Retrieving Cuneiform Structures in a Segmentation-free Word Spotting Framework

Leonard Rothacker
Department of Computer Science XII
TU Dortmund, Germany
leonard.rothacker@udo.edu

Denis Fisseler
Department of Computer Science VII
TU Dortmund, Germany
denis.fisseler@udo.edu

Gerfrid G.W. Müller
Department of Ancient Cultures
Ancient Near Eastern Studies
University of Würzburg, Germany
gerfrid.mueller@uni-wuerzburg.de

Frank Weichert
Department of Computer Science VII
TU Dortmund, Germany
frank.weichert@udo.edu

Gernot A. Fink
Department of Computer Science XII
TU Dortmund, Germany
gernot.fink@udo.edu

ABSTRACT

Cuneiform tablets are an invaluable documentation of early human history. Efforts are being made in digitizing large tablet collections for preserving their content and making them available to a global research community. However, there are hardly any automated computer aided methods for supporting philologists in their analysis. In this paper we present an approach for automatically retrieving cuneiform wedge constellations from digitized cuneiform tablet collections. Compelling results could be achieved in our qualitative and quantitative evaluation on a challenging benchmark consisting of 3D-scanned cuneiform tablets.

CCS Concepts

•**Applied computing** → **Digital libraries and archives; Document analysis;** *Document searching;*

Keywords

word spotting, cuneiform analysis

1. INTRODUCTION

Cuneiform script is the oldest writing system besides Egyptian hieroglyphs and was developed in the second half of the 4th millennium BC, most likely by the Sumerians in southern Mesopotamia. It was in use over more than three millennia until it was finally replaced by alphabet-based writing systems which were easier to learn. Cuneiform texts were written on all kinds of materials but preferably on tablets made of clay which was readily available almost everywhere. Traditionally a stylus from reed served as writing tool which was impressed into moist clay (cf. [5]). In retrospect a big advantage of clay tablets over papyrus and other perishable writing materials was that it can survive over millennia once it comes to rest under the surface and is thereby protected from erosive effects. As a result of their durability, cuneiform tablets are the most important source for our knowledge of the earlier political and intellectual history of mankind. Far more than 500.000 cuneiform documents have been unearthed so far, most of which are still waiting to be studied by philologists. This clarifies the need for efficient methods to search and process the huge amount of cuneiform manuscripts, which until now has been mainly done by manual examination and comparison of the artifacts.

In order to understand the special challenges in processing cuneiform script, basic knowledge about its visual and philological properties are required. Unlike Latin script, cuneiform writing started as pictographic script. Over time the sign inventory was reduced from more than 1500 signs in the early stage to about 600 signs and the shapes became more and more abstract. This resulted in a simultaneous reduction of the visual core components to only vertical, horizontal, diagonal and so-called *winkelhaken* wedge types. The individual wedges resemble tetrahedron shaped holes in the clay surface which introduces a high degree of local self similarity on wedge level and cuneiform sign level. Therefore, small parts of cuneiform signs often can not be identified correctly without knowledge of their geometric context.

Additional challenges of processing cuneiform script can be found on the semantic level. To mention only a few aspects, cuneiform signs are polyvalent and can usually be read in several ways and used in various functions. They have a logographic value, sometimes denote various words and possess one or more syllabic values (cf. [1]). Furthermore signs and groups of wedges can be combined to express a new meaning or incorporated into other signs as elements. As in other writing systems, an individual grapheme may be represented in a wide variety of distinct ways, all represent-

HIP '15 August 22, 2015, Nancy, France

ISBN 978-1-4503-3602-4/15/08…$15.00

DOI: http://dx.doi.org/10.1145/2809544.2809562

ing the "same" grapheme (allographs of a grapheme). The allograph choice may be influenced by the writing instrument, the style of the scribe, neighboring graphemes in the text, the writing speed, the intended target audience, and even unconscious features of individual handwriting, all of which can not be easily handled by established automated transcription methods (cf. [3]).

The script properties described above induce philological text analysis methods that rely on searching occurrences of groups of wedges in cuneiform manuscripts as a starting point. Beyond that, the recent availability of 3D-scans allows a quantitative study of individual writing for the first time. This brings up the need for automatically retrieving particular cuneiform structures like signs, sign groups and words, including assessing and comparing them. Being able to perform a fast and accurate search for groups of wedges in a large number of different cuneiform manuscripts is, therefore, an essential desideratum in philological research.

For over a century, cuneiform tablets were published almost exclusively as autographies. These hand-copies present a readable text in a printable form. The production of photographic images for the ten-thousands of excavated tablets were unaffordable. With the turn of the millennium, digital imaging and the internet made an increasing number of pictures easily accessible. Large collections of cuneiform 2D image databases include the Hethitologie Portal Mainz [14] and the British Museum Collection [6]. Hitherto these pictures are in the best case linked to some descriptive data or a transliteration and translation of the text and may be consulted with it for collation. The digital availability of the documents opens up new paths for computer assisted analysis, especially in the field of pattern recognition. Yet, promising solutions for optical character recognition (OCR) on cuneiform manuscripts are still unavailable.

The main contribution of this paper is an approach for searching cuneiform structures in collections of cuneiform tablets. The search is implemented as a special image retrieval problem that is referred to as query-by-example word spotting (cf. [10]). The philologist selects an exemplary occurrence of a group of wedges in a cuneiform tablet image as query. With our approach it is possible to retrieve occurrences of similar wedge groups in a collection of cuneiform images. The result can be visualized by highlighting detections and listing them ranked according to similarity to the query. This poses two major challenges: As traditional photographic 2D representations of cuneiform script exhibit large variations regarding the shot modalities, they are usually not suitable for direct application of comparative image retrieval methods on a large number of data sets. Therefore, we convert 3D-scans of cuneiform tablet fragments into a more suitable 2D image representation. Secondly, the variability in the visual appearance of cuneiform script is extremely high, especially across tablets, due to different writing styles and writing materials. This is addressed by modeling cuneiform queries with BoF-HMMs (Bag-of-Features Hidden Markov Models). These statistical sequence models have successfully been applied for spotting different scripts like Roman or Bangla [18], and are very flexible and adjustable with respect to their modeling capabilities [20].

The remainder of the paper is organized as follows. Section 2 reviews related work with respect to word spotting and computer aided cuneiform tablet analysis. In section 3 we describe how BoF-HMMs are applied to cuneiform structure retrieval in rasterized 2D cuneiform images. How this method is applied in a typical philological use case is presented in section 4. In section 5 we show qualitative and quantitative results on a dataset of cuneiform tablets. Finally, a conclusion is given in section 6.

2. RELATED WORK

Word spotting methods are very popular for the analysis of historical documents. This is due to their ad-hoc applicability. The use of a recognizer in order to perform a full transcription is usually impossible because of special characteristics of historic documents. Often their script is very particular which requires annotated training material from the same document collection. Additionally, document images suffer from degenerations due to storage, fading ink or ink bleed through, requiring specialized preprocessing methods. In contrast, query-by-example word spotting methods assume relatively uniform appearance of the script and directly match an exemplary instance of a query word with document image regions. These regions are ranked according to similarity with the query and are presented to the user. Word spotting is therefore robust with respect to recognition errors. As long as the relevant detections are among the top results, the method still aids the user.

A prominent method for word spotting in historic document collections was presented by Rath et al. [17]. They relied on a given segmentation of document images into words. These word images were encoded with specialized features, like projection profiles and ink background transitions. Similarity between their feature representations was computed with Dynamic Time Warping. Recently, methods in word spotting became influenced by advances in Computer Vision. These techniques are typically able to be adapted automatically to the problem domain. One of the first methods in this regard was presented in [22]. Rusiñol et al. used BoF (Bag-of-Features) representations for segmentation-free word spotting. BoF represent local document image regions as orderless collections of image features. This approach has mainly two advantages. BoF are estimated from sample data in an unsupervised manner. Thus, no manual effort is required for designing features. As the scripts' appearance varies largely across historic document collections this is a useful property. Furthermore, BoF representations can be applied in a segmentation-free manner as no prior assumptions about the spatial location of script is required. Especially, with respect to analysing historic documents this is desirable. Heuristic segmentations on word level are often hard to obtain because of dense writing, inhomogeneous spacings and document degenerations (cf. [23]).

In our previous work (cf. [20]), we presented an integration of statistical sequence models and BoF, so-called BoF-HMMs, for segmentation-free word spotting in historic documents. HMMs serve for modeling the horizontal variabilities of script, while BoF representations make the matching process robust with respect to vertical displacements of the script in our segmentation-free framework.

Digital extraction and processing of cuneiform script is covered by only few recent publications which are focused almost entirely on the processing of digitized 3D cuneiform datasets. In [12] Mara et al. describe a method to extract a 2D spline representation of 3D cuneiform characters using integral invariant filtering and a skeletonization approach. The integral invariants are used to compute robust curvature metrics and to detect relevant features on the outline of local connected wedge components for the construction of skeleton branches. However, the method requires a perfect mesh which implies that 2-manifoldness has to be ensured through various mesh repair methods in advance. In [3] Bogacz et al. convert the spline representation from [12] into a graph representation and extend it with triangulation information to evaluate a graph similarity based metric for cuneiform characters. Bogacz et al. conclude that the structural information of cuneiform characters can be captured well using a graph representation and claim that conventional OCR methods are not suitable for handling cuneiform script.

As we demonstrated in [7] a model based cuneiform wedge extraction approach using a modified watershed algorithm for wedge separation yields robust extraction results even on incompletely scanned cuneiform tablets and non 2-manifold meshes without the prerequisite of mesh repair operations. We used the extracted wedges and wedge components from [7] successfully in [8] and [4] to analyze statistical low-level script properties. These were employed for identifying similar script types originating from different scribal traditions in order to join candidate identification of cuneiform tablet fragments. The statistical wedge features were evaluated on a non-semantic level as semantics in cuneiform script require the analysis of wedge constellations. In [15] we evaluated a spatial database-centered approach for statistical analysis of syntactic and semantic cuneiform features by extending the data from [8] with local spatial connectivity data. Thus, specific wedge constellations could be retrieved from preprocessed 3D data sets stored in a spatial database.

Unlike our segmentation-free method presented here, approaches in [3] and [15] require a reliable prior segmentation of individual cuneiform wedges. This is difficult to compute due to the properties of cuneiform tablets described above.

3. METHOD

Automatically searching for cuneiform structures is a challenging task. Suitable digital cuneiform tablet representations are hard to obtain, the writing variability is large and cuneiform structures are written closely next to each other, thus no reliable a-priori segmentation can be obtained. Furthermore, cuneiform characters consist of few reappearing wedge types what makes them hard to discriminate. Therefore, we propose a method that approaches these challenges on multiple levels. In order to avoid unwanted variabilities found in photographic reproductions of cuneiform tablets, we obtain our image representations from 3D scans (section 3.1). Images are then represented in terms of quantized local features that can be combined to BoF histograms (section 3.2). Cuneiform queries can then be modeled with BoF-HMMs. These allow for a dynamic probabilistic modeling of the cuneiforms spatial sequential structure and have already been successfully applied for modeling Roman, Bangla and Arabic script [20, 18, 21] in word spotting and handwriting recognition (section 3.3). Using this model the query is spotted in cuneiform tablet images in a patch-based segmentation-free framework (section 3.4).

3.1 Image generation

As discussed in section 1, traditional photographic 2D representations of cuneiform script are usually not suitable for direct application of comparative pattern recognition approaches on a large number of data sets. This is not only

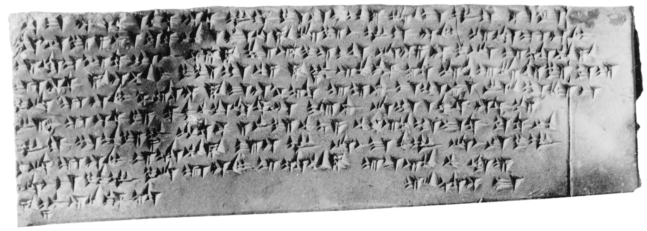


Figure 1: Representative sample section of a traditional photographic reproduction of the cuneiform fragment 44/a, with multiple disadvantageous shot modalities. (Source: Hethitologie Portal Mainz [14])

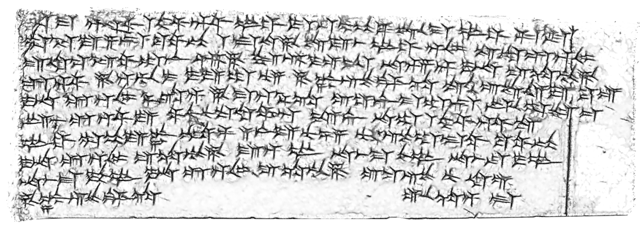
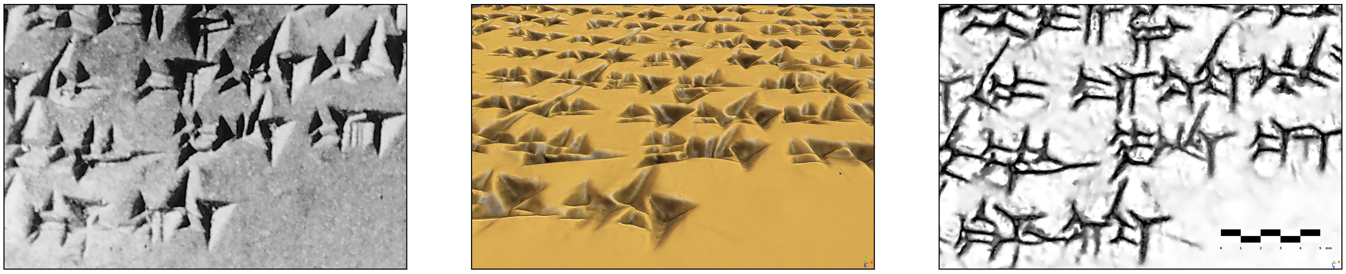


Figure 2: Unified 2D representation generated from a 3D scan of the cuneiform fragment 44/a.

due to large variations in the shot modalities between individual images, but also due to large variations in each individual shot. Figure 1 shows a typical example of a traditional photographic documentation that exhibits multiple disadvantageous shot modalities. Problematic image properties resulting from inappropriate lighting conditions may include low overall contrast, low contrast in wedge shadows, inappropriate exposures and large variations in local lighting conditions. On the other hand, there may also be material induced discolorations on the tablet surface and projective distortions. In order to address this, we use high resolution 3D-scans of cuneiform tablet fragments taken with a Breuckmann optoTOP structured light scanner (AICON 3D Systems, http://www.aicon3d.de) with a maximum resolution of $30\mu m$ to generate unified 2D representations suitable for processing by our word spotting system. We ensure the 2D representations contain a sufficient amount of data to be able to identify cuneiform signs and constellations by targeting a similar appearance like the Y-shaped stylized wedge representations used in traditional cuneiform transcriptions (see [9] for details) like shown in figure 6. As these Y-shaped wedge representations are characterized by the concave three inner edges of a wedge, we optically enhance these features using a differential operator from [13] to approximate the maximum curvature on the triangle meshes of the digitized tablet fragments. The resulting curvature values are morphologically filtered to remove noise and the effects of geometric artifacts and then color mapped to produce a high contrast, sketch like appearance.

The final rasterized 2D output image for processing by the word spotting system is then generated by constructing an orthographic projection matrix from an approximated normal of the cuneiform tablet surface using Principle Component Analysis and a global estimate of the writing direction. This ensures, that the 2D representation of the cuneiform text contains minimal projective distortions and the writ-



(a) (b) (c)

Figure 3: (a) Section from a traditional photographic reproduction of a cuneiform fragment (Source: Hethitologie Portal Mainz [14]), (b) corresponding 3D-scan and (c) computed stylized 2D representation.

ing direction is aligned horizontally as expected by the word spotting system. As 3D-scans are usually saved using real world units, maintaining a consistent scale for all images requires no additional efforts. At this point it is essential to account for holes of various sizes in the mesh which may result from the scanning process either by coloring the background white or tag it with a transparency attribute. The word spotting method may yield wrong recognition results, otherwise. Figure 2 shows an example of a respectively processed cuneiform tablet fragment matching the photographic reproduction in figure 1. As can be seen, the parts of the tablet containing cuneiform writing now exhibit a uniform and yet distinctive rendering of the cuneiform wedges. A detailed side by side comparison of a photographic reproduction, a corresponding 3D-scan and the generated representation at wedge scale is shown in figure 3. This emphasizes the beneficial properties of our 2D representation compared to the original photographic reproduction.

3.2 Image representation

The 2D image representation of the cuneiform tablet from section 3.1 has to be converted to quantized local image features that can be used within BoF-HMMs for word spotting. In order to accomplish this, we extract a dense grid of local image descriptors. This is done using SIFT [11] descriptors that are capturing wedge orientations and their local arrangements. The descriptor size is manually adapted to the typical cuneiform character height and the dense grid ensures that all image regions are represented uniformly and no relevant areas are accidentally omitted. Figure 4a exemplarily shows a dense grid of SIFT descriptors that are indicated by blue patches. It has to be noted that, unlike shown in figure 4a, descriptors exhibit a high amount of overlap in practice.

For obtaining quantized local image features, the descriptors are associated with their most similar representative from a visual vocabulary, containing 4096 visual words. This vocabulary is estimated in an unsupervised prior training step. Figure 4b shows an example of these quantized image features visualized as colored points. Even in the coarse grid depicted here, a correlation between similar color patterns and similar cuneiform structures can be seen. For this specific application to cuneiform script we use descriptors of size 32×32 pixels in a dense grid of 2×2 pixels.

3.3 Cuneiform query modeling

In our query-by-example scenario, queries are specified by tight bounding boxes containing cuneiform structures. Based on these regions, the quantized local image features are modeled with a BoF-HMM. For this purpose a sequence of BoF representations is obtained in a sliding window manner. At each window position the respective feature representatives are accumulated in a BoF histogram. By using the Baum-Welch algorithm (cf. [16]) the HMM is estimated such that the likelihood of generating the sequence is maximal. It is worth noting, that only a single instance of the query cuneiform structure is required in this regard. In the estimation step, a probability for observing a local feature representative within each HMM state is obtained. This probability relates to observing the respective features in the corresponding section of the query. An example of the query modeling process is given in figure 4c where the query image (c1) is represented by the grid of quantized features (c2). Like in figure 4b the quantized features are again shown as colored points. Below, the estimated BoF-HMM is shown (c3). In this example, the *blue* feature has a rather high occurrence frequency in the beginning and in the middle of the query where, for example, the *yellow* feature is rather frequent in the end. This is also reflected in the feature probabilities in the HMM states (c3). For this reason the number of states is important for the specificity of the model. Since few types of cuneiform wedges appear in different cuneiform structures, mostly their order is the discriminating property. This is encoded using a relatively high number of states with respect to the length of the sequence of BoF representations, when compared to the numbers of states that have been chosen for other scripts, like Roman or Bangla (cf. [19, 18]).

3.4 Cuneiform query spotting

For spotting a query in the cuneiform tablet image, we compute a similarity measure between its model and image regions in a segmentation-free approach. As shown in figure 4d a patch is slid through the image. For each patch position a sequence of BoF representations is computed from the underlying grid of feature representatives. The likelihood of generating a sequence with the query BoF-HMM is obtained through probabilistic inference with the Viterbi algorithm (cf. [16]).

The final result of the word spotting system is obtained by extracting regions-of-interest for patches with the highest scores in their respective local neighborhood. These regions are ranked by similarity and presented to the user. An exemplary result is visualized in figure 5. Similarity scores are shown in a heat map next to the retrieval list of ranked regions-of-interest.

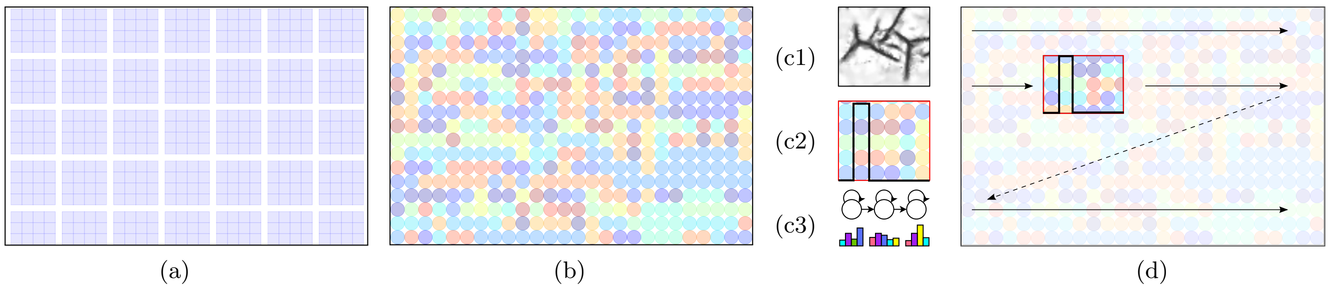


Figure 4: (a) grid of SIFT-Operators, (b) corresponding quantized image features, (c1) search query, (c2) BoF sequence, (c3) BoF-HMM and (d) patch-based model decoding process.

4. SOFTWARE DEMONSTRATOR

In this section we present the visual front-end of our segmentation-free word spotting software demonstrator. This word spotting front-end allows for a seamless integration of our method with a typical philological use case, where text passages of similar content in different manuscripts are compared. For this purpose we show the front-end in figure 5. It consists of three main parts. The left side contains a list based cuneiform fragment explorer that shows small size previews of available cuneiform fragments in the database for easy navigation. The fragments can be selected for simple exploration or visualization of potential query results. The center area of the front-end contains a zoomable and moveable visualization of the currently selected cuneiform fragment with three possible visualization modes. Figure 5 shows these visualization modes next to each other from left to right, framed in green, red and blue. The first mode on the left side displays the generated 2D representation of the plain fragment. This view can be used for fragment exploration and for selecting queries. The second view shows detected regions-of-interest colored by similarity to the query, where blue corresponds to least and red to most similar. The last mode on the right side displays an intuitive heat map visualizing the similarity scores for a query on the entire fragment, with blue to red colors where blue corresponds to least and red to most similar patches, again. Aside from fragment visualization the center area of the front-end can also be used for specifying search queries. In order to define a query, a cuneiform structure can be selected by drawing a tight bounding-box around it. Finally, the right hand side of the front-end contains a compact list based visualization of extracted individual query results. It contains thumbnails of all detected regions for a query across all fragments in the database. While this list allows a quick and easy visual inspection of the query results by a philologist, the small color bar on the left hand side of the thumbnails refers to the similarity to the query. This corresponds to the colors shown in the regions-of-interest visualization on the fragment. The entries of the retrieval list can also be used to directly navigate to the respective locations on the cuneiform fragments for further examination of their context.

5. EVALUATION

In this section we will evaluate our approach based on a sample set of cuneiform fragments and cuneiform sign queries. First, we will define a philologically motivated benchmark in order to evaluate the performance of the approach in section 5.1. Subsequently, in section 5.2 the benchmark results are discussed quantitatively and qualitatively and are put into the philological context.

5.1 Benchmark

In order to evaluate our word spotting method's performance when searching for cuneiform structures, we consider a sample set of cuneiform tablet fragments matching a typical philological use case. Based on the 2D rasterized image representations we created ground truth annotations on the fragments that can be used as a benchmark. The ground truth consists of bounding boxes of selected relevant cuneiform wedge constellations. This is combined with a benchmark protocol that is already established for segmentation-free word spotting and has first been presented in [22]. Afterwards, it has been applied, for example in [19, 18, 23]. In this protocol every ground truth annotation is used as a query and the method's ability to retrieve relevant occurrences of the query is measured.

The cuneiform tablets examined in this paper originate from the ancient city of Hattusa, capital of the Hittites, dating to the second half of the second mill. BC. About 30.000 fragments of tablets are documented in almost 70.000 images and two thousand 3D-scans. This vast amount of data is so far not searchable. However, the text on the tablets has to be examined for collation. As big tablets often have to be recorded on multiple images or 3D-scans, the spotting of sought text passages gets even more time consuming.

For the benchmark we use 3D-scans of 11 cuneiform tablet fragments from the data pool described above. Close to a typical philological use case, the fragments were chosen to cover a variety of different writing styles from different scribes, but also to each contain a sufficient amount of text. As we were interested in how the performance is affected by the visual quality of the input images, we also payed attention to include samples with damage or noise related deficiencies in tablet surface quality. In order to compile a suitable set of queries, we chose 4 wedge constellations of different complexity that represent the real hittite signs *an*, *na*, *ti* and *zi* as shown in figure 6. The signs consist of

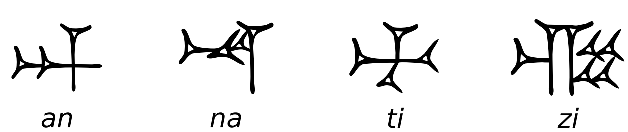


Figure 6: Stylized drawings of the hittite signs *an*, *na*, *ti* and *zi* rendered using the hittite unicode font Ullikummi_A.

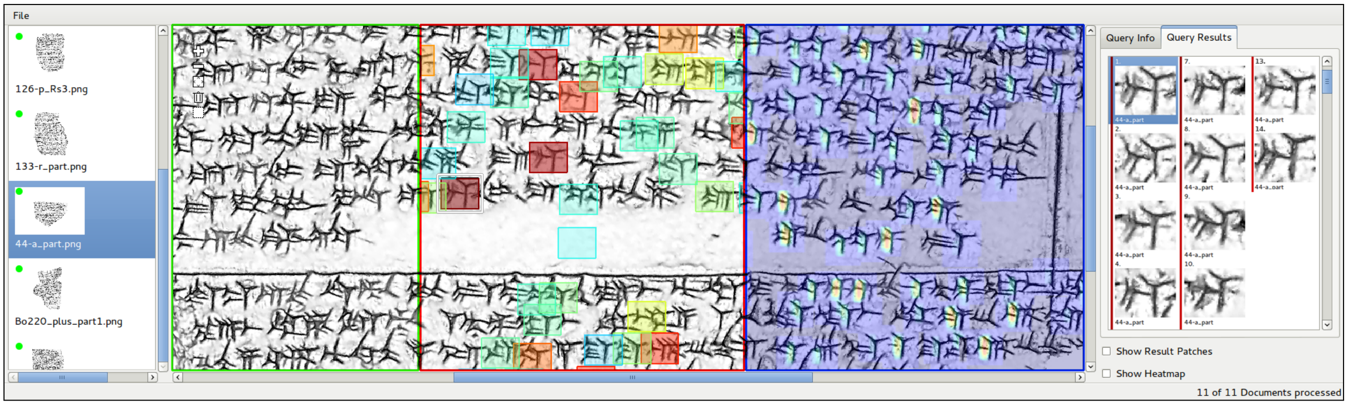


Figure 5: Software Demonstrator. (Left: fragment list, Green: fragment view, Red: retrieval regions-of-interest, Blue: retrieval heat map, Right: retrieval list)

3(an), 4(na, ti) and 7(zi) wedges and share a high frequency of occurrence as well as a comparatively high distinctiveness, the last of which is important to minimize occurrences of the queries as parts of other signs. The complete sample set of 11 tablet fragments contains 500 query instances.

5.2 Results

In the considered philological use case, a philologist defines a query on a cuneiform tablet fragment in order to retrieve all similar regions on the same tablet fragment or on all tablet fragments in the database. As mentioned in section 3.4, we visualize the retrieval results in form of a list of tablet image regions that are ranked according to their similarity with the query (cf. figure 5). In this context it is important for the philologist that the most relevant regions are listed first and that the list contains all relevant instances. Both criteria can be evaluated by computing the average precision that refers to the area below the precision-recall curve. While the precision describes the fraction of retrieved results that are relevant to the query, the recall reflects the fraction of relevant retrieved results and relevant results that can be retrieved for the query at all. Since the precision-recall curve expresses the precision at different recall levels, the area under the curve reflects how well the list is sorted. For a perfect retrieval result, listing all relevant detections first, the precision is maximal for all recall levels. In order to express the second criterion individually, we examine recall separately from average precision. The overall results are obtained from the scores of all queries and are reported as mean average precision (mAP) and mean recall (mR), cf. [2]. Regarding the segmentation-free scenario considered, it is essential to define a notion of relevance. Therefore, we consider a detection as relevant if it overlaps with an occurrence of the query in the ground truth by 50% or more. It has to be noted that the query itself is also retrieved and is typically the first result in the retrieval list.

Table 1: Quantitative results

Tablets	Fragments	Queries	mAP(%)	mR(%)
all tablets ○	11	500	41.6	65.9
avg. per tablet △			79.9	94.3
126-p	3	126	85.2	95.1
133-r	1	36	91.6	99.6
44-a	1	50	78.2	98.6
Bo220_plus	2	73	73.0	92.5
Bo2515	1	12	78.0	97.2
Bo604	1	64	60.8	83.2
Bo49	2	139	85.4	96.6

In the remainder of this section we will discuss the specific results obtained. First we present our quantitative evaluation followed by a presentation of qualitative results. The qualitative analysis gives an intuitive impression of typical difficulties and achievements encountered in practice while the quantitative analysis allows for an interpretation of results over all queries. For a more detailed understanding of the word spotting method's parameters, we refer to our previous work, especially [19] and [20].

Results of the quantitative evaluation are shown in table 1. One major challenge in spotting cuneiform structures is the high variability resulting from different writing styles and degenerations of the tablets. As these properties are expected to vary mainly between tablets, our evaluation contains results from performing queries on all tablet fragments as well as results obtained only from single tablets. The results show clearly that the variability of cuneiform structures within a single tablet is typically low, while a significant negative influence in the evaluation over all tablets can be observed.

The first result in table 1 shows the evaluation on all tablets. Due to the high variability, the mean recall is relatively low. Instances of the same query that differ largely in appearance and size will not be retrieved at all in the segmentation-free query-by-example scenario. In addition to that, also the mean average precision is affected. With increasing variability it becomes more and more difficult to rank relevant instances of a query first. This is confirmed in the per tablet evaluation. The average performance, weighted according to the number of queries per tablet, improves considerably in mean average precision and mean recall, where the mean average precision is almost two times as high. The noticeably lower values for the fragment Bo604 originate from a high level of physical damage on the tablet surface and an increased variability in writing style.

The interpolated precision-recall curve (cf. [2]) for the evaluation over all tablets and per tablet is shown in figure 7. Precision is obtained at 11 recall levels by maximum in-

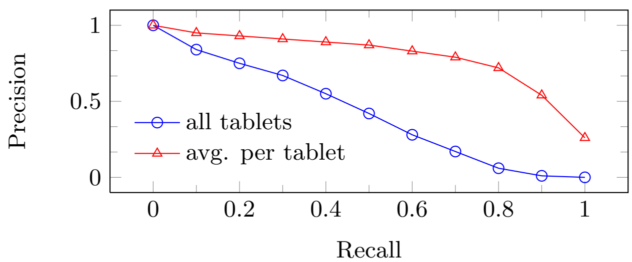


Figure 7: Interpolated precision - recall

terpolation and averaged over all queries. The curves shows that spotting performance is typically very high in the top ranks but continuously decreases from there. The decrease is more rapid when retrieving from all tablets than retrieving only from single tablets. In the former case, for hardly any query all relevant results could be retrieved. Based on the results from table 1 it is reasonable to assume that mostly the detections from the same tablet result in good precision at lower recall levels, while high variabilities across tablets result in worse precision at higher recall levels.

In the qualitative evaluation some interesting properties can be observed with respect to the results typically achieved in the top ranks of the retrieval list. Figure 8 shows examples for all four cuneiform signs that have been considered in the benchmark. On the left we show queries with overall good retrieval performance. In the middle we show queries with overall medium performance and on right queries that can be considered as failure cases.

The first thing that can be observed in all cases is that the top results presented to the philologist are visually very similar to the query. This refers to the texture as well as to the size and aspect ratio of the detections. Consequently, relevant occurrences differing from the query in this regard will not be spotted well. This is due to the query-by-example scenario. From a single exemplary instance of the query cuneiform structure it is hard to generalize and cope with the variability found in the cuneiform tablets.

Further challenges can be seen in the top ranks of queries with medium and poor performance. Cuneiform tablet degenerations, that are part of the selected query, are modeled in the same way as cuneiform wedges. They can have heavy influence on the detections. In figure 8 this can especially be observed for the medium performance query of *an* and the low performance queries of *ti* and *zi*.

It is worth mentioning that cuneiform structures for different signs can be very similar and only be distinguishable by single wedges. Due to their overall high similarity with the query, these cuneiform structures will be retrieved in the top ranks, especially if the count of differing wedges is low compared to the total number of wedges contained in the cuneiform sign. In figure 8 this especially applies to the low performance query of sign *an* and the medium performance query of sign *zi*.

Regarding possible philological use cases, our method is able to provide a significant speed-up for searching cuneiform structures. While manual and correct identification of all occurrences of a single cuneiform sign on a single tablet fragment took at least several minutes, the same task with the additional availability of a heat map could be performed much faster. The list based query results are especially useful when searching for signs with specific writing characteristics across larger collections of fragments, because the possibility of direct visual comparison in many cases eliminates the time consuming need to switch back and forth between photographic documentations or hand-copies. In addition, the generated stylized 2D representation can help with deciphering cuneiform manuscripts. It enhances subtle details, is easily printable and resembles the traditional style of handwritten transliterations.

6. CONCLUSION

In this paper we presented a novel approach for segmentation-free query-by-example retrieval of cuneiform structures. Suitable 2D rasterized image representations were generated from 3D-scan data allowing us to substantially reduce unwanted variabilities that can be found in photographic documentations of cuneiform tablets. We presented a software demonstrator that integrates seamlessly with typical philological research methods. In our benchmark evaluation we achieved remarkable results when retrieving cuneiform signs from single tablets. Due to the high variability observed across different tablets, retrieval performance decreased as expected for queries executed on the complete benchmark data set. In the future we will cope with these challenges by increasing the generalization capabilities of our method.

7. ACKNOWLEDGMENTS

This work is partially supported by the German Federal Ministry of Education and Research within the BMBF project '3D-Joins und Schriftmetrologie'.

8. REFERENCES

[1] L. Adkins. Cuneiform and the Bible. *Empires of the Plain*, 2003.

[2] R. A. Baeza-Yates and B. A. Ribeiro-Neto. *Modern Information Retrieval - the concepts and technology behind search, Second edition*. Pearson Education Ltd., Harlow, England, 2011.

[3] B. Bogacz, M. Gertz, and H. Mara. Cuneiform Character Similarity Using Graph Representations. In P. Wohlhart and V. Lepetit, editors, *Proc. of the 20th Computer Vision Winter Workshop*. Graz University of Technology, 2015.

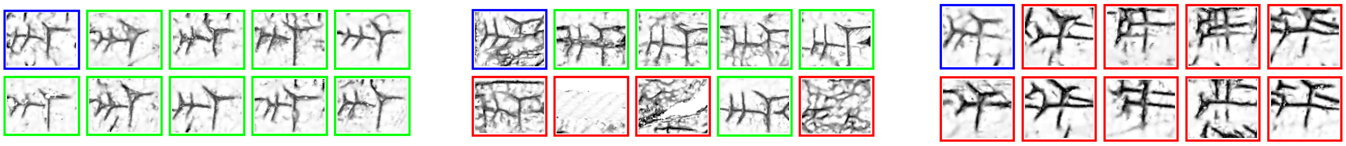
[4] M. Cammarosano. 3D-Joins und Schriftmetrologie: A Quantitative Approach to Cuneiform Palaeography. In *Current Research in Cuneiform Paleography. Proc. of a Workshop held at the 60th Rencontre Assyriologique Internationale*. University of Warsaw, 2014.

[5] M. Cammarosano. The Cuneiform Stylus. *Mesopotamia*, XLIX/2014, 2014.

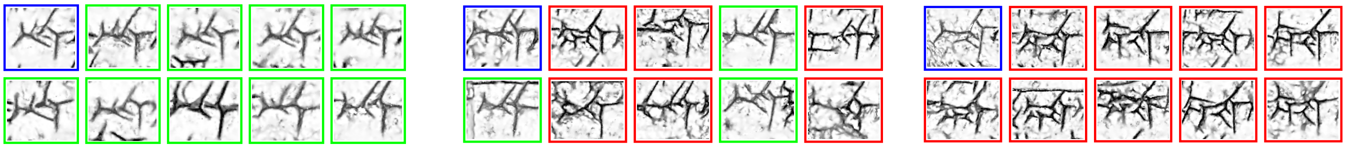
[6] L. Department of the Middle East of the British Museum and Cuneiform Digital Library Initiative (CDLI). British museum collection. Available at cdli.ucla.edu, last seen 2015-06-09.

[7] D. Fisseler, F. Weichert, M. Cammarosano, and G. G. W. Müller. Towards an interactive and automated script feature analysis of 3D scanned cuneiform tablets. In *Scientific Computing and Cultural Heritage*, 2013.

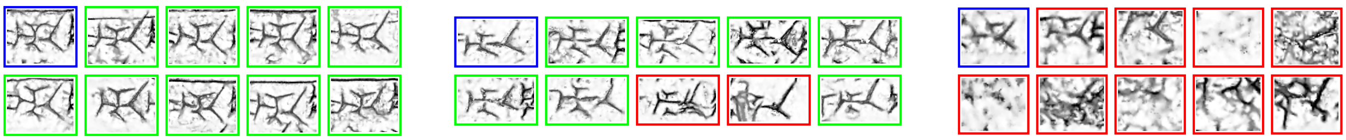
[8] D. Fisseler, F. Weichert, G. G. W. Müller, and M. Cammarosano. Extending Philological Research with Methods of 3D Computer Graphics Applied to



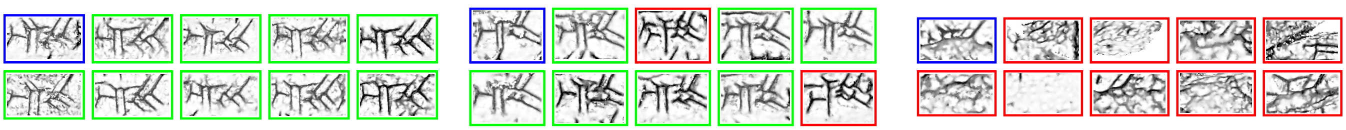
(a) Results for the sign *an*



(b) Results for the sign *na*



(c) Results for the sign *ti*



(d) Results for the sign *zi*

Figure 8: Top query results for a selection of three samples for each of the hittite signs *an*, *na*, *ti* and *zi*. The query is tagged with a blue frame, whereas relevant results are tagged using a green frame and non-relevant results are tagged with a red frame.

Analysis of Cultural Heritage. In R. Klein and P. Santos, editors, *Eurographics Workshop on Graphics and Cultural Heritage*. The Eurographics Association, 2014.

[9] R. Labat. *Manuel d'épigraphie akkadienne. Signes, Syllabaire, Idéogrammes.* Librairie orientalisle P. Geuthner, Paris, 1948.

[10] J. Lladós, M. Rusiñol, A. Fornés, D. F. Mota, and A. Dutta. On the influence of word representations for handwritten word spotting in historical documents. *Int. Journal of Pattern Recognition and Artificial Intelligence*, 26(5), 2012.

[11] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *Int. Journal of Computer Vision*, 60:91 – 110, 2004.

[12] H. Mara and S. Krömker. Vectorization of 3d-characters by integral invariant filtering of high-resolution triangular meshes. In *Proc. of the Int. Conf. on Document Analysis and Recognition*. IEEE Computer Society, 2013.

[13] M. Meyer, M. Desbrun, P. Schröder, and A. H. Barr. Discrete differential-geometry operators for triangulated 2-manifolds. In *Visualization and Mathematics III*. Springer-Verlag, 2002.

[14] G. G. W. Müller. Hethitologie Portal Mainz, 2000. Available at www.hethiter.net, last seen 2015-06-09.

[15] G. G. W. Müller, F. Weichert, D. Fisseler, M. Cammarosano, D. Bachmann, and J. Günnewig. Modellierung einer raumbezogenen Datenbank zur Repräsentation und Analyse syntaktischer und semantischer Merkmale von Keilschrifttafeln. *Datenbank-Spektrum*, 15(1), 2015.

[16] T. Plötz and G. A. Fink. Markov Models for Offline Handwriting Recognition: A Survey. *Int. Journal on Document Analysis and Recognition*, 12(4), 2009.

[17] T. M. Rath and R. Manmatha. Word spotting for historical documents. *Int. Journal on Document Analysis and Recognition*, 2007.

[18] L. Rothacker, G. A. Fink, P. Banerjee, U. Bhattacharya, and B. B. Chaudhuri. Bag-of-Features HMMs for Segmentation-free Bangla Word Spotting. In *Proc. of the Int. Workshop on Multilingual OCR*, 2013.

[19] L. Rothacker, M. Rusiñol, and G. A. Fink. Bag-of-Features HMMs for Segmentation-Free Word Spotting in Handwritten Documents. In *Proc. of the Int. Conf. on Document Analysis and Recognition*, 2013.

[20] L. Rothacker, M. Rusiñol, J. Lladós, and G. A. Fink. A Two-Stage Approach to Segmentation-Free Query-by-Example Word Spotting. *manuscript cultures*, 7(7), 2014.

[21] L. Rothacker, S. Vajda, and G. A. Fink. Bag-of-Features Representations for Offline Handwriting Recognition applied to Arabic Script. In *Proc. of the Int. Conf. on Frontiers in Handwriting Recognition*, 2012.

[22] M. Rusiñol, D. Aldavert, R. Toledo, and J. Lladós. Browsing heterogeneous document collections by a segmentation-free word spotting method. In *Proc. of the Int. Conf. on Document Analysis and Recognition*, 2011.

[23] M. Rusiñol, D. Aldavert, R. Toledo, and J. Lladós. Efficient segmentation-free keyword spotting in historical document collections. *Pattern Recognition*, 48(2), 2015.