

# Automatisierte Aufbereitung archäologischer Grabungsfotos mittels Computer Vision

Simon Metzger

## **Masterarbeit**

zur Erlangung des akademischen Grades Master of Arts im  
Studiengang Digitale Methodik der Geistes- und  
Kulturwissenschaften

Johannes-Gutenberg-Universität Mainz und Hochschule Mainz

## **Zusammenfassung**

Im Format abstract

# Inhaltsverzeichnis

<b>1 Allgemeines</b>	<b>2</b>
1.1 Unterpunkt . . . . .	2
<b>2 Einleitung</b>	<b>3</b>
2.1 Datenlage Archäologie . . . . .	3
2.2 Grabung Kapitol . . . . .	3
2.3 Datengrundlage und Vorhaben . . . . .	3
2.3.1 Datensatz vorstellen . . . . .	3
<b>3 Tafeldetektierung</b>	<b>4</b>
3.1 Die Tafeln und ihre Tücken . . . . .	4
3.1.1 Die Tafeln . . . . .	4
3.1.2 Tafelvergleiche . . . . .	5
3.2 Detektierungsmöglichkeiten . . . . .	6
3.2.1 CNN . . . . .	6
3.2.2 Ergebnisse CNN . . . . .	6
3.2.3 Ergebnisse Feature Detection . . . . .	8
3.2.4 Ergebnisse Feature Detection . . . . .	8
3.2.5 Contours . . . . .	8
3.2.6 Rectangle Detection . . . . .	13
3.2.7 Hough . . . . .	13
3.2.8 Ergebnisse Kantenerkennung /Contours . . . . .	14
<b>4 Texterkennung</b>	<b>15</b>
4.1 Texterkennung allgemein . . . . .	15
4.2 Das Ausgangsmaterial . . . . .	15
4.3 Tesseract . . . . .	15
4.4 Vorgehen . . . . .	15
<b>5 SIFT</b>	<b>16</b>
5.1 SIFT allgemein . . . . .	16
5.2 Ausgangsmaterial und Probleme . . . . .	16
5.3 Vorgehen . . . . .	16
<b>6 Fazit</b>	<b>17</b>

# **1 Allgemeines**

Vom Allgemeinen

## **1.1 Unterpunkt**

zum Speziellen [1]

## **2 Einleitung**

Einleitung und Fragestellung

### **2.1 Datenlage Archäologie**

viele Grabungen, bei denen keine solide Datenbasis vorhanden ist (Quelle!)  
Digitalisierungsrückstand (Quelle!)

### **2.2 Grabung Kapitol**

Grabungsverlauf bis 2014 (recherchieren)  
Übernahme durch DAI (recherchieren)

### **2.3 Datengrundlage und Vorhaben**

#### **2.3.1 Datensatz vorstellen**

Herkunft

Umfang

Fragestellungen des Projektes

drei Schritte: Tafeln/ Objekte erkennen und ausschneiden, Texterkennung,  
SIFT

## 3 Tafeldetektierung

Das folgende Kapitel befasst sich mit dem ersten Schritt in der automatisierten Analyse der Grabungsfotos: der Erkennung der Schiefertafeln.

Zunächst sollen die Tafeln vorgestellt und die Probleme bei der Detektion erörtert werden. Im Anschluss werden verschiedene Möglichkeiten der Erkennung präsentiert. Schließlich werden mehrere angewandte Methoden erörtert und die erzielten Ergebnisse vorgestellt.

### 3.1 Die Tafeln und ihre Tücken

#### 3.1.1 Die Tafeln

Die Verwendung von Tafeln zur Dokumentation von Fund- und Grabungsarealen ist in allen, im weitesten Sinne grabenden, Wissenschaften weit verbreitet (Vgl. Bildquellen). So setzt auch die Archäologie diese Methode ein. Dabei werden neben den zu dokumentierenden Gebieten verschiedenste Formen von Tafeln oder Schildern platziert, auf denen Zeit und Ort der Aufnahme sowie weitere bild- und motivbezogene Informationen festgehalten werden können. Der Vielfalt von Form und Material der Tafeln ist dabei keine Grenze gesetzt. Bei den Tafeln, die Gegenstand dieses Projektes sind, handelt es sich um Schiefertafeln mit einem Holzrahmen, die mit Kreide beschriftet wurden. Für die Detektion der Tafeln ergeben sich daraus folgende Faktoren:

1. Die Tafeln haben grundsätzlich eine rechteckige Form.
2. Durch die Breite des Rahmens können bis zu zwei Rechtecke erkannt werden, ein Inneres und ein Äußeres.
3. Durch die große Differenz zwischen dem hellen Holzrahmen und der dunklen Schieferplatte sollte der innere Rand in der Regel gut detektierbar sein.

Die im Beispielbild gezeigte Tafel stellt gewissermaßen ein Idealbild dar: Die Tafel nimmt einen relativ großen Teil des Originalbildes – bei der Darstellung hier handelt es sich um einen Ausschnitt – ein. Sie ist frontal vor der Kamera positioniert. Die Beleuchtung ist gut und indirekt. Keines der weiteren Bildelemente verdeckt die Tafel. Diese Beschreibung impliziert schon die Problemfelder, die bei der Detektion beachtet werden müssen:

1. Die Tafel ist unter Umständen stark rotiert.
2. Die Distanz der Tafel zur Kamera und damit ihre Größe im Bild kann stark variieren.



Abbildung 1: Beispiel eines Fotos der verwendeten Tafel. GOT bezeichnet die Kampagne, darunter folgt das Datum. US ist die Abkürzung für *unità stratigrafica*, die stratigrafische Einheit.

3. Der Rahmen der Tafel kann teilweise verdeckt oder anderweitig durch Gegenstände überlagert sein.
4. Die Farbe des Tafelrahmens kann dazu führen, dass sie sich nicht klar vom Hintergrund abhebt, was die Detektion des äußeren Randes erschweren kann.
5. Unregelmäßigkeiten im Rahmen, die auf grobe Verarbeitung oder Abnutzung zurückzuführen sind, können die Detektion erschweren.
6. Die Beleuchtung kann zu Problemen führen. Grundsätzlich sind alle Fotos hell und gut ausgeleuchtet, direktes Licht kann sich aber negativ auf die Kontraste auswirken.
7. Weitere Gegenstände, die den Spezifika der Tafeln entsprechen, können im Bild vorhanden sein.

Teilweise werden die hier genannten Probleme auch bei der Texterkennung wieder relevant. Auf diese und auf weitere wird an geeigneter Stelle zurückgegriffen.

### 3.1.2 Tafelvergleiche

Im Rahmen der Arbeit wurden weitere Tafeln exemplarisch dem Algorithmus unterzogen. Dabei handelte es sich um Aufnahmen der späteren Grabungen des Deutschen Archäologischen Instituts am Kapitol in Rom sowie um vergleichbare Fotos von Bodenuntersuchungen der Gruppe Terrestrische Ökohydrologie der Friedrich-Schiller-Universität Jena. Der ursprüngliche Gedanke dahinter war eine möglichst universale Detektion von Tafeln aller Art anzustreben. Während dieses Vorhaben aus Zeit- und Komplexitätsgründen ohnehin zum Scheitern verurteilt

war, warf das weitere Material die Frage auf, wo die Grenze des technisch möglichen liegt, vor allem mit der hier letztlich gewählten Methodik.

Die Tafeln beider Projekte sollen im Folgenden kurz vorgestellt werden, um das Spektrum der Komplexität evtl. Vergleiche zu Tafeln aus späterer Grabung als Positivbeispiel:

besser gearbeitete Tafeln

besser lesbare Schrift

evtl. Vergleiche zu Tafeln der Bodenkunde als Negativbeispiel:

Tafel schwierig durch Form und Farbe

Klarsichthülle: Reflektion und Formveränderung

oft verdeckt

## 3.2 Detektierungsmöglichkeiten

### 3.2.1 CNN

CNN

Convolutional Neural Networks (CNN) Kurzdefinition

Coco und Coco bzw. Yolo Weights erklären

Code Herkunft erklären (Rücksprache Sellent bzgl. Quelle und Zitation etc.)

### 3.2.2 Ergebnisse CNN

Der Ansatz bei der Arbeit mit CNNs bestand in der Überlegung, dass die Tafeln bestimmten Objekten, wie beispielsweise Bücher oder Müslipackungen, ausreichend ähneln, um als solche erkannt zu werden. Prinzipiell wäre es auch möglich, ein eigenes Modell zu trainieren, dass auf die Erkennung der Tafeln zugeschnitten ist. Dieser Ansatz wurde hier nicht weiter verfolgt, da einerseits die vorliegende Datenmenge von knapp 1500 Bildern gering für ein solches Vorhaben ist und andererseits der Aufwand sehr groß wäre für ein Problem, das sich mit klassischen Methoden der Computer Vision ohne Weiteres lösen lässt.

Entsprechend wurden sowohl der YOLO, als auch der COCO object detector auf den Datensatz angewandt. Die Ergebnisse sind dabei wenig überzeugend.



Abbildung 2: Eine beispielhafte Auswertung mit COCO-Weights: Es werden zwar durchaus Objekte erkannt, die Tafel ist aber nicht darunter. Die Objekte werden nicht korrekt erkannt, was aber bei dieser untypischen Fotografie nicht weiter verwunderlich ist.



Abbildung 3: Auffällig häufig ist die Klassifizierung des Nordungspfeils als Surfboard. Diese ist aber nicht häufig und zuverlässig genug, um COCO zur Erkennung des Pfeils einzusetzen.

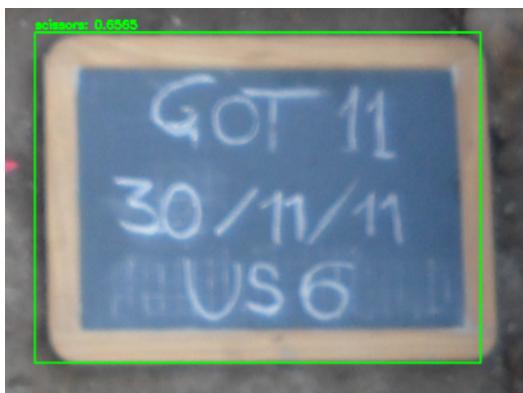


Abbildung 4: Ähnlich wie bei COCO klassifiziert auch YOLO die Tafeln nur auf wenigen Bildern. Dann allerdings als Scheren...

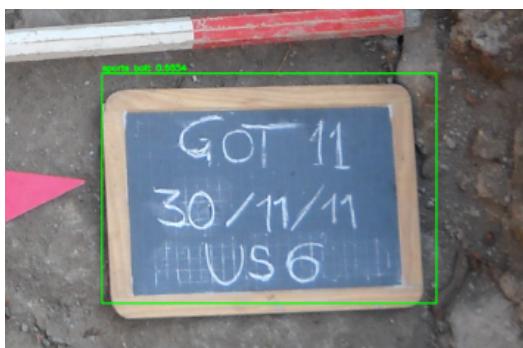


Abbildung 5: ... oder als Sportgerät. Die erhoffte Ähnlichkeit mit beschrifteten, rechteckigen Objekten wie Büchern besteht somit also nicht.

### 3.2.3 Ergebnisse Feature Detection

theoretische Grundlagen Feature Detection

### 3.2.4 Ergebnisse Feature Detection

alten Code raussuchen, aufbereiten und präsentieren

### 3.2.5 Contours

Die Contours-Funktion von OpenCV basiert darauf, dass alle benachbarten Punkte mit gleicher Farbe oder Intensität als Teil einer Kontur betrachtet werden (Quelle OpenCV doc). Der Algorithmus liefert also eine Liste von Punkten, die Grundlage einer Vektorgrafik sind. Der Parameter `cv2.CHAIN_APPROX_SIMPLE` vereinfacht die Kontur, indem redundante Punkte entfernt werden. Als Input für Contours werden binarisierte Bilder empfohlen. Das sind Bilder, die in eine Grauskala umgewandelt und anhand eines Thresholds, eines Grenzwertes, in ein reines Schwarz-Weiß-Bild transformiert wurden. Zur Binarisierung wurden im Verlauf der Entwicklung zwei Verfahren geschrieben, von denen eines de facto nicht mehr in Verwendung ist, der Vollständigkeit halber hier aber aufgeführt werden soll. Das Kernproblem, das zu der parallelen Entstehung zweier Konzepte führte, waren vor allem falsch-positive Detektionen von Tafeln, also Fälle, in denen korrektweise Rechtecke erkannt wurden, die aber keine Tafeln waren.

**Adaptive Ansatz** Der adaptive Ansatz heißt so aufgrund der Verwendung eines adaptiven Thresholds, der auf Bildausschnitten vorher festgelegter Größe gewissermaßen lokale Thresholds festlegt und so gut geeignet ist, um Bilder mit großen Unterschieden in der Helligkeit zu binarisieren, ohne dass wichtige Informationen verloren gehen (Quelle OpenCV doc). Da es bei dem vorliegenden Material, vor allem aufgrund von Schatten und direktem Sonnenlicht, große Helligkeitsunterschiede sowohl zwischen den Bildern als auch innerhalb eines einzelnen Bildes gibt, ist die Arbeit mit einem fixen Threshold schwierig und dieser Ansatz bot sich an. Allerdings entstand dadurch eine relativ große Zahl an falsch-positiven, wie oben zu sehen. So kam es zur Entwicklung des zweitens Ansatzes.

Die Umsetzung des adaptiven Ansatzes sieht wie folgt aus: Als Input für die entsprechende Funktion werden das zu bearbeitende Bild und der dazugehörige Dateiname übergeben, der für die Nachverfolgung und den Output wichtig ist. Das Bild wird in eine Grauskala umgewandelt und in mit der Funktion `scaleImage`



Abbildung 6:  
Falsch-Positive:  
Hier werden korrekt  
erweise Rechtecke  
detektiert, die  
allerdings keine  
Tafeln und so  
mit uninteressant  
für die weitere  
Verarbeitung sind.

auf eine Größe von 1000 Pixeln skaliert. Dieser Schritt erfolgt, um die Detektion durch kleinere Datenmengen zu beschleunigen und um durch die dadurch einheitliche Größe der Bilder präzisere Kriterien für die Rechtecksdetektion formulieren zu können. Das skalierte Bild wird jetzt mittels adaptiven Threshold binarisiert. Es folgt die eigentliche Rechtecksdetektion durch den Aufruf der Funktion `rect_detect`, auf die später genauer eingegangen wird. Die so gewonnenen Konturen und entdeckten Rechtecke werden auf das skalierte Bild übertragen und der `output`-Funktion übergeben. Die Rechtecke selbst, im Code als `rois` (region of interest) bezeichnet, werden auf die ursprüngliche Bildgröße zurückskaliert und in das Hauptprogramm übergeben, damit die Bilder in ihrer höheren Ursprungsauflösung weiter bearbeitet werden können.

```

1 def rect_detect_adaptive(img, fileName):
2
3     #convert to grayscale
4     gray = cv2.cvtColor(img, cv2.COLOR_BGR2GRAY)
5
6     gray = normalizeImage(gray)
7
8     #scale image
9     scaled = scaleImage(gray)
10
11    #set threshold

```

```

12     binary = cv2.adaptiveThreshold(scaled, 255, cv2.
13                                     ADAPTIVE_THRESH_GAUSSIAN_C, cv2.THRESH_BINARY, 11, 1)
14
15     #detect contours and rectangles
16     contours, rois = rect_detect(binary)
17
18     scaled = cv2.cvtColor(scaled, cv2.COLOR_GRAY2BGR)
19
20     #add contours in red to image
21     roisImg = cv2.drawContours(scaled, contours, -1, (0, 0, 230))
22
23     #rescale rois
24     scaledrois = [rescale(gray, rect) for rect in rois]
25
26     #add the found rectangles in green to image
27     roisImg = cv2.drawContours(scaled, [cv2.boxPoints(rect).
28                                astype('int32') for rect in rois], -1, (0, 230, 0), 2)
29
30     #send the modified images in the output function
31     output('output', roisImg, fileName, 'adaptive')
32
33     return(scaledrois)

```

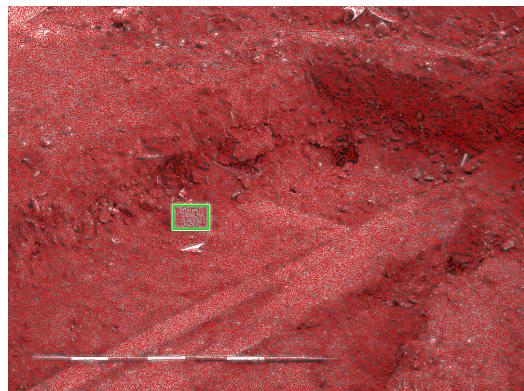


Abbildung 7:  
Detektion mittels adaptiven Ansatz:  
Aus allen gefundenen Konturen (rot) werden die Rechtecke ausgewählt (grün).

**Iterative Ansatz** Der Iterative Ansatz ist danach benannt, dass durch ein Spektrum an Thresholds iteriert wird. In jedem dieser so entstandenen binären Bilder werden anschließend die Contours gesucht und aus diesen wiederum die Rechtecke ausgewählt. Die Idee dahinter war, dass die Falsch-Positiven oft weniger konstant in den Bildern detektiert wurden als die tatsächlichen Tafeln. Würde man also ein Bild mit verschiedenen Thresholds binarisieren sollte das Rechteck, dass in den meisten dieser binären Bilder entdeckt wird, die Tafel sein. In der Implementierung dieses Ansatzes werden von 20 bis 200 in Fünferschritten 37 Thresholds durchlaufen. Die Gleichheit zweier Rechtecke wird dabei mittels

*Intersection over Union* berechnet, deren Ergebnis sich der 1 nähert, je gleicher die Rechtecke in Position und Größe sind. Im Detail sieht das wie folgt aus: Aus den globalen Flags wird ein Wert für den minimalen Threshold übernommen. Wie bereits in dem adaptiven Verfahren wird das Bild in eine Grauskala umgewandelt. Beginnend beim minimalen Threshold (default = 20) werden die Bilder binarisiert und mit `rect_detect` auf Rechtecke untersucht. Diese werden in der Liste `allRois` gespeichert und der Threshold um 5 erhöht, bis der Wert von 200 erreicht ist. Im Anschluss werden die Rechtecke in eine Liste von Dictionaries überführt. Das hat den Hintergrund, dass sich hier leicht und übersichtlich ein Keyword einfügen lässt, mit dem die Zahl ähnlicher Rechtecke gemessen werden kann. In zwei Schleifen wird diese Liste durchlaufen und so je zwei Rechtecke mittels `intersection_over_union` miteinander verglichen. Wird ein Wert von über 0.9, also eine hohe Übereinstimmung, erzielt, wird das Keyword `same` um eins erhöht. Schließlich wird das Rechteck mit dem höchsten `same`-Wert als ein Tafelfund betrachtet. Das weitere Verfahren ist wie beim adaptiven Ansatz: Das gefundene Rechteck wird auf das Bild übertragen und ausgegeben und dann ins Hauptprogramm zurückgegeben.

```

1 def rect_detect_iterative(img, fileName):
2
3     thresh = THRESHOLD_MIN
4     allRois = []
5
6     #convert to grayscale and normalize
7     gray = cv2.cvtColor(img, cv2.COLOR_BGR2GRAY)
8     gray = normalizeImage(gray)
9
10    #search for rectangles with increasing threshold, max 200
11    while thresh <= 200:
12
13        rois = []
14        contours = []
15
16        ret, binary = cv2.threshold(gray, thresh, THRESHOLD_MAX,
17                                     cv2.THRESH_BINARY)
18
19        contours, rois = rect_detect(binary)
20
21        if len(rois) > 0:
22
23            allRois.append(rois)
24
25            thresh += 5
26
27    #new rois list
28    rois_list = []

```

```

28 #go through the found rectangles and add them to an array of
29 #dictionaries
30 for r in allRois:
31
32     for i in range(len(r)):
33         (x,y), (w,h), angle = r[i]
34         rois_dict = {
35             }
36         rois_dict["x"] = x
37         rois_dict["y"] = y
38         rois_dict["w"] = w
39         rois_dict["h"] = h
40         rois_dict["angle"] = angle
41         rois_dict["same"] = 0
42
43         rois_list.append(rois_dict)
44
45 #find and count rectangles in the same area
46 for i in range (len(rois_list)):
47     for j in range (len(rois_list)):
48         recta = rois_list[i]["x"], rois_list[i]["y"],
49         rois_list[i]["w"],rois_list[i]["h"]
50         rectb = rois_list[j]["x"], rois_list[j]["y"],
51         rois_list[j]["w"],rois_list[j]["h"]
52         if intersection_over_union(recta, rectb) > 0.9:
53             rois_list[i]["same"] = rois_list[i]["same"] + 1
54             rois_list[j]["same"] = rois_list[j]["same"] + 1
55
56 #new rectangle list
57 rects = []
58 #same is the number of same rois in the area, 0 is default
59 same = 0
60
61 #if there are dictionaries in the list, search for the one
62 #with the highest number of same rectangles in the area
63 if len(rois_list) > 0:
64     roi = rois_list[0]
65     for i in range (len(rois_list)):
66         if rois_list[i]["same"] >= roi["same"]:
67             roi = rois_list[i]
68
69 #add contours in red to image
70 if roi["same"] >= 6:
71     #roisImg = cv2.drawContours(gray, contours, -1, (0,
72     0, 230))
73     rect = (roi["x"],roi["y"]),(roi["w"],roi["h"]),roi[
74     "angle"]
75     rects.append(rect)
76     same = roi["same"]
77
78 #convert to colored img for output
79 gray = cv2.cvtColor(gray, cv2.COLOR_GRAY2BGR)

```

```

71 #add contours in red to image
72 roisImg = cv2.drawContours(gray, contours, -1, (0, 0, 230))
73 #add the found rectangles in green to image
74 roisImg = cv2.drawContours(gray, [cv2.boxPoints(rect).astype
75 ('int32') for rect in rects], -1, (0, 230, 0),3)
76
77 #send the modified images in the output function
78 output('output', roisImg, fileName, str(same))
79
80 return(rects)

```

Tatsächlich können mit diesem Verfahren die Zahl der Falsch-Positiven reduziert werden, ohne das Problem allerdings ganz zu lösen. Generell lässt sich sagen, dass je öfter ein Rechteck in einer der Iterationen erkannt wird, desto größer ist die Wahrscheinlichkeit, dass es sich tatsächlich um eine Tafel handelt. Ab einem Wert von 20 liegt die Quote bei 100%. Es gibt aber auch korrekt erkannte Tafeln, die nur in 6 der Iterationen detektiert werden und umgekehrt Falsch-Positive, die in bis zu 14 Iterationen vorkommen. Zwar handelt es sich hier jeweils um Ausreißer, das Problem bleibt jedoch bestehen. Empirisch hat sich gezeigt, dass alles unter 6 Iterationen mit Sicherheit keine Tafel ist und somit aussortiert werden kann. Wenige Ausreißer verhindern, dass diese Grenze nach oben gesetzt werden kann. Auch dieser Ansatz führte also nicht zum gewünschten Ergebnis.



Abbildung 8:  
Falsch-Positive  
beim Iterativen An-  
satz. Hier wurden  
in 14 Iterationen  
die Holzbret-  
ter als Rechteck  
identifiziert.

### 3.2.6 Rectangle Detection

### 3.2.7 Hough

Später schreiben, je nachdem ob sich der Algorithmus verbessern lässt und präzisere Ergebnisse zu erzielen als simple crop oder nicht. In letzterem Falle Ansatz und Fehlschlag vorstellen.



Abbildung 9:  
Direkte Sonnen-einstrahlung macht die Erkennung schwierig, vor allem, da sich der Rahmen nicht mehr stark vom Schiefer abhebt. Nur in 6 Iterationen wurde diese Tafel erkannt.

### 3.2.8 Ergebnisse Kantenerkennung /Contours

Detaillierte Beschreibung des Vorgehens  
Vor- und Nachteile aufzählen  
konkrete Probleme benennen

## 4 Texterkennung

### 4.1 Texterkennung allgemein

Tesseract: was kann es, wie funktioniert es

### 4.2 Das Ausgangsmaterial

Probleme benennen wie: verwischte Kreide, das Karomuster der Tafeln, Handschrift per se, Licht und Beleuchtung

Funktion der Tafeln und Beschreibung  
rechteckig -& klare Umrisse  
doppelter Rand  
grobe Verarbeitung  
unterschiedliche Positionierung (Entfernung, Winkel)  
wechselnde Beschriftung  
wechselnde Beleuchtung  
rechteckiges Muster im Hintergrund

Problematiken aufzeigen: unterschiedliche Größe, Licht und Beleuchtung, Winkel, Verdeckung  
andere Objekte die als Tafeln erkannt werden könnten

### 4.3 Tesseract

evtl ALternativen

### 4.4 Vorgehen

Evaluation: Vorgehen, Überlegungen

Preprocessing: besondere Herausforderungen, vorgehen, beide Varianten vorstellen

normales Modell

eigenes Modell

Vergleich: Tafeln aus späterer Grabung (gesetzte Lettern)

evtl. Vergleich Tafeln Bodenkunde

## **5 SIFT**

### **5.1 SIFT allgemein**

### **5.2 Ausgangsmaterial und Probleme**

### **5.3 Vorgehen**

Ausschneiden der anderen beweglichen Elemente (Zollstock, Nordungspfeil)

## **6 Fazit**

## **Literatur**

- [1] Gerald Teschl and Susanne Teschl. *Mathematik für Informatiker*, volume 2. Springer, 2014.