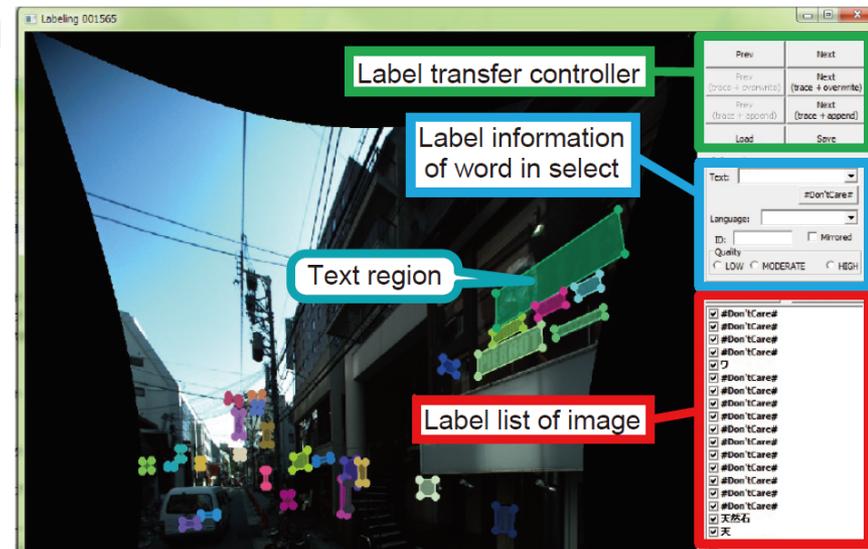
Ground Truthing Policy

- **Basic unit**
  - Word or Bunsetsu (in Japanese)

    Bunsetsu: the smallest unit of words that sounds natural in a spoken sentence

  - Proper noun is not divided

- **Bounding box**
  - Basic unit is represented by its four corners

Ground Truthing Policy

- **Transcription**
  - transcription consists of visible characters

- **Quality**
  - High, mid or low
  - Low corresponds to "Don't care" regions

- **ID**
  - The same ID is assigned to a sequence of same basic units as long as it can be traced
  - Trace ends when a basic unit completely goes out from the frame

Distribution of lengths of image sequences

Length of sequence	#sequence
5001	2
3181	1
2000 - 2009	4
1951	1
1500 - 1501	2
101-582	6
-100	9
Total	27

Agenda

1. Introduction
2. Unique Features of DOST Dataset
3. Construction of DOST Dataset
4. Known Issues
5. Evaluation
6. Conclusion

Known Issues

- **Ground truths are not perfect**
  - Bounding boxes of text regions are not tight enough
  - Ground trothing "Don't care" is not comprehensive

    "Don't care" is marked in illegible regions

  - Some word sequences are broken

- **Relationship between other cameras**
  - Word images in other cameras are not followed

Agenda

1. Introduction
2. Unique Features of DOST Dataset
3. Construction of DOST Dataset
4. Known Issues
5. Evaluation
6. Conclusion

Evaluation: Methods

- **Text detection**
  - OpenCV API
  - Matsuda's method based on NAT method

- **End-to-end text recognition**
  - Google Vision API

Evaluation: Datasets

- **Image datasets**
  - ICDAR2003
  - ICDAR2013 Chal. 2
  - ICDAR2015 Chal. 4
  - SVT
  - COCO-Text

- **Video datasets**
  - ICDAR2015 Chal. 3
  - YVT
  - DOST
  - DOST Latin

Subset of DOST which contain words consisting of alphabets and digits

Data were sampled

Text Detection by OpenCV API

Image DB
Video DB

F-measure [%]

Image DB	F-measure [%]
ICDAR2003	18.7
ICDAR2013 Chal. 2	6.1
ICDAR2015 Chal. 4	13
SVT	19
COCO-Text	11.9
ICDAR2015 Chal. 3	8.5
YVT	28.5
DOST	2.4
DOST Latin	1.2
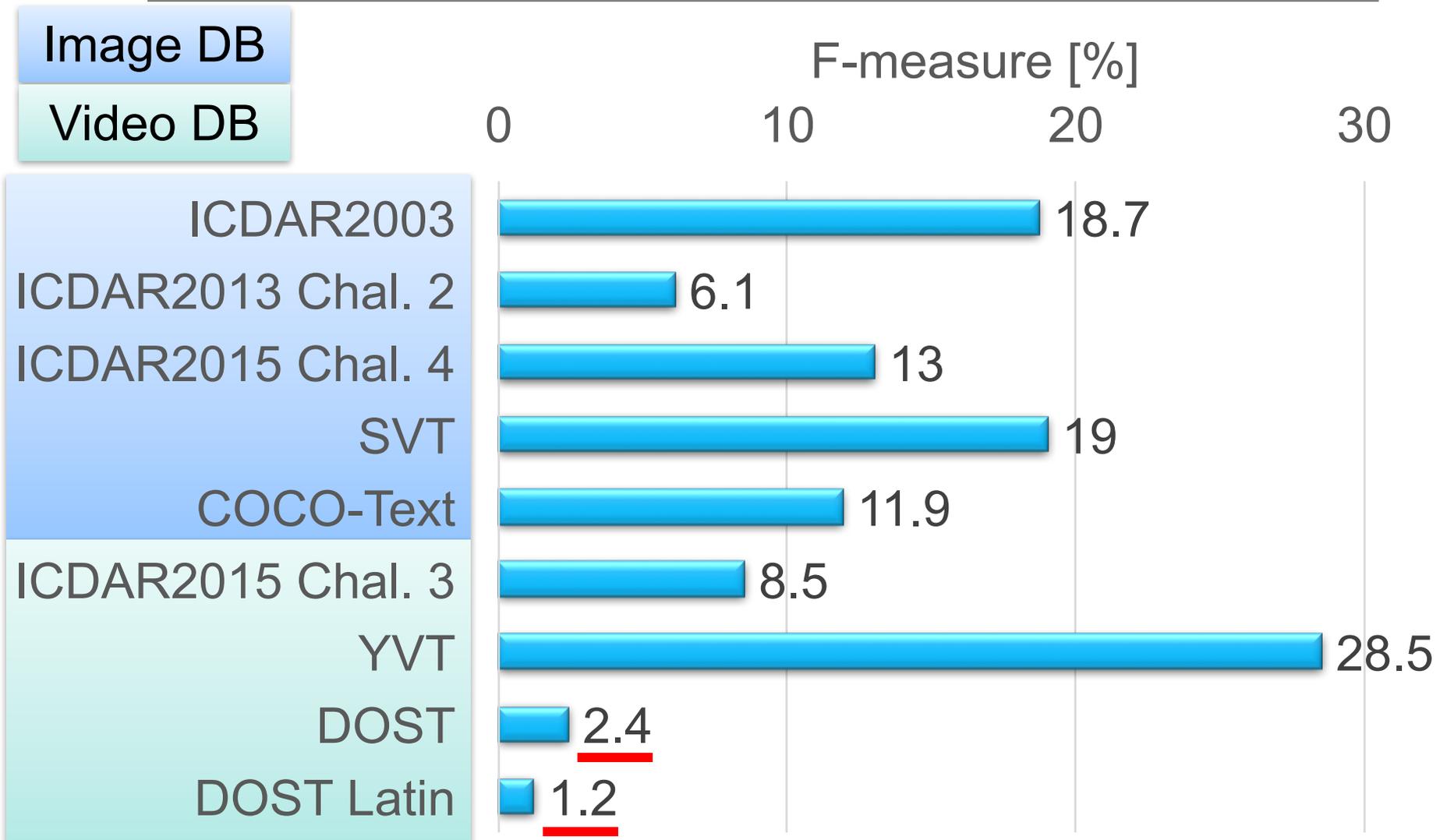
Text Detection by Matsuda's method

F-measure [%]

Database	F-measure [%]
ICDAR2003	47.5
ICDAR2013 Chal. 2	4.8
ICDAR2015 Chal. 4	6.3
SVT	29.1
COCO-Text	1.5
ICDAR2015 Chal. 3	3.9
YVT	1.9
DOST	2.8
DOST Latin	2.1
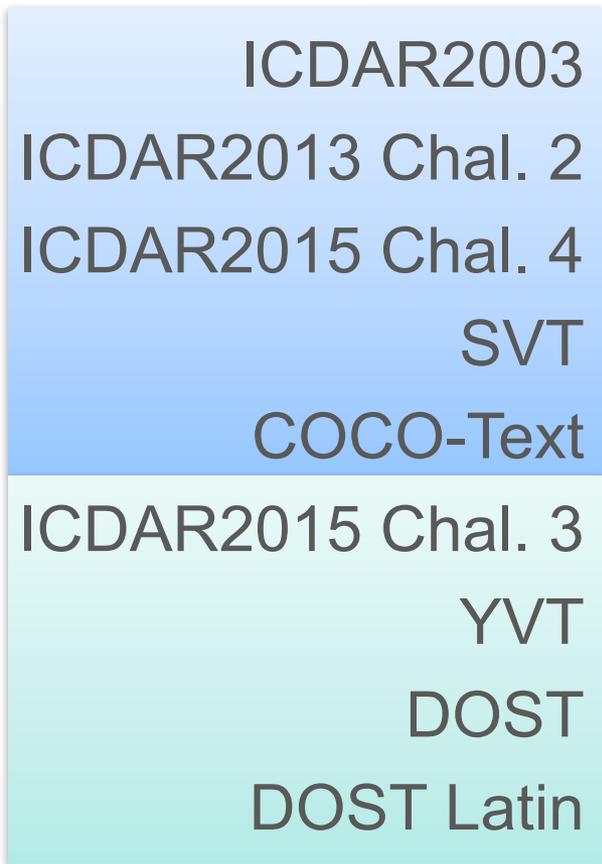
End-to-end Text Recognition by Google Vision API

Image DB

Video DB

F-measure [%]

ICDAR2003	81.8
ICDAR2013 Chal. 2	71.3
ICDAR2015 Chal. 4	48.5
SVT	24.2
COCO-Text	17.1
ICDAR2015 Chal. 3	44.1
YVT	37.7
DOST	2.7
DOST Latin	11.2

Recognized in Japanese mode

Agenda

1. Introduction
2. Unique Features of DOST Dataset
3. Construction of DOST Dataset
4. Known Issues
5. Evaluation
6. Conclusion

Conclusion

- **DOST dataset is presented**
  - Has unique features
  - More challenging than existing datasets

Thank you for your attention!!