

16th International Conference on Frontiers in Handwriting Recognition

ICFHR 2018

Niagara Falls, USA, August 5 - 8, 2018



Competition on Document Image Analysis Tasks for Southeast Asian Palm Leaf Manuscripts

**Made Windu Antara Kesiman^{1,2}, Dona Valy^{3,4}, Jean-Christophe Burie¹, Erick Paulus⁵, Mira Suryani⁵,
Setiawan Hadi⁵, Michel Verleysen³, Sophea Chhun⁴, and Jean-Marc Ogier¹**

¹Laboratoire Informatique Image Interaction (L3i), Universite de La Rochelle, La Rochelle, France

²Laboratory of Cultural Informatics (LCI), Universitas Pendidikan Ganesha, Singaraja, Bali, Indonesia

³Institute of Information and Communication Technologies, Electronic, and Applied Mathematics (ICTEAM), Universite Catholique de Louvain, Louvain, Belgium

⁴Department of Information and Communication Engineering, Institute of Technology of Cambodia, Phnom Penh, Cambodia

⁵Department of Computer Science, Universitas Padjadjaran, Bandung, Indonesia





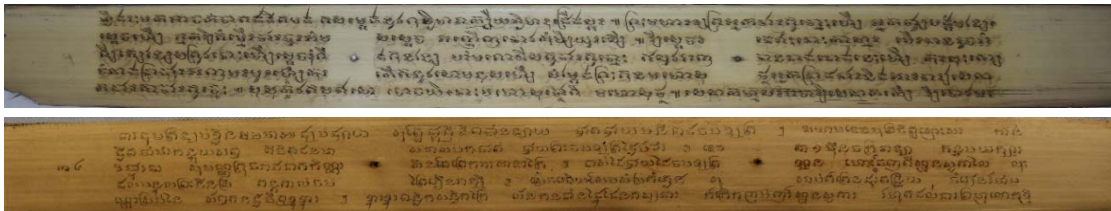
Palm Leaf Manuscripts Corpus

- Balinese Palm Leaf Manuscripts – Bali - Indonesia



23 different collections from 5 different locations (regions)

- Khmer Palm Leaf Manuscripts - Cambodia



from Buddhist temples in different locations throughout Cambodia: Phnom Penh, Kandal, and Siem Reap

- Sundanese Palm Leaf Manuscripts – West Java - Indonesia



from Situs Kabuyutan Ciburuy, Garut, West Java



Challenges: Document Image Analysis

- Challenge A: Binarization
 - 1 Track Mixed
- Challenge B: Text Line Segmentation
 - 1 Track Mixed
- Challenge C: Isolated Character/Glyph Recognition
 - 3 Single Tracks
- Challenge D: Word Transliteration
 - 3 Single Tracks and 1 Track Mixed

PALM LEAF MANUSCRIPT DATASETS FOR BINARIZATION TASK

Manuscripts	Train	Test	Dataset
Balinese	50 pages	50 pages	Extracted from AMADI_LontarSet ¹ [3], [9]
Khmer	23 pages	23 pages	Extracted from EFEO ² [13]
Sundanese	31 pages	30 pages	Extracted from Sunda Dataset ICDAR2017 [11]

PALM LEAF MANUSCRIPT DATASETS FOR TEXT LINE SEGMENTATION TASK

Manuscripts	Train	Test	Dataset
Balinese	47 pages	49 pages	Extracted from AMADI_LontarSet [9]
Khmer	50 pages	200 pages	Extracted from SleukRith Set [10]
Sundanese	31 pages	30 pages	Extracted from Sunda Dataset [11]

PALM LEAF MANUSCRIPT DATASETS FOR WORD TRANSLITERATION TASK

Manuscripts	Train (images)	Test (images)	Text	Dataset
Balinese	15,022 from 130 pages	10,475 from 100 pages	Latin	AMADI_LontarSet [3], [9]
Khmer	16,333 (part of 657 pages)	7,791 (part of 657 pages)	Latin and Khmer	SleukRith Set [10]
Sundanese	1,427 from 20 pages	318 from 10 pages	Latin	Sunda Dataset [11]

PALM LEAF MANUSCRIPT DATASETS FOR ISOLATED CHARACTER/GLYPH RECOGNITION TASK

Manuscripts	Classes	Train (images)	Test images)	Dataset
Balinese	133	11,710	7,673	AMADI_LontarSet [3], [6], [9]
Khmer	111	113,206	90,669	SleukRith Set [10]
Sundanese	60	4,555	2,816	Sunda Dataset [11]

The Winners !



- Challenge A: Binarization

Deepak Kumar,

from *Department of Electronics & Communication Engineering, Dayananda Sagar Academy of Technology and Management (DSATM), Bengaluru, India*

- Challenge B: Text Line Segmentation

No winner

- Challenge C: Isolated Character/Glyph Recognition

Zi-Rui Wang, Jun Du and Wen-Chao Wang,

from *University of Science and Technology of China, China*

- Challenge D: Word Transliteration

Jianshu Zhang, Jun Du, and Lirong Dai,

from *National Engineering Laboratory for Speech and Language Information Processing, University of Science and Technology of China, China*