

A Fast System for the Retrieval of Ornamental Letter Image

Mathieu Delalandre^{*} and Jean-Marc Ogier⁺ and Josep Lladós^{*}

^{}CVC, UAB, Barcelona, Spain*

{mathieu,josep}@cvc.uab.es

⁺ L3i, La Rochelle, France

jean-marc.ogier@univ-lr.fr

Abstract

This paper deals with the retrieval of document images especially applied to the digitized old books. In these old books our system allows the retrieval of graphical parts and especially the ornamental letters. The aim of our system is to process large image databases. For this purpose, we have developed a fast approach based on a Run Length Encoding (RLE) of images. We use the RLE in an image comparison algorithm using two steps: one of image centering and then a distance computation. Our centering step allows to solve the shifting problems usually met between scanned images. We present experiments and results about our system according to criteria of processing time and recognition precision.

Keywords: Image Retrieval, Ornamental Letter, Complexity, Run Length Encoding, Shape Recognition.

This paper deals with the topic of the image retrieval and especially the document images. During the last years many works have been done for the retrieval of journals, forms, maps, drawings, musical scores ... In this paper we focus ourselves on a new retrieval application: the one of old books. Indeed, since the Digital Libraries development in the years 90's numerous works of digitization of historical collections have been done. Nowadays, large Digital Libraries of old books are available on the Web and will still grow in the future. However, few systems have been proposed to retrieve graphical parts from these collections. Only the works of [1] [2] [3] [4] exist and each of them is dedicated to specific kind of retrieval: [1] retrieves similar stroke based illustrations inside old books, [2] tracks common sub-parts in old figures, [3] and [4] retrieve ornamental letters according to graphical style and image layout criteria.

In this paper we propose a new retrieval application of old graphics: the wood plug tracking. Indeed, from the 16th to the 17th centuries the plugs, used to print graphics in the old books, were mainly in wood. The Figure 1 gives examples of printings produced by a same wood plug. Most of these wood plugs were used to print ornamental letters. These wood plugs could be re-used to print several books, be exchanged between printing houses, or reproduced in the case of damage. So to retrieve, in automatic way, printings produced by a same wood plug could be very useful for the historian people. It could solve some dating problems of books as soon as to highlight the existing relations between the printing houses.



Figure 1. Examples of printings produced by a same plug

This retrieval application can be viewed as a classical image comparison. Indeed the images produced by a same wood plug present similarities at pixel level. However, this raises a complexity problem. First in regard to the amount of data, building a comparison index between thousands of image can require days of computation. Next in regard to the copyright aspects. The images belongs to specific Digital Libraries or private collections; in order to allow crossed queries between these databases real-time retrieval processes are needed. In regard to these specificities we have developed a system to perform a fast retrieval of images. This one is presented in the Figure 2.

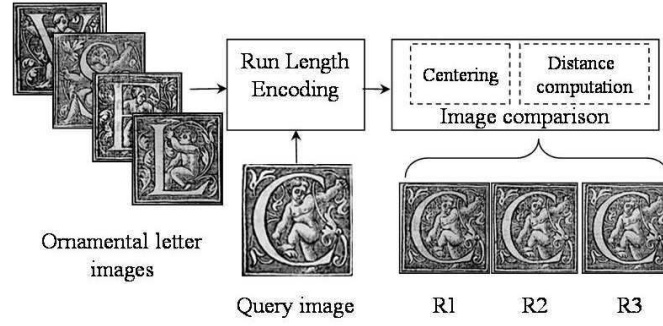


Figure 2. System overview

Which interest us in our approach is to process, in a fast way, the image databases for the retrieval. So it is necessary to decrease the processing times of our algorithms. To do it in our system we have used a run based representation of images. The run is well known data structure. It encodes successive pixels of same intensity into a single object. The conversion of a raster image to a set of run is called Run-Length Encoding (RLE). We use the RLE in our comparison algorithm using two steps: one of image centering and then a distance computation. Our centering step allows to solve the shifting problems usually met between scanned images. We use next the RLE to perform the image comparison. This one is done by a “simple” difference pixel to pixel between two images. However, it is based obviously on the RLE handling. For that purpose we have developed a specific algorithm detailed in the next Figure 3.

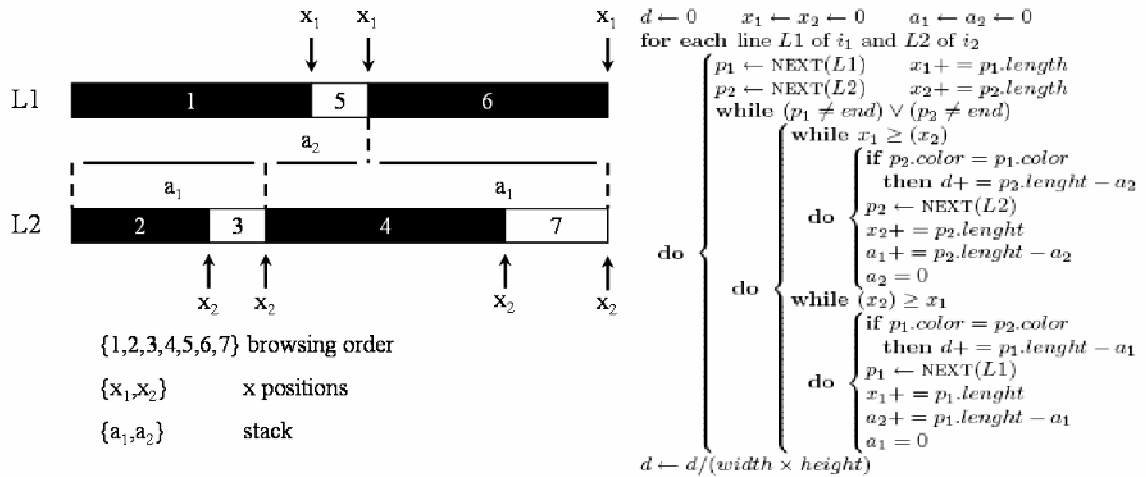


Figure 3. RLE comparison

We present different experiments and results about our system. These experiments show how our system allows to compress from 8 to 9 times the image sizes, and therefore to reduce the needed retrieval times. We also illustrate the retrieval precision of our system through examples of query result.

References

- [1]. J. Bigun, S. Bhattacharjee, and S. Michel. Orientation radiograms for image retrieval: An alternative to segmentation. In *International Conference on Pattern Recognition (ICPR)*, volume 3, pages 346-350, 1996.
- [2]. E. Baudrier, G. Millon, F. Nicolier, R. Seulin, and S. Ruan. Hausdorff distance based multiresolution maps applied to an image similarity measure. In *Optical Sensing and Artificial Vision (OSAV)*, pages 18-21, 2004.
- [3]. R. Pareti and N. Vincent. Global discrimination of graphics styles. In *Workshop on Graphics Recognition (GREC)*, pages 120-128, 2005.
- [4]. S. Uttama, M. Hammoud, C. Garrido, P. Franco, and J. Ogier. Ancient graphic documents characterization. In *Workshop on Graphics Recognition (GREC)*, pages 97-105, 2005.