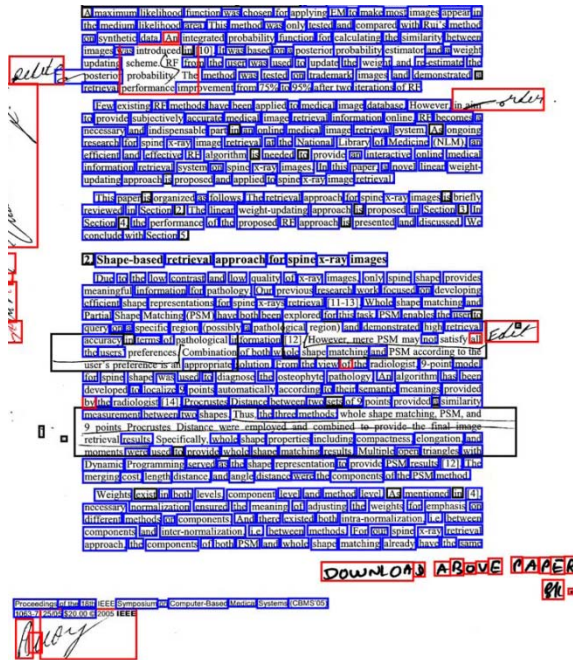


ICDAR 2009 Best Paper Award

Markov Random Field Based Text Identification from Annotated Machine Printed Documents

Xujun Peng¹, Srirangaraj Setlur¹, Venu Govindaraju¹, Ramachandhula Sitaram² and Kiran Bhuvanagiri²

¹Center for Unified Biometrics and Sensors, University at Buffalo ²HP Labs, Bangalore, India

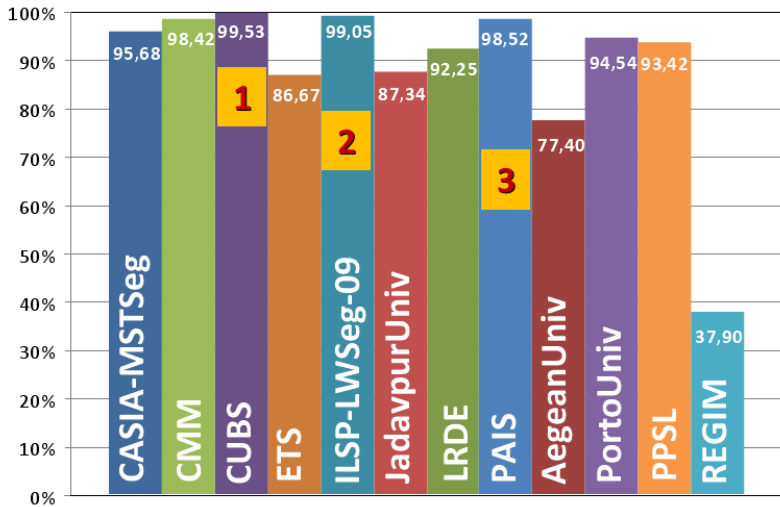


Conference Statistics:
Total submissions: 430
Oral papers: 87 (20%)

Joint work with HP Labs under the
HP Innovation Research Award 2008/09 for
**Intelligent Processing of Hand-Annotated
Documents**

Labeled result from system

ICDAR 2009 Handwriting Segmentation Contest



- CASIA-MSTSeg** - Chinese Academy of Sciences, Beijing, China
- CMM** - Center of Mathematical Morphology, Paris, France
- CUBS** - Center for Unified Biometrics and Sensors, Buffalo, USA
- ETS** - Ecole de Technologie Superieure, Montreal, Canada
- ILSP-LWSeg-09** - Institute for Language and Speech Processing, Athens, Greece
- JadavpurUniv** - Jadavpur University, Kolkata, India
- LRDE** - EPITA Research and Development Laboratory, France
- PAIS** - ECNU-SRI Lab for Pattern Analysis and Intelligence Systems, Shanghai, China
- AegeanUniv** - University of Aegean, Samos, Greece
- PortoUniv** - University of Porto, Porto, Portugal
- PPSL** - University of Mysore, Mysore, India
- REGIM** - University of Sfax, Tunisia

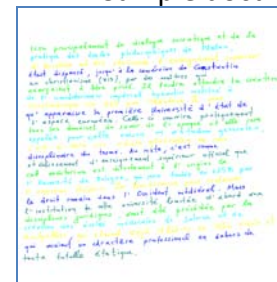
Segmentation contest test data:

- 200 handwritten documents
- Multiple writers
- Multiple languages

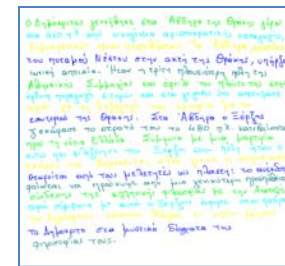
CUBS method (Z. Shi, S. Setlur, V. Govindaraju)

- Best line segmentation performance among 12 entries

Sample documents from the test set



French



Greek



Similar performance seen on Arabic documents (DARPA data)