

Learning to Represent Handwritten Shapes and Words for Matching and Recognition

Jon Almazán

Centre de Visió per Computador, Universitat Autònoma de Barcelona

Advisor/s: Ernest Valveny and Alicia Fornés

21st October 2014, Autonomous University of Barcelona

Received 24th July 2015; accepted 27th July 2015

Abstract

Writing is one of the most important forms of communication and for centuries, handwriting had been the most reliable way to preserve knowledge. However, despite the recent development of printing houses and electronic devices, handwriting is still broadly used for taking notes, doing annotations, or sketching ideas. In order to be easily accessed, there is a huge amount of handwritten documents, some of them with uncountable cultural value, that have been recently digitized. This has made necessary the development of methods able to extract information from these document images.

Transferring the ability of understanding handwritten text or recognizing handwritten shapes to computers has been the goal of many researches due to its huge importance for many different fields. However, designing good representations to deal with handwritten shapes, *e.g.* symbols or words, is a very challenging problem due to the large variability of these kinds of shapes. One of the consequences of working with handwritten shapes is that we need representations to be *robust, i.e.*, able to adapt to large intra-class variability. We need representations to be *discriminative, i.e.*, able to learn what are the differences between classes. And, we need representations to be *efficient, i.e.*, able to be rapidly computed and compared. Unfortunately, current techniques of handwritten shape representation for matching and recognition do not fulfill some or all of these requirements.

Through this thesis we focus on the problem of learning to represent handwritten shapes aimed at retrieval and recognition tasks. Concretely, on the first part of the thesis, we focus on the general problem of representing any kind of handwritten shape. We first present a novel shape descriptor based on a deformable grid that deals with large deformations by adapting to the shape and where the cells of the grid can be used to extract different features. Then, we propose to use this descriptor to learn statistical models, based on the Active Appearance Model, that jointly learns the variability in structure and texture of a given class. Then, on the second part, we focus on a concrete application, the problem of representing handwritten words, for the tasks of *word spotting*, where the goal is to find all instances of a query word in a dataset of images, and *recognition*. First, we address the segmentation-free problem and propose an unsupervised, sliding-window-based approach that achieves state-of-the-art results in two public datasets. Second, we address the more challenging multi-writer

Correspondence to: <jon.almazan@gmail.com>

Recommended for acceptance by Jorge Bernal

DOI <http://dx.doi.org/10.5565/rev/elcvia.728>

ELCVIA ISSN:1577-5097

Published by Computer Vision Center / Universitat Autònoma de Barcelona, Barcelona, Spain

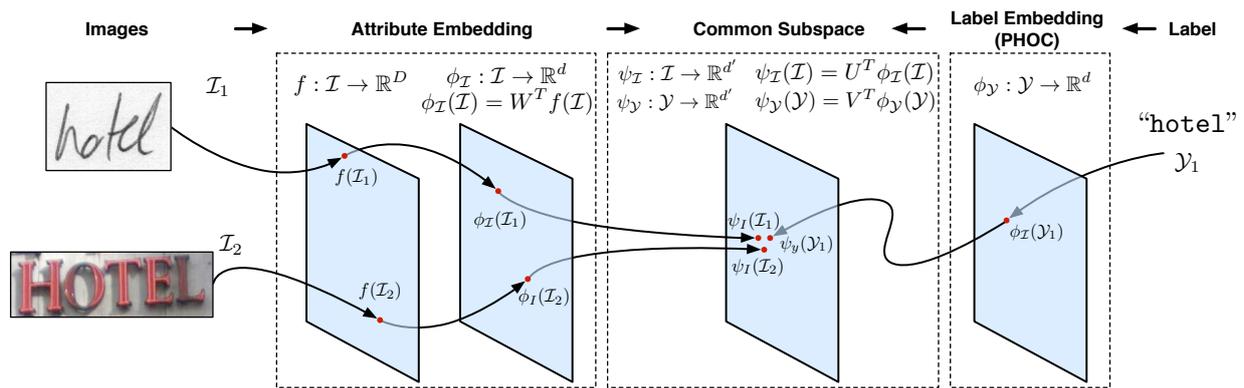


Figure 1: Overview of the word embedding method of images and strings in a common vectorial space.

problem, where the variability in words exponentially increases. We describe an approach in which both word images and text strings are embedded in a common vectorial subspace, and where those that represent the same word are close together. This is achieved by a combination of label embedding and attributes learning, and a common subspace regression. This leads to a low-dimensional, unified representation of word images and strings, resulting in a method that allows one to perform either image and text searches, as well as image transcription, in a unified framework. An overview of the method can be seen in Figure ???. We evaluate our methods on different public datasets of both handwritten documents and natural images showing results comparable or better than the state-of-the-art on spotting and recognition tasks.

Selected Publications

Journals

- Almazán J., Gordo A., Fornés A., Valveny E., Word Spotting and Recognition with Embedded Attributes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2014.
- Almazán J., Gordo A., Fornés A., Valveny E., Segmentation-Free Word Spotting with Exemplar SVMs. *Pattern Recognition*, 2014.
- Almazán J., Fornés A., Valveny E., A non-rigid appearance model for shape description and recognition. *Pattern Recognition*, 2012.

International Conferences

- Almazán J., Gordo A., Fornés A., Valveny E., Handwritten Word Spotting with Corrected Attributes. In *International Conference on Computer Vision*, 2013.
- Almazán J., Fornés A., Valveny E., Deformable HOG-based Shape Descriptor. In *International Conference on Document Analysis and Recognition*, 2013.
- Almazán J., Gordo A., Fornés A., Valveny E., Efficient Exemplar Word Spotting. In *British Machine Vision Conference*, 2012.
- Almazán J., Fornés A., Valveny E., A Non-Rigid Feature Extraction Method for Shape Recognition. In *International Conference on Document Analysis and Recognition*, 2011.