Creating an Atlas over Handwritten Script Signs

Anders Hast[1], Lasse Mårtensson[2], Ekta Vats[1], and Raphaela Heil[1]

[1] Department of Information Technology, Uppsala University, Sweden
`anders.hast@it.uu.se, lasse.martensson@su.se, ekta.vats@it.uu.se,`
`raphaela.heil@it.uu.se`
[2] Department of Swedish Language and Multilingualism, Stockholm University

Abstract. A framework for interactive visualization of script charac-
teristics, as present in the form of handwritten letters, is proposed in
this work. The basic idea behind this investigation is to lay the foun-
dations for creating a comprehensive atlas over letter forms extracted
from a large collection of handwritten documents, with minimal human
guidance. The visualization of the results is based on the atlas metaphor
and uses the t-SNE visualization method for creating island-like clusters
that can be investigated using the proposed visualization framework. By
changing a scale parameter one can investigate the dataset on different
levels, i.e different sizes of the clusters.

Keywords: Handwritten script, atlas, visualization, t-SNE

1 Introduction

Handwritten script contains an enormous amount of variation in form. In nor-
mally executed handwritten script, each individual instance of a script sign, each
graph, is to some extent unique, even those produced by the same scribe in the
same document. However, in this variation certain regularities exist, and some
graphs display a larger degree of similarity than others. Even though each graph
produced by one specific scribe is unique, as stated, they can still be assumed
to be similar to the extent that a reader can often observe them as being pro-
duced by the same person. The similarities can furthermore be manifested on
different levels. The individual level, i.e. similarities between graphs produced
by the same person is one, but one can also account for similarities manifested
in script produced by different persons, but that have received the same train-
ing and that have been active at the same time period. The latter category of
similarities becomes more relevant when we go back in time. During the Middle
Ages, certain script types existed, such as the Carolingian Minuscule and the
different Gothic variants (Textualis, Cursiva, Hybrida etc.). The time and place
largely determined which styles a scribe learned, but then the script of a certain
scribe also had individual features, unique to this individual.

In this paper, we will address the issue of identifying handwritten script signs,
displaying similarities in a large set of handwritten documents. The documents
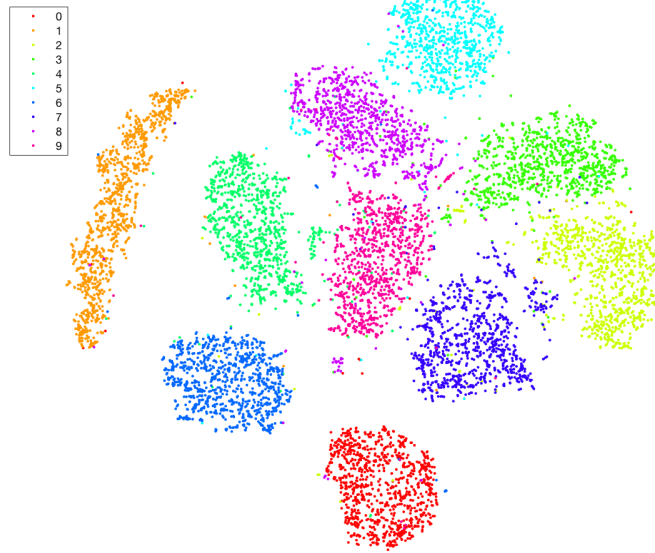are modern, consisting of numbers, but this set has been chosen mainly for the

Fig. 1: The t-SNE based visualization of the MNIST dataset [1] with different numbers depicted using different colours, as shown in the colour bar to the left. One can note that some numbers are miss-classified due to the fact that they are written in a way hard to recognize. Figure best viewed in color.

purpose of evaluating the method as such, and the next step will be to use this method on other material. This is further described below, and the present investigation should be seen as a proof of concept rather than as a concluded task. The documents investigated here have been produced by different scribes, and our aim is to let the computer cluster graphs that share characteristics in form, even though they have been produced by different scribes. What is being identified in the dataset is, thus, classes of graphs that share certain features in form. The ultimate goal for this research track is to lay the foundations for creating an atlas over script signs extracted from a large set of handwritten documents, where the script signs have been clustered into classes on the basis of their similarity to each other.

It should be noted that the clustering process works on two levels of likeness. Firstly, the computer registers the similarity regarding the basic shape of the signs, in our case the different numbers. This means that, for instance, '0, 1, 2,' etc. are divided into basic clusters on the basis of their basic form. Secondly, the separate instances of the numbers are clustered on the basis of the more detailed form, so that those instances of e.g. '7' that share characteristics in form are clustered together. This will be demonstrated further below.

2   Visualization framework

To begin with, the proposed approach allows an end-user to select one or several letters of interest using a simple drag-and-drop gesture to be searched in the handwritten document collection. Typically, the user-marked bounding box rectangle captures more letters than intended, and the background noise in the document renders the user interaction to be more challenging. Therefore, a two band-pass filtering approach based on [2] is employed for automatic background noise removal. Inspired by our previous work [3], the algorithm adjusts the user-marked rectangle based on the input from the user in deciding which letters are to be truly selected. The extent of the letters of interest is located, such that the letters are perfectly encapsulated in the bounding box rectangle and are noise-free. This semi-automatic process initiates faster collection of handwritten letters for creating a comprehensive atlas over a variety of letter forms.

As a proof of concept, the first iteration of the atlas will be created on the basis of the MNIST database [1], which consists of 70,000 grey-level images (28x28 pixels) of handwritten digits (0-9), collected from more than 500 writers. This dataset serves well as a first use case for the evaluation of different clustering, classification and visualization techniques, due to the limited grapheme count and the high variance in writing styles, as a result of the large number of writers. Once a working prototype has been established on this basis, the system can be extended to more complex use cases, such as a higher number of graphemes (i.e. whole alphabets), and a higher variance between writing styles.

We used 10,000 images from the training dataset of MNIST and computed the histograms of oriented gradients (HOG) [4] for each of the samples. HOG features are based on the directions of gradients (i.e. changes of colour or intensity in an image) and encode images into vectors with significantly fewer dimensions. This dimensionality reduction, as well as their previous usage in the context of text recognition (e.g [5]) render HOG features interesting for our application.

Following the conversion of images to HOG features, the t-distributed stochastic neighbor embedding (t-SNE) [6] is applied to the data. This machine learning technique reduces the number of dimensions further (to 2D or 3D), while maintaining a representation of similarities among datapoints (i.e. similar points are represented close and dissimilar points further from each other). Visualizing the transformed datapoints in 2D (respectively 3D) results in a clustering, an example of which is shown in Fig. 1 (for the aforementioned MNIST samples). Ten rather distinct island-like clusters are formed where each point represents a digit from 0-9. Typically, some instances end up in the wrong cluster since they are written in such way that t-SNE is unable to distinguish their feature representations from each other. Some write '1' as others would write '7' etc. In other words, some digits are written quite sloppily and even a human cannot always tell what number they actually represent. This will also be the case for some graphemes such as 'c' and 'e'.
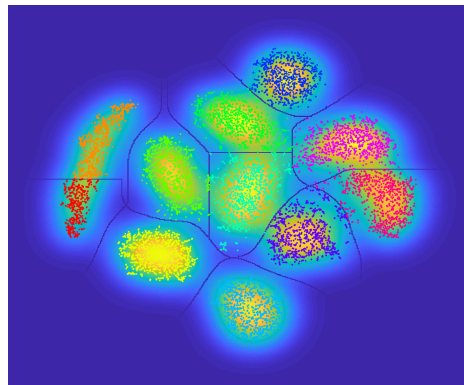
In depth account of the technical details of the proposed approach is not discussed in this paper, as the method is still in the experimental stages and under development. To give an overview, we process the points resulting from
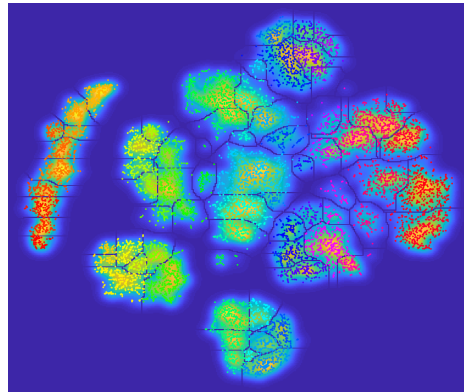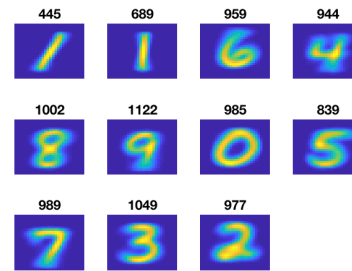
(a) The "islands" clustered on a rather fine scale.
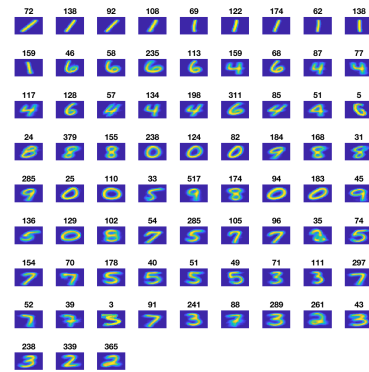
(b) Heat-maps of images in each cluster. Note how the same digit appears in different shapes.

(c) The same data visualized using a coarser scale, generating fewer clusters.

(d) The resulting 11 clusters.

Fig. 2: Visualization of the MNIST dataset. In (d), the resultant 11 clusters captured the 10 digits even if '1' appears twice due to its elongated cluster, which in turn is because of the two main ways of writing the digit. The number above each image in (b) and (d) refers to the number of digits overlaid in the heat-map. Figure best viewed in color.

t-SNE in order to be able to extract clusters using a certain level of detail or scale, defined by the user. This makes it possible to look at different parts of the "island" appearing, and thereby also different clustering of the digits/graphemes, depending on the desired scale. The result of varying the level of detail is shown in Fig. 2. Each "island" is depicted by a varying colour that goes from blue to yellow, depending on the density of the points in the cluster. The blue river-like curves are showing the borders between the obtained clusters. Heat-maps are

also created where all digits in each cluster are aggregated and the average is shown using different colours to indicate the so called "heat", going from blue (low) to yellow (high), via green and orange, just as for the "islands". The more yellow it is, the more similar the digits in the cluster in question are.

By selecting one of the "islands" the user can examine it further. In Fig. 3 all numbers annotated as '1' have been chosen. The shape changes since only a subset of the total number of features is used. The main clusters are shown together with two cells in the heat-map, one that contains quite different looking variants of the same digit while the other depicts quite similar looking digits.

3   Conclusion and future work

This paper presented a visualization framework as a proof of concept that needs to be further investigated as future work. The future aim is to build an interactive visualization tool capable of handling user interaction effectively.

We intend to make use of the proposed approach on the medieval Swedish handwritten documents, with the aim of creating an atlas of the script in this geographic area, and during the period from the appearance of written records (i.e. the 12th century), until the end of the middle ages (i.e. approximately 1520s). Such a catalogue is a very comprehensive yet crucial task, and will be investigated in depth as future work. The details of the classification and the structuring of the atlas, however, remain yet to be solved.

References

1. LeCun, Y., Cortes, C., Burges, C.: Mnist handwritten digit database. AT&T Labs [Online]. Available: http://yann. lecun. com/exdb/mnist **2** (2010)
2. Vats, E., Hast, A., Singh, P.: Automatic document image binarization using bayesian optimization. In: Proceedings of the 4th International Workshop on Historical Document Imaging and Processing, ACM (2017) 89–94
3. Vats, E., Hast, A.: On-the-fly historical handwritten text annotation. In: Document Analysis and Recognition (ICDAR), 2017 14th IAPR International Conference on. Volume 8., IEEE (2017) 10–14
4. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). Volume 1. (June 2005) 886–893 vol. 1
5. Almazn, J., Fernndez, D., Forns, A., Llads, J., Valveny, E.: A coarse-to-fine approach for handwritten word spotting in large scale historical documents collection. In: 2012 International Conference on Frontiers in Handwriting Recognition. (Sept 2012) 455–460
6. Maaten, L.v.d., Hinton, G.: Visualizing data using t-sne. Journal of machine learning research **9**(Nov) (2008) 2579–2605
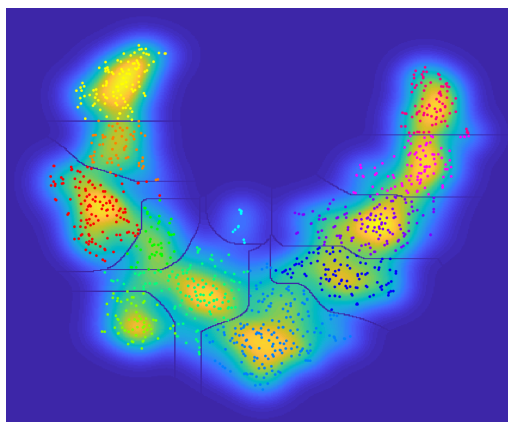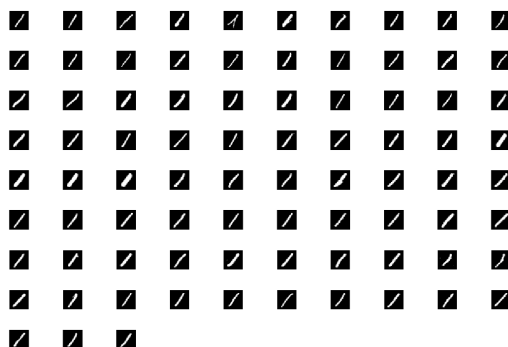
(a) The shape of the "island" becomes a bit different when it is not affected by the digits/clusters.
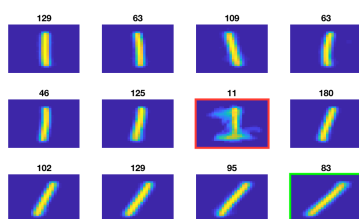
(b) Heat-map of '1', showing the main shapes of that number.
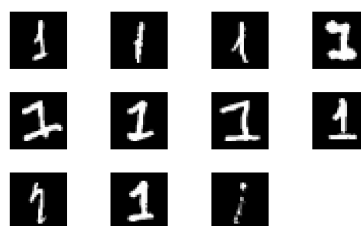
(c) All digits in cell #7 (marked in red in (b)), showing that they are quite different.

(d) All digits in cell #12 (marked in green in (b)), showing that they are indeed quite similar.

Fig. 3: Separate visualization of the number '1'. In (b), note how cell #7 is noisy while the others are quite distinct. Figure best viewed in color.