READ
RECOGNITION & ENRICHMENT OF ARCHIVAL DOCUMENTS

D8.2
Open Innovation Forum P2 (DocScan and ScanTent)

Florian Kleber, Markus Diem, Stefan Fiel and Günter Mühlberger
CVL

Distribution: http://read.transkribus.eu/

READ
H2020 Project 674943

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 674943

Project ref no.	H2020 674943
Project acronym	READ
Project full title	Recognition and Enrichment of Archival Documents
Instrument	H2020-EINFRA-2015-1
Thematic priority	EINFRA-9-2015 - e-Infrastructures for virtual research environments (VRE)
Start date/duration	01 January 2016 / 42 Months

Distribution	Public
Contract. date of delivery	31.12.2017
Actual date of delivery	28.12.2017
Date of last update	11.12.2017
Deliverable number	D8.2
Deliverable title	Open Innovation Forum P2 (DocScan and ScanTent)
Type	Report, Demonstrator
Status & version	in progress
Contributing WP(s)	WP8
Responsible beneficiary	CVL
Other contributors	CVL, UIBK
Internal reviewers	UCL, StAZh
Author(s)	Florian Kleber, Markus Diem, Stefan Fiel and Günter Mühlberger
EC project officer	Martin MAJEK
Keywords	Crowd-Scanning, Android App, Transkribus

Contents

1 Executive Summary

The Open Innovation Forum focus on the development of the ScanTent and the DocScan app due to the positive feedback of scholars and project partners. An open source Android app for the scanning of documents has been developed. Transkribus DocScan detects automatically the page area of a document and gives feedback on the image according to defined quality criteria (page area, perspective distortion, rotation, sharpness). The main innovative feature is the auto-shoot mode, which captures a picture if all criteria are fulfilled and a page is turned over. The app will integrate batch processing to allow the scanning of entire documents.Transkribus DocScan is connected to Transkribus to allow the user to log in and upload the images to the Transkribus Cloud. In addition to the app a Transkribus ScanTent has been developed which is a universal scanning device providing homogeneous conditions (e.g. uniform illumination) for digitizing documents with a mobile phone. The development of the Crowd-Scanning app was originally foreseen as a sub-contract for the Russian company ABBYY. This plan was changed due to the fact that CVL was able to reuse recently developed technology. The app is be provided as open source on GitHub. The task of improving the functionality of DocScan and the ScanTent has been continued under Task 8.2.

Section 2 describes the first prototype of the ScanTent developed at the end of 2016 and during 2017. DocScan and an evaluation of its results are presented in Section 3. Future work is presented in Section 5.

2 ScanTent

Based on the first experiments in 2016, the characteristics of the prototype developed in 2017 are described in this section. Also images of the prototype are shown.

Short Summary: If a digital camera is used to digitize documents, apiece of equipment is needed to fulfill the following criteria: (1) the distance should be kept constant if multiple pages are scanned (results in the same resolution for every image), (2) the camera position should be fixed to avoid motion blur, and (3) a non-destructive homogenous illumination should be present (since sufficient light is often not given in a usual reading room). Additionally, since documents are often present in a bound form and cannot be opened in the same way as a modern book (due to the condition), the use of two hands make it easier to fix the documents position and flatten the pages. Thus, the ScanTent[1], a cheap (costs will be about 100 Euro) transportable device, has been developed, which fulfills the aforementioned criteria.

An image of the ScanTent prototype is shown in Figure 1. The bottom part of the ScanTent consists of a ruled base to provide a homogeneous background and to detect the dimensions of a document. The ScanTent fullfils the following properties: (1) portability, (2) fixed distance between document and camera, (3) illumination system with polarized light and polarization filter in front of the camera to avoid specular reflections, (4) a base with a ruling and (5) allows to shoot hands-free.

[1] `https://scantent.caa.tuwien.ac.at/en/`

Figure 1: The ScanTent in the library.

The Depth-of-Field (DoF) is dependent on the mobile phone's camera and the distance of the object to the camera. For example, the camera of the Samsung Galaxy S6 has a focal length of 4.3 mm, aperture size of F/1.9 (sensor diagonal of about 6.8mm). The Circle-of-Confusion (CoC) is estimated as $sensor - diagonal/1500$. The distance between the document and camera is about 45cm because of the ScanTent. According to Vaquero et al. [1] the depth of field limits D_{near} and D_{far} can be calculated as follows:

$$D_{near} = \frac{sf^2}{f^2 + Nc(s-f)}; D_{far} = \frac{sf^2}{f^2 - Nc(s-f)} \quad (1)$$

where f is the focal length, s is the focused distance, N is the lens' f-number and c is the CoC. This results in depth of field limits of $D_{near} = 37cm$ and $D_{far} = 56cm$ for the given example. Thus, the entire area of a document is in focus, even if the document is curved.

The full workflow and the selection of all parts for the manufacturing of the ScanTent have been realized in 2017. The first 12 prototypes for testing have been delivered to:

- University Archive Greifswald
- Niederösterreichisches Landesarchiv
- Staatsarchiv des Kantons Zürich (StAZh)
- Goethe-Uni Frankfurt
- BBAW Berlin

- UIBK
- scholars (historians) 2x
- UCL
- NAF
- CVL 2x

Experiments of the historians in Innsbruck showed, that 800 pages could be scanned in 4 hours by the combination of the ScanTent and DocScan. The digitisation department of StAZH has analyzed the ScanTent and written a report about the setting and the lighting which is available in the Wiki.

3 DocScan

DocScan is a document scanner app, which has a live view and detects in real time the document page. Furthermore, it evaluates if a picture is in focus to give the user feedback and to assure a certain picture quality which will be used for further processing with document image analysis methods. Additionally, it learns a background model to detect page turns and automatically take new pictures. Thus, a book can be scanned by flipping through the pages. The DocScan app is directly accommodated to the ScanTent.

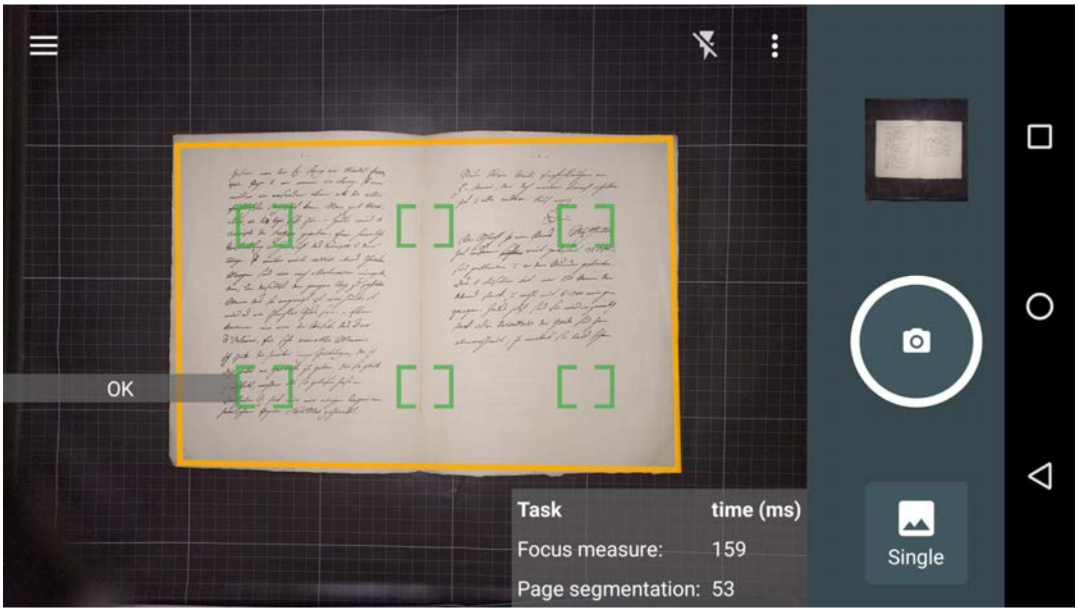

Figure 2: Screenshot of the DocScan app with page detection and focus measure visualized.

Figure 2 shows a screenshot of DocScan. The yellow rectangle shows the detected page and the green markers show areas in focus (red indicates out-of-focus sections in the document area). The time to calculate the page detection is on average 72 ms and for the focus measure 62 ms (on a Samsung S6). The app is developed in Android Studio (Version 2.1). To allow efficient image processing OpenCV library (Version 3.1.0) is used. The OpenCV functions are called via Java Native Interfaces (JNI). A (Gradle) build script is used to compile the native C++ source files and OpenCV functions with Android SDK. The camera is addressed using the android.hardware.camera class which is backward compatible with Android 4.0. DocScan is open source and available on GitHub[2]. The following sections describe the page detection, the focus measure and the background model and the evaluation of the page detection. For a detailed description see [2].

3.1 Page Detection

The page segmentation is blob based and is designed in favor of speed rather than accuracy. The polygons of white blobs which are generated using multiple thresholds are analyzed. Quasi rectangular polygons are then compared to find the best hypothesis page region.

The page detection is evaluated on the ICDAR2015 SmartDoc competition dataset [3]. The detailed evaluation is presented in Section 3.4.

3.2 Focus Measure

To determine if a document region is in focus, the image is subdivided into smaller regions and based on a Focus Measure Operator (FMO) [4] the in-focus-value is calculated. The FMO determines for every pixel in the image the level of focus. Pertuz et al. [4] have defined 6 operators for Shape-from-Focus: (1) gradient-based, (2) laplacian-based, (3) wavelet-based, (4) statistics-based, (5) DCT-based and (6) miscellaneous operators. For a detailed description see [4]. Due to computational restrictions on a mobile phone Brenner's FMO [5, 6] (gradient based approach) is used.

3.3 Series Mode

To detect if a page is turned a background model is learned. This is based on the "*Improved adaptive gaussian mixture model for background subtraction*" from Zivkovic [7, 8]: In Zivkovic a background subtraction method is proposed, that makes use of Gaussian mixture models (GMM). Each pixel is modeled as a mixture of Gaussians and the parameters and number of components of the GMM's are constantly adapted in an online procedure. In our implementation two background models are learned: The first background model is used to determine changes stemming from sudden movements - such as page flipping. If no change is detected, the current image is added to the second background model. This, the second background model recognizes a page flip if 10% of the image content are foreground pixels. In this case the series mode shoots a new image and the two background models are reinitialized with the new image.

[2] `https://github.com/TUWien/DocScan`

Table 1: Jaccard index of the proposed (CVL) page segmentation and the winner of SmartDoc (LRDE).

	CVL JI	LRDE JI
Overall JI	0.7572	0.9716
Background 01	0.8907	0.9869
Background 02	0.8013	0.9775
Background 03	0.8463	0.9889
Background 04	0.8277	0.9837
Background 05	0.0138	0.8613
datasheet	0.8189	0.9758
letter	0.7852	0.9718
magazine	0.7397	0.9707
paper	0.7605	0.9715
patent	0.7227	0.9698
tax	0.7086	0.9696

3.4 Results

The page segmentation is evaluated on the ICDAR2015 Competition on Smartphone Document Capture and OCR (SmartDoc) dataset [3]. Challenge-1 deals with the detection of page outlines on images acquired with mobile devices. The dataset comprises 6 different document types and videos of the documents are taken with 5 different backgrounds. The videos are Full HD (1920 x 1080) and comprise about 25.000 frames in total. For a detailed description see [3]. As evaluation measure the Jaccard Index (JI) was proposed, which is a measure for the overlapping of the detected quadrilaterals:

$$JI = \frac{area(GT \cap DP)}{area(GT \cup DP)} \tag{2}$$

where GT defines the Ground Truth (GT) polygon of the page and DP defines the Detected Polyon. The JI has a range from 0 to 1, where 1 is the best segmentation possible. The results of the page detection for the proposed method (CVL) and the winner of SmartDoc (LRDE) [3] are shown in Table 1.

Background 01 to Background 04 have a JI in the range from 80% to 89%. Only Background 05 has a JI from 1.38%. The reason that most images of Background 05 fail is the obstruction of the document with pens and cables. Due to the occlusions no documents are detected in the images.

3.5 Connection to Transkribus

The DocScan app has been connected to Transkribus, using the Transkribus rest API. The app gives the user the ability to create or select collections or documents to which the captured images are added. The user can log in to Transkribus via the app and select the documents that should be uploaded to the Transkribus server. The upload process is performed in the background. Thus, the user can take images with DocScan while uploading documents in parallel. Additionally, it is not necessary that the app is opened while the documents are uploaded. If the mobile device is offline the upload starts after the network connection is available again, regardless if the app is running or not.

4 Resources

The Transkribus DocScan App is OpenSource and available at the Google Playstore, as well as in the Transkribus Github repository: `https://github.com/TUWien/DocScan`.

All information regarding the ScanTent are presented at `https://scantent.caa.tuwien.ac.at/en/`.

5 Future Work

In 2017 the material selection for the ScanTent and the workflow for the production has been defined. In 2018 about 20 more prototypes will be realized. It is further planned to test the ScanTents together with the DocScan app in different archives (e.g. Archive Greifswald), and to organize the first Scanathon to show the potentials of crowd scanning together with the ScanTent and DocScan. Based on the feedback (Usability testing), the professional production of about 500-1000 pieces will be planned.

Additionally, the DocScan app will be further developed and improved: in a next step a QR code reader will be integrated into the DocScan app. We are planning to scan QR codes that provide details about documents (such as title or signature). For this purpose QR codes should be added to books / documents by libraries and archives, and the DocScan users should scan such a QR code before capturing the corresponding document. Thus, details about the document are provided to the app and errors by the human operator are prevented, such as mistyping the document title or refering to incorrect stack numbers. This functionality could be of particular use to archives and libraries by allowing them to access copies of the content digitised by DocScan users. The next major release of the DocScan app is planned in mid-2018.

References

[1] D. Vaquero, N. Gelfand, M. Tico, K. Pulli, and M. Turk, "Generalized autofocus," in *2011 IEEE Workshop on Applications of Computer Vision (WACV)*, Jan 2011, pp. 511–518.

[2] F. Kleber, M. Diem, F. Hollaus, and S. Fiel, "Mass digitization of archival documents using mobile phones," *The 4th International Workshop on Historical Document Imaging and Processing (ICDAR 2017)*, 2017.

[3] J. Burie, J. Chazalon, M. Coustaty, S. Eskenazi, M. M. Luqman, M. Mehri, N. Nayef, J. Ogier, S. Prum, and M. Rusiñol, "Icdar2015 competition on smartphone document capture and ocr (smartdoc)," in *13th International Conference on Document Analysis and Recognition (ICDAR)*, Aug 2015, pp. 1161–1165.

[4] S. Pertuz, D. Puig, and M. A. Garcia, "Analysis of focus measure operators for shape-from-focus," *Pattern Recognition*, vol. 46, no. 5, pp. 1415–1432, May 2013. [Online]. Available: http://dx.doi.org/10.1016/j.patcog.2012.11.011

[5] J. Brenner, B. Dew, and J. Horton, "An automated microscope for cytologic research," *J Histochen Cytochem*, no. 24, 1976.

[6] L. Firestone, K. Cook, K. Culp, N. Talsania, and K. Preston, "Comparison of autofocus methods for automated microscopy," *Cytometry*, vol. 12, no. 3, pp. 195–206, 1991. [Online]. Available: http://dx.doi.org/10.1002/cyto.990120302

[7] Z. Zivkovic, "Improved adaptive gaussian mixture model for background subtraction," in *Proceedings of the 17th International Conference on Pattern Recognition (ICPR'04)*, ser. ICPR '04. Washington, DC, USA: IEEE Computer Society, 2004, pp. 28–31. [Online]. Available: http://dx.doi.org/10.1109/ICPR.2004.479

[8] Z. Zivkovic and F. van der Heijden, "Efficient adaptive density estimation per image pixel for the task of background subtraction," *Pattern Recognition Letters*, vol. 27, no. 7, pp. 773–780, May 2006. [Online]. Available: http://dx.doi.org/10.1016/j.patrec.2005.11.005