



LEHRSTUHL FÜR
MUSTER-
ERKENNUNG

Master Thesis

Image Detection and Text Localization in Historical Documents

The analysis of historical documents typically involves a segmentation of all its structural parts. However, given a large document dataset, a complete image segmentation for a query like “give me all documents containing an image” might not be necessary and more efficient solutions may exist.

Goals of this Thesis:

- Classify each document image in either consisting of text only, or containing an image. Instead of segmenting the image, the approach suggested is to follow a common retrieval pipeline, i. e. to compute global image descriptors which are eventually classified.
- Localize the text in the document image. Various methods for detecting text in street images have been proposed recently. One such approach is going to be adapted to historical documents.

Requirements:

- Good knowledge in either C++ and / or Python.
- Helpful: Knowledge in Computer Vision / Pattern Recognition.

Associated Projects:

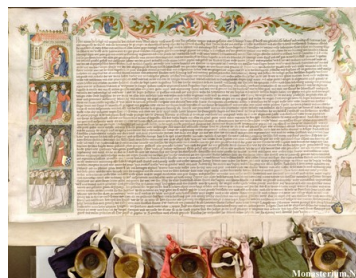
- *Illuminated Charters*, <http://illuminierte-urkunden.uni-graz.at/>
- *Papal Charters from the Middle Ages*, <http://www5.cs.fau.de/puhma>

Interested?

Please contact Vincent Christlein <vincent.christlein@fau.de>.



Stiftsarchiv Altenburg, Urkunden
Altenburg, 1288 IV 05,
monasterium.net



Stiftsarchiv Herzogenburg, Urkunden
Dürnstein, Augustiner-Chorherren,
1410 II 17, monasterium.net