Symbol spotting based alignment of transliteration to generate hyperlinks between annotations and image segments

Suman kr. Ghosh
Computer Vision Center, UAB, Barcelona

Introduction:

The goal of this task is to use computer vision to make the current display system of CDLI more interpretable and user friendly. Currently, user needs to read a text to absorb visual and text information simultaneously, and to interpret the mapping between them, since image and transliteration are shown side by side (example: https://cdli.ucla.edu/P423472). This type of annotation is enough for experts in cuneiform studies. However, non-experts and informal learners have no direct means of connecting image and annotation content. Automatic generation of hyperlinks to connect text annotation with corresponding delineation in images can be very helpful to get a richer annotation for informal user and non experts. This will benefit cuneiform paleography.

There has been significant research work in recent years in computer vision for object detection, text detection etc. The input in most works is a scene image represented by a two dimensional picture,. This usually contains natural texture and boundaries which can be effectively used for detecting artifacts inside an image. The cuneiform tablets offers significantly different problem in two aspect first, In contrary to the two dimensional objects, the tablets are inherently three-dimensional, secondly due to ageing in most cases tablets are damaged, which increases the noise in the training algorithm.

The study and annotation of cuneiform data is closer to that of annotation and retrieval of handwriting (on paper) in historical documents. The historical documents are usually scanned from a paper which enables them to be represented by a two dimensional form. But, tablets are inherently three dimensional structures which get distorted when represented in two dimensional form. Another significant difference between handwritten historical documents and tablets is availability of annotated data at desired level of granularity. This aspect demands a change in methodology in case of tablets from historical documents. In case of most computer vision tasks like object detection, semantic segmentation, etc. availability of training material is one of the crucial factor for success. In the study of cuneiform data though there is enough tablets with annotation but they are annotated at semantic level which does not necessarily match with visual boundary. Moreover, the boundary positions are also not available. Thus the annotation is not enough to directly apply state of the art computer vision models.

Hence, generating better annotation for tablets at various granularity and using this information to create hyperlink between image and collected text content necessitates some novel techniques.

Methodology:

Normally bounding box or segmentation mask is used for segmentation. In absence of these data one possible approach could be generating some ground truth based on open-source tools like DIVADIA [1]. However these processes are time consuming and need manual labour inducing human error. A second alternative is to generate a rough segmentation by using unsupervised techniques for example based on projection profile. Similar techniques are popularly used in handwriting segmentation where bounding box at different level is not available during training. These techniques are very inexpensive and multiple segmentation mask can be obtained. In this work we follow one such pipeline.

In brief the proposal is to first use a learning free algorithm to generate a rough segmentation. In this work we use seam carving [2] to generate a rough segmentation of the tablet images. Once segment boundaries are detected, symbol spotting will be employed to make the alignment with the transliteration. Symbol spotting will be used for every symbol in annotation. Note that symbol spotting can be done without any segmentation at all, however a segmentation mask (or region) will help in localizing the search in a given region.

Finally we combine the result of symbol spotting and segmentation mask (output of seam carving) to generate annotation at glyph level.

The following figure illustrates the different components involved in this project.

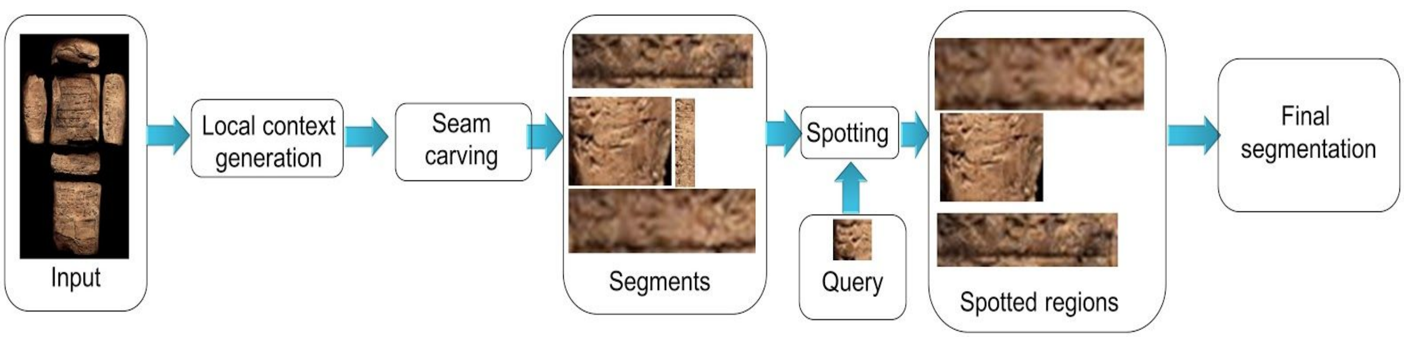

Figure 1: The proposed pipeline

Following sections describe different proposed components as individual methods:

Seam Carving*:*

A seam is a connected path of low energy pixels in an image. Seam carving has been used for image resizing by [2] and for text line segmentation from handwritten documents in [3, 4]. The basic idea is to follow the lowest energy path from left to right (or right to left), where energy map can be obtained by calculating the gradient image for the input image. For a given gray scale image, this can be obtained as

Applying basic seam carving approach has two problems in case of cuneiform tablets

1) An image of Cuneiform tablets constitutes of 6 components representing six sides of a cuboid
2) Usually a seam finds a path between two opposite ends of an image, this way of finding a path can sometimes cut through lines in presence of skew or irregularity in line structure, which is a common phenomenon.

Unlike [3] we can apply seam carving locally to overcome above problems. The hypothesis behind this is though the structures in cuneiform writing can be irregular they must have some regularity in a local context. Secondly all the components (different sides) can be processed independently and later we can merge the outputs. This leads to another problem to define the local context where seam carving would be applied. The local context can be either generated by some heuristic based on projection profile similar to many line segmentation algorithms [4] or can be generated for a small set of samples manually and used as training set.

Glyph (symbol) Spotting:

A zero shot query by example symbol spotting approach is then applied to generate further segments at glyph level. A similar method has been adopted in for handwritten word spotting. We create a list of query examples from a given dataset. For a given dataset we extract one example symbol for every unique glyph present in the dataset. Finally the symbol spotting procedure works as follows:

First regions are extracted by the above seam carving method. Now a tablet can be represented as a set of regions. For each tablet we have access to a list of annotations. For each symbol appearing in the list of annotations we query the symbol in these list of regions (in the same tablet). Symbol spotting will return a list of bounding boxes containing the query symbol which can be sorted according to the similarity with the query, the best retrieved region is used as segmentation mask. Final segmentation is generated by union of all these spotted regions over all queries from one annotation line.

Spotting a query symbol in an image (or image segment), can be done in multiple approaches. different state of the art techniques are available and can be employed. As we do not have annotated data at glyph level our choices will be limited to learning free techniques or zero shot techniques. My proposal is to so called adaptive training scheme, i.e. start with a learning free technique and then gradually move to training based methods by using the generated output as training data.

State of the art image representation like **HOG**, **SIFT**, **BOVW** or **Fisher Vector** can be used. In addition to that deep learning based feature descriptor can also be used, Autoencoder can be used to learn a deep feature representation unsupervised way and can be used for this task.

Training pipeline:

The above procedure will lead to generation of annotated samples at glyph level, However this can be noisy as this is learning free (normally learning free spotter are less accurate than their learning based counterpart)
Once we have significant amount of training data we can use training based symbol spotting for refinement of the end result. Same procedure can be repeated multiple times with user in loop system. Data needs to be segmented in training test and validation sets, the test sets can be in display system to get user feedback.

Generating Hyperlinks:

Once segmentation at glyph level is available its trivial to generate hyperlinks for each glyph and corresponding region in image

Efficacy and Feasibility:

In this section a brief discussion about efficacy and feasibility of proposed approach is described. The efficacy of the proposed approach is inferred from the fact that in handwriting segmentation, annotation and retrieval similar pipelines are used for example [5] proposed a zero shot word spotting, local and global feature based word spotting are very popular and popular examples are [6,7,8]. For segmentation projection profile based and seam carving based approaches are used, which is already mentioned in other sections.

Usability of depth information: Another factor that can influence the digitization process through computer vision is having access to depth information. In current dataset depth information is not available but there are other datasets where depth information is available and can be used to learn a model to infer depth information for the current datasets. Now the question remains that how we can use this information in this proposed framework. Though seam carving framework is designed for images with 2 dimensional input but suitable energy function can be derived which can take depth information while calculating the energy.

Deliverables:

Based on the above proposals the following deliverables are envisioned

1) A segmentation algorithm implemented for cuneiform tablets
2) One or more symbol spotting framework for cuneiform scripts

3) Implementation of a automatic hyperlinking framework for a given set of symbols (script)
4) A user feedback loop, where user can give feedback and rectify errors made by the algorithm (which can be later used in training)

Timeline :

The whole implementation can be divided in two phases 1) Initial segmentation 2) Symbol Spotting. The outline with major subtasks for each phase is given below followed by a gantt chart with detailed plan.

May : Initial Segmentation.

- Set up data and annotation pipeline
- Evaluate efficacy of projection profile based local context generation with respect to manual
- Implement seam carving approach for different energy function
- Generate initial segmentation

I

Jun : Symbol Spotting

- Manually create a list of symbols (used as queries)
- Evaluate different descriptors (SIFT, HOG etc) for symbol spotting in segments generated

July: Training based model

- Combine the outputs to generate segmentation mask
- Use the segmentation mask as training data
- Use the training data for segmentation free symbol spotting and symbol detection

Aug:

- Creation of hyperlink
- Use user feedback in training

About Me : I am a PhD student in my fourth year of studies in the Computer Vision Center at the Autonomous University of Barcelona. Before joining CVC, I worked for IBM India Pvt. Ltd for 3.5 years and Computer Vision and Pattern recognition Unit for 2 years. The topic of my studies is word spotting in heterogeneous environment. I am proficient in C++, python, and matlab/octave. I took part in GSOC 2017 and contributed to text module in OpenCV. I am also a regular user of deep learning frameworks such as theano, caffe, pytorch. My google scholar profile summarises my works and experiences. My github account summarises different projects to which I contribute

References:

[1] http://diuf.unifr.ch/main/hisdoc/divadia

[3] Ryu, Jewoong, Hyung Il Koo, and Nam Ik Cho. "Language-independent text-line extraction algorithm for handwritten documents." *IEEE Signal processing letters* 21.9 (2014): 1115-1119.

[4] Arvanitopoulos, Nikolaos, and Sabine Süsstrunk. "Seam carving for text line extraction on color and grayscale historical manuscripts." *Frontiers in Handwriting Recognition (ICFHR), 2014 14th International Conference on*. IEEE, 2014.

[5] Almazán, Jon, Albert Gordo, Alicia Fornés, and Ernest Valveny. "Segmentation-free word spotting with exemplar SVMs." *Pattern Recognition* 47, no. 12 (2014): 3967-3978.

[6] Likforman-Sulem, Laurence, Abderrazak Zahour, and Bruno Taconet. "Text line segmentation of historical documents: a survey." *International Journal of Document Analysis and Recognition (IJDAR)* 9, no. 2-4 (2007): 123-138.

[7] Ghosh, Suman K., and Ernest Valveny. "A sliding window framework for word spotting based on word attributes." In *Iberian Conference on Pattern Recognition and Image Analysis*, pp. 652-661. Springer, Cham, 2015.

[8] Giotis, Angelos P., et al. "A survey of document image word spotting techniques." *Pattern Recognition* 68 (2017): 310-332.