Litis | L3i | 〈BnF

http://icdar2017hba.litislab.eu/

ICDAR 2017 Competition on Historical Book Analysis

Context
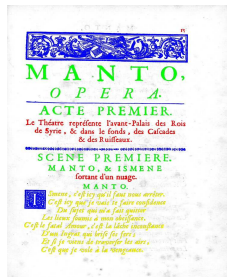
In conjunction with ICDAR'17, the **HISTORICAL BOOK ANALYSIS COMPETITION (HBA)** is organized. The HBA competition will address a thriving topic of major interest of many researchers in different fields including **(historical) document image analysis**, **image processing**, **pattern recognition** and **classification**.

The HBA competition will provide a **large experimental corpus** and a **thorough evaluation protocol** to ensure **a consistent comparison of image processing methods** for historical document image analysis.

A new dataset which is called the **HBA dataset** will be released at this occasion. The HBA dataset is composed of **4436 real scanned ground-truthed one-page historical document images** (**2435 and 2001 printed and manuscript pages, respectively**) from **11 books** (**6 manuscripts** and **5 printed books**) in different languages and scripts published between the 13th and 19th centuries.

Challenges

Two nested challenges are proposed in the HBA competition.

1- The HBA competition will aim at evaluating how image analysis methods could **discriminate the textual content from the graphical one** at **pixel level**.

2- It will aim at assessing the capabilities of the participating methods to **separate the textual content according to different text fonts** (e.g. lowercase, uppercase, italic…) at **pixel level**.

Ground truth

The documents of the HBA dataset are **gray-scale or color** images which were digitized at **300 or 400 dpi** and saved in the **TIFF format** which provides a high resolution of digitized images. Each selected foreground pixel is marked by a color that symbolizes the corresponding content type. **The ground truth information is currently available at the pixel level.**

Evaluation protocol

The HBA dataset is divided into 2 sub-datasets, the **sample dataset** and the **evaluation dataset**.

The sample dataset will contain 2 books while the evaluation dataset will be composed of the 9 remaining books. Each dataset will be composed of a set of **training images** and a set of **test images**.

Several per-pixel classification accuracy metrics, including *precision, recall, classification accuracy* and *F-measure* will be performed.

Important dates

Date	Description
April 01, 2017	– Opening of the registration to competition – Publication of the sample dataset
June 01, 2017	– Closure of the registration to competition – Publication of the evaluation dataset – Beginning of the competition
July 31, 2017	– Deadline of the results submission – Deadline of the submission of the description of the participating methods
August 31, 2017	– Sending the competition results to the participants

The results of the HBA competition will be announced at **HIP'17** and **ICDAR'17**.

How to participate

1- Register your interest through the **registration form** (**before 01 June 2017**).

2- Specify clearly in which challenge you would like to participate (**challenge 1 only, challenge 2 only, both challenges 1 and 2**).

3- Download the **sample dataset** (**available from 01 April 2017**).

4- Download the **evaluation dataset** (**available from 01 June 2017**).

5- Submit the **description** and the **results** of your methods (**before 31 August 2017**).

Organizers

Maroua Mehri[1]
Pierre Héroux[1]
Rémy Mullot[2]
Jean-Philippe Moreux[3]
Bertrand Coüasnon[4]
Bill Barrett[5]

[1] LITIS, University of Rouen Normandy, France

[2] L3i, University of La Rochelle, France

[3] BnF - French national library, France

[4] Intuidoc, IRISA Rennes, France

[5] Family History Technology Lab, Brigham Young University, USA

Contact us

icdar2017hba@litislab.eu

Follow us

Twitter: @icdar2017hba