A multi-layer separation based system for camera-based complex map image retrieval

Q.B. Dang[(1)], M.M. Luqman[(1)], M. Coustaty[(1)], N. Nayef[(1)], C.D. Tran[(2)], J.M. Ogier[(1)]

[(1)] L3i Laboratory, University of La Rochelle, France
[(2)]College of ICT, Can Tho University, Vietnam

Abstract. In this paper, we present a method of camera-based document image retrieval for heterogeneous-content document using a multi-layer separating approach. We use Locally Likely Arrangement Hashing (LLAH) extracting text features on the layer which contains text. In addition, we employ a technique of reducing the memory required for storing the hash table. Experiment result show that our approach is efficient in term of accuracy result and real-time retrieval for heterogeneous-content document camera-based retrieval.

Keywords: camera-based document image retrieval, automatic indexing, text/graphic separation, feature extraction.

1. Introduction and related work

Camera-based document image retrieval is a task of searching document images relevant to user's query that is captured by a digital camera. This task requires not only to tackle the problem of "perspective distortion" of images, but also to establish a way of matching document images efficiently. Recently the method called Locally Likely Arrangement Hashing (LLAH) has been known as the efficient and real-time camera-based document image retrieval method. It is based on local combination of affine invariant calculated from feature points which are extracted from centroid of each word connected component (Nakai et al., 2007). Because of this, accuracy of retrieval will reduce when it is applied to rich graphics document.

Text/graphics separation is a process segmenting a document image into two layers, one containing text and the other containing graphics. From several decades, many methods have been proposed to solve this problem. Color-based has been used for separating an image into many layers (Dhar et al., 2006; Ebi et al, 1994). Furthermore, connected component (CC) analysis was widely used for this separation. For instance, Karl Tombre (Tombre et al., 2002) proposed a size-histogram analysis from the bounding boxes of all CCs. Winfried Höhn (Höhn, 2013) used density of CC that is a ratio between the area of the convex hull and the number of pixels in CC. Further, they used diameter ratio that is the ratio between minimum diameter of CC and maximum diameter of CC. In this paper, we propose a method of multi-layer separating for camera-based document analysis and retrieval of complex map images which are composed of heterogeneous-content.

3. Proposed method

In this section, we describe our system for camera-based complex map image retrieval using a multi-layer separating approach. Our method aims to separate document image into 2 layers using attributes of CC and extract LLAH features from text layer for storing in hash table and retrieval.

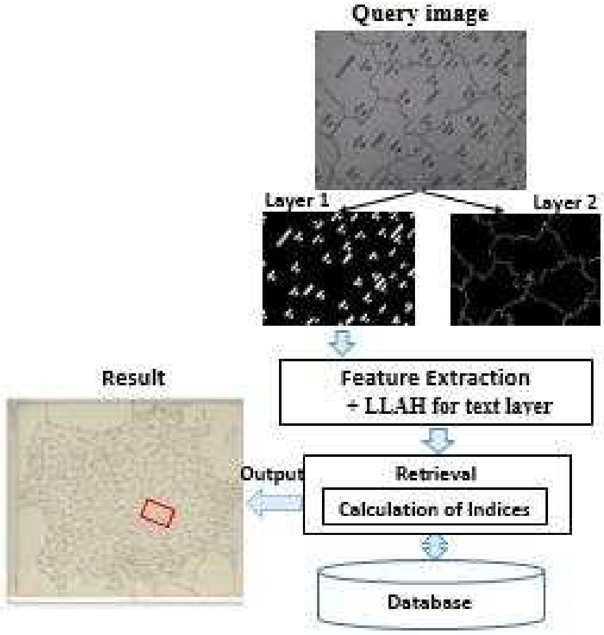

Figure 1: Retrieval phase

3.1. Multi-layer separating

Our method is outlined in **Figure 1**. In both indexing phase and retrieval phase, document image is separated into 2 layers. For complex maps data, the linguistic maps of France, attributes of CC are used for separating the image into 2 layers. Layer 1 contains word CCs**,** and layer 2 contains graphics CCs which are the borders of the map, **Figure 1**. In order to extract word CCs, the image is converted into binary image firstly. Next, the binary image is blurred using the Gaussian filter whose parameters are determined based on an estimated character size (the square root of a mode value of areas of CCs). Then, the blurred image is applied adaptive-threshold again (Nakai et al., 2007). Finally, all CCs are extracted. Consequently, dashed and dotted lines are also joined to large CCs or long CCs which is a base for separating graphics layer. Owing to the attributes of CC such as CC area, area of CC's bounding box, maximum diameter of CC, large CCs and long CCs can be extracted into the layer 2. Those whose attributes are bigger than thresholds are considered as large CCs. The thresholds are determined by choosing mean value multiply by 2 (for each attribute). Moreover, very small CCs which are noise need to be discarded.

3.2. Feature extraction

We extract feature vector of LLAH within text layers (layer 1), which can be obtained even under perspective distortion, noise, and low resolution. For the reason that we have not used the layer 2 (borders of map), only feature vectors of LLAH were extracted from layer 1(text layer). We use affine invariant as a discrete value of the invariant that is defined using 4 coplanar points, each point is the centroid of each word region bounded by word CC. So, LLAH vector is calculated from the arrangement of m points that is described as a sequence of invariants calculated from all possible combinations of 4 points taken from m points (Nakai et al., 2007). We also add additional feature using rank of area ratios of word regions at each m points (Takeda et al., 2011) so that it can deal with fewer words case. As a result, LLAH feature vector has $C_m^4 + m$ dimensions, where m can be set greater than or equal 6. According to Nakai, all combinations of m points from n nearest neighbor points (following clockwise order) from one point p are examined. Because of this, there are C_n^m LLAH feature vectors computed from one point (n >m).

3.3. Indexing phase

One of the usefulness of auto indexing method is that when adding a new document into hash table, the existent database does not need to be reconstructed or recomputed (Kise et al., 2007). The document is separated into 2 layers firstly. Then for layer 1 (text layer), all feature vectors of LLAH are extracted and indexed using hashing function following formula (1). Aiming to reduce the required amount of memory, we do not store feature vectors in hash table. So as to index, the real-valued feature vectors need to be converted into integers. The index H_{index} of the hash table is calculated by the following hash function:

$$H_{index} = \left(\sum_{i=0}^{n-1} r_{(i)} k^i\right) \bmod H_{size} \tag{1}$$

Where r is a LLAH vector, n is the number of dimension of r, k is the level of quantization (e.g. k=15), H_{size} is the size of hash table (Nakai et al., 2007).

3.4. Retrieval phase

Figure 1 shows the block diagram for retrieval phase of the system. For text layer separated from captured query image, LLAH feature vectors are extracted firstly. Next, for each LLAH feature vector, it is used for searching in hash table and voting for document ID containing it. Finally, the document with majority vote is returned as the retrieval result. We employ the RANSAC algorithm to find a perspective transformation T between set of query point IDs and set of document point IDs matching together after voting, and transformation T is used for spotting region of interest.

4. Experimentation

The experimentation is performed on a dataset of the linguistic map of France. There are 12 images of resolution 9800 x 11768 pixels in the dataset. The map dataset along with the ground truth information is made publically available for academic research. It can be downloaded from http://navidomass.univ-lr.fr/MapDataset. For online retrieval phase, Samsung document camera SDP-760 is used. For LLAH parameters, we set n=8 and m=6, and for hash table H_{size} is set equal 128 x 10^8. In order to perform experiments of real-time document image retrieval, the camera was fixed at 20 cm above surface of the captured map. Size of the captured images is 640 x 480. Each map was captured by 6 videos recorded at different regions in the map (top left, top right, middle left, middle right, bottom left and bottom right). We use first 20 frames of each video. So, there are 1440 frames used for testing. For each frame, the retrieval is correct if the ID of a returned map image is correct and region of interest retrieval is correct in the map**.** In order to evaluate our proposed method, we also tested in case of unseparated graphics layer. Our method outperformed the unseparated method. In overall, the accuracy of retrieval of our method reached a level of 87% while unseparated method reached level of 72%.

5. Conclusion

We have presented our work on a multi-layer information spotting system for camera-based heterogeneous content document image retrieval. The multi-layer separating approach has produced promising initial results for camera-based heterogeneous content complex map images retrieval. We are working on to take forward our system using dedicated feature like PCA-SIFT or SIFT to extract feature vector from the graphics layer. Work is in progress to extend our system to multi-layer (>2) for automatic indexing and retrieval of scanned newspapers.

Acknowledgement

This work has been partially supported by the LabEx PERSYVAL-Lab (ANR-11-LABX-0025), by the CNRS PEPS Project CartoDialect, and by the Program 165.

Bibliography

Nakai Tomohiro, Koichi Kise, and Masakazu Iwamura. "Camera based document image retrieval with more time and memory efficient LLAH." *Proc. CBDAR* (2007).

Kazutaka Takeda, Koichi Kise, and Masakazu Iwamura. "Memory reduction for real-time document image retrieval with a 20 million pages database." In *Proceedings of the 4th International Workshop on CBDAR*. 2011.

Tombre Karl, Salvatore Tabbone, Loïc Pélissier, Bart Lamiroy, and Philippe Dosch. "Text/graphics separation revisited." In *DAS*. Springer Berlin Heidelberg, 2002.

Winfried Höhn. "Detecting Arbitrarily Oriented Text Labels in Early Maps." In *Pattern Recognition and Image Analysis*. Springer Berlin Heidelberg, 2013.

Dhar Deeptendu Bikash, and Bhabatosh Chanda. "Extraction and recognition of geographical features from paper maps. *IJDAR* 8, no. 4, 2006.

N. Ebi, Bernd Lauterbach, and Walter Anheier. "An image analysis system for automatic data acquisition from colored scanned maps." *Machine Vision and Applications* 7, no. 3, 1994.