

Konzept: Projekttitel

EDBV WS 2019/2020: AG_C_3

Anand Eichner (11808244)

Laurenz Edmund Fiala (11807869)

Anna Nieto-Berezhinskaya (01223066)

Aleksandar Vucenovic (01635282)

Jansen Wu (01226578)

28. Oktober 2019

1 Ziel

Das Projekt soll Bücher in einem Bücherregal erkennen, in Bücher-Koordinaten umwandeln und nach ihrem Label abspeichern. [1]

2 Eingabe

JPG-Bild eines Bücherregals mit Büchern, auf denen eindeutige Labels (schwarz-auf-weiß) kleben.

3 Ausgabe

visuell:

TODO: Bild mit gerade-stehenden Büchern.



Abbildung 1: Input-Bild

textuell:

Strukturierte Klartext-Datei mit Inhalt:

- Standort der Bücher (in Büchern zum Ursprung - links-oben)
- Vier Pixel-Vektoren, die ein Buch in einem Viereck umschließen

4 Voraussetzungen und Bedingungen

- Die Bücher müssen gerade ($+/- 5^\circ$) stehen.
- Das Bild darf nicht mehr als 30° von der Waagrechte abweichen.
- Das Bild muss eine für die Texterkennung der Labels ausreichende Auflösung aufweisen (Abhängig von der Entfernung).
- Das Bild muss farbig sein.

Voraussetzungen für Eingabe definieren, anhand dieser werden Datensätze erstellt
Kommentar: Welche Eigenschaften muss der Input erfüllen, damit ihr damit arbeiten könnt (zB.: Kameraeinstellungen, Hintergrundeigenschaften, etc.)?

5 Methodik

Methodik- Pipeline

1. Hough-Transformation
Finden der Regalfächer zum Korrigieren der Perspektive
2. Perspektivenkorrektur
Mittels Transformationsmatrix aus HT berechnet
3. Kantenfilter & Finden von Interest Points
Finden der Ecken von Labels

4. Integral imaging auf IPs
Finden von Labels innerhalb eines Akzeptanzbereichs
5. Eigene Heuristik
Einordnen von Labels in Buch-Koordinaten
6. Optical Character Recognition
Erkennen von Text auf den Labels in den zuvor erkannten Bereichen

Kommentar: Die folgenden Fragen sollten hier bedacht und beantwortet werden: Welche Arbeitsschritte sind notwendig um für den gegebenen Input den entsprechenden Output zu berechnen? Wozu sind die jeweiligen Methoden notwendig – d.h. welche konkrete Methoden wird für diese Arbeitsschritte verwendet?

6 Evaluierung

- **Werden die Regalfächer korrekt erkannt?**

Voraussetzungen:	Seite von Regalfächer, die zur Kamera zeigt, ist schwarz.
Ergebnis:	An jedem Fach liegt eine Gerade an.

- **Wird die Perspektive korrekt angepasst?**

Voraussetzungen:	Korrekte Geraden der Regalfächer.
Ergebnis:	Bücher und Labels sind im Bild weitestgehend rechteckig.

- **Werden alle Labels erkannt?**

Voraussetzungen:	Perspektivenkorrigiertes Bild
Ergebnis:	Es müssen mindestens alle Labels erkannt werden (false-positives sind erwartet da die Filterung mittels OCR passiert)

- **Sind die Bounding Boxes der gefundenen Labels korrekt? Ist der gesamte Text darin enthalten?**

Voraussetzungen:	Korrekt erkanntes Label oder ein false-positive.
Ergebnis:	Vier Vektoren, die den gesamten Text umschließen (und nicht mehr). Bei false-positives ist das Ergebnis nicht relevant, jedoch sollte es nicht zu groß sein (z.B. das gesamte Bild überdecken).

- **Werden die Labels korrekt in Bücher-Koordinaten umgewandelt?**

Voraussetzungen:	Bounding Boxes der Labels sind korrekt.
Ergebnis:	Bücher-Koordinatensystem als 2D-Array mit Ursprung links oben.

- **Werden die Labels der TU-Bibliothek korrekt gelesen und in Text umgewandelt?**

Voraussetzungen:	Label Bounding-Boxes wurden korrekt berechnet (enthalten keinen unnötigen Text).
Ergebnis:	String-Repräsentation des Labels. Erwartete Korrektheit: > 90% für typische Datensätze.

- **Wird die Wahrscheinlichkeit der Label-Korrekttheit angemessen berechnet?**

Voraussetzungen: Korrekt in Text umgewandeltes Label.

Ergebnis: Floating-point Wert im Intervall [0, 1]. Alle Labels mit Wahrscheinlichkeit 0 wurden entfernt.

Kommentar: Eine qualitative Evaluierung basiert auf der subjektiven Wahrnehmung einer Person (Ist ein Ergebnis gut oder schlecht?). Ihr sollte hier aber vor allem auch eine quantitative Evaluierung durchführen, d.h. eine objektive Evaluierung durch Vergleich eurer (Zwischen-)Ergebnisse mit ground truth (ein klassisches Beispiel: Für wie viele der Test-Datensätze wurde für Aufgabe xy ein korrektes Ergebnis erzielt?).

7 Datenbeispiel



Abbildung 2: Input-Bild

8 Zeitplan

Meilenstein	abgeschlossen am	Arbeitsaufwand in h
Prototyp	3.11.	30
Hough-Transformation	15.11.	50
Perspektivenkorrektur	15.11.	6
Labelerkennung	25.11.	80
Labels in Buch-Koordinaten	1.12.	4
Optical Character Recognition	10.12.	110
Labels filtern	15.12.	6
Daten in Output-Format umwandeln	15.12.	4
Tests	18.12.	5
Evaluierung	20.12.	5 [2]

Kommentar: Definiert euch „Meilensteine“. Die vorgegebenen Termine (zB. Zwischenpräsentation) sind hier nicht von Interesse, stellt euch eher die Frage: Wann rechnet ihr mit einem fertigen Prototyp (mit Hilfe von Matlab-Toolboxes)? Wann soll danach ein gewisser Arbeitsschritt (entsprechend eurer Methodik-Pipeline) fertig implementiert sein? Plant auch Zeit für zB. Tests, Evaluierung etc. ein. Gebt weiters pro Arbeitsschritt an, wieviel Arbeitsaufwand (Stunden) eurer Meinung nach zur Umsetzung notwendig sind. Bedenkt, dass es sich bei EDBV um eine Übung im Ausmaß von 3.0 ECTS handelt. Für 1.0 ECTS rechnet man mit 25h Arbeitsaufwand pro Semester. Auf Teil1 (die Gruppenphase von EDBV) entfallen 2.4 dieser 3.0 ECTS und somit 60h Arbeit pro Gruppenmitglied. Wir rechnen daher für Teil 1 mit 300h Arbeit pro Gruppe.

Kommentar: Gebt eine relevante Literaturquelle (Bücher bzw. Kapitel, Konferenz- oder Journal-Papers) pro Gruppenmitglied (im kompilierten Bibtex-Format - Beispiele für Referenzen im Bibtex-Format: <http://verbosus.com/bibtex-style-examples.html?lang=de>). Diese Quellen sollten für euch bei zB. der Implementierung einer Methode, der Wahl von Parametern, etc. helfen. Können aber auch ein ähnliches Problem behandeln und motivieren, warum ihr euch für gewisse Methodik entschieden habt.

Literatur

- [1] Peter Adams. The title of the work. *The name of the journal*, 4(2):201–213, 7 1993.
An optional note.
- [2] David G Lowe. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2):91–110, 2004.