2010 International Conference on Pattern Recognition

Fast Logo Detection and Recognition in Document Images

Zhe Li[1], Matthias Schulte-Austum[1], and Martin Neschen[2]

[1]*Siemens AG, Industry Sector, Mobility Division, Infrastructure Logistics*
Konstanz, Germany
zhe.zl.li@siemens.com, matthias.schulte-austum@siemens.com

[2]*Recosys GmbH*
Aachen, Germany
Martin.Neschen@recosys.de

***Abstract*—The scientific significance of automatic logo detection and recognition is more and more growing because of the increasing requirements of intelligent document image analysis and retrieval. In this paper, we introduce a system architecture which is aiming at segmentation-free and layout-independent logo detection and recognition. Along with the unique logo feature design, a novel way to ensure the geometrical relationships among the features, and different optimizations in the recognition process, this system can achieve improvements concerning both the recognition performance and the running time. The experimental results on several sets of real-word documents demonstrate the effectiveness of our approach.**

***Keywords*-logo recognition; logo detection; image processing; document retrieval**

I. Introduction

Digitization provides an efficient way to preserve, process, and transport all kinds of information. On the other hand the question arises how to find the relevant information in an ocean of data. Regarding the increasing size of data to be searched, precision is no longer the only criterion for performance. The speed of search plays an important role as well. Traditionally document image processing is dominated by a technique named Optical Character Recognition (OCR), which has achieved very good results on text reading in documents. However, beside text information some documents also contain information in graphical symbols. OCR engines, subsequent word correction algorithms, and text interpretators cannot derive this information completely and sometimes it takes high latency time to extract complete text information of complex documents. This process can be improved significantly by means of logo recognition. For example, logo recognition can support applications where it is required to search the documents from a certain company by identifying a certain graphical logo. In addition to company logos, certain unique graphical objects, and even certain text lines in the document images can be considered as symbols by which the documents can be identified and categorized. The recognition of them provides a fast way to explore the documents.

The detection and recognition of logos faces many difficulties in the field of object recognition in the computer vision (CV) community. For example, logos have large intra-class variations. Different printers and scanners can generate images with varying degrees of quality and degradations. The surrounding text, diagrams, manual remarks, and labels on the documents form a clutter background.

Most of the former research has focused only on logo recognition [1], [2], [3], where it is assumed that the segmentation of the logos has been done by a different module. In contrast to that, Zhu and Doermann use a unified framework which combines logo detection and recognition based on a multi-scale strategy [4]. The initial classification is performed on each connected component using the Fisher classifier at a coarse scale level, regarding the logo region as a gray-scale blob. Each logo candidate region is further classified at successively finer scales by a cascade of simple classifiers. Recently, more papers have been dedicated to the combined solutions [5], [6], [7]. Wang and Chen used a similar framework as [4]. Instead of a multi-scale approach, they used a method named boundary extension of feature rectangles to group the logo components together. Both approaches assume that the components of a logo are near to each other and that they can be grouped together by scaling the images or finding their neighborhoods. However this assumption is too restrictive because of the clutter image background or the large variance of the logo design.

Enlightened by the object recognition strategy in the CV community, Rusinol and Llados chose a feature-based approach [6]. Using features as “words”, the logos are represented by a bag-of-words model [8]. An important drawback of the bag-of-words model is that the spatial arrangement of the features is ignored. In order to address this drawback, [6] used an “opening” operation to find the features which are in clusters. They explored two kinds of features, SIFT [9] and shape context descriptor [10], which are both computationally expensive.

In contrast to the previous work, we are aiming at a fast and truly segmentation-free and layout-independent solution. To this end, we define a logo as a group of features with restrictive geometrical relationships. In order to detect and recognize this kind of logos, the following contributions have been done:

- A generic logo detection and recognition framework with three layers;
- A simple and appropriate feature design;
- A novel geometrical reconstruction method.

With all these contributions, our approach can work on a large variety of logos and other graphic objects, even texts in the document images, and achieve improvements concerning both recognition performance and running time.

In the next section, we first introduce the system architecture. Then the feature design, the geometrical reconstruction of logos and the verification are addressed in detail. In Section 3, the databases used in the experiments are explained, and the results are analyzed and compared to those of other approaches.

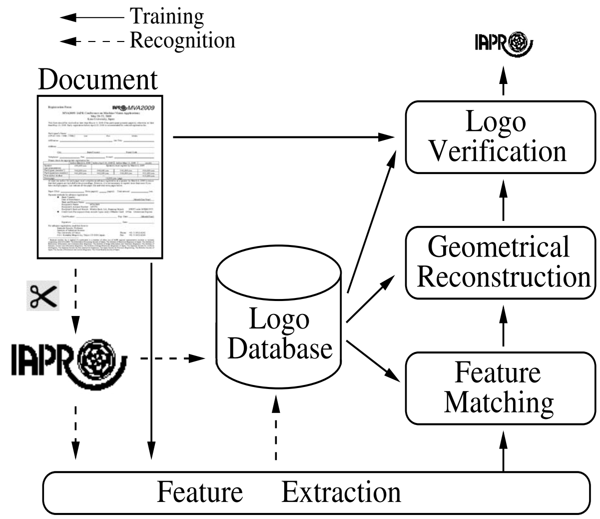

Figure 1. System architecture

II. System for Detection and Recognition

The system architecture is shown in Figure 1. In the training phase, only one instance of each logo model is required. The region of this prototype is selected in the document image. The features in this region are extracted and saved in a database together with a normalized image of this logo. During the training phase, the prototype features are consolidated, which means that the same features used by different prototypes are saved only once. All features are indexed in a hyper-cubic structure, which leads to a fast searching and matching. In the detection and recognition phase, all features on the target image are extracted at first. Then they are matched to the features in the database. The resulting matches include both correct feature matches to the existing logo prototypes and spurious mismatches. In order to check the consistency of pairwise distances between feature candidates which are matched with the same logo prototype, a novel concept named *anchor lines* is introduced. Using the consistent mapping of input features with logo prototype features, we derive the position of the logo bounding box on the input image, which refers to the region chosen in the training phase. Finally, a verification is done by comparing the normalized image in the bounding box with the saved prototype image to ensure a correct matching. The following subsections will address these points in detail.

A. Features

Local descriptors are popular features used in CV community for object detection. Their calculation normally includes two steps. First the points of interest are determined using corner detectors or as local maxima/minima of differences of Gaussians. Then local descriptions are built based on neighboring pixels of these points. Because of the large amount of points of interest and the complexity of the local descriptors, their calculation and matching are quite expensive. Therefore, we introduce two kinds of features for logo recognition. The first one is the scale and rotation invariant shape descriptor of the connected components, which are the simplest objects and can be quickly extracted from a binary image. In order to better describe a logo prototype using connected components, several "layers" of a logo image are defined. These layers differ in resolution, color channel, scale, contrast and binarization method in order to be able to generate robust features. At the end, every resulting binary image is one "layer" and the connected components are extracted from it. For every connected component, its convex hull is calculated. Figure 2 gives an example. The black area is a connected component and the dashed line is the convex hull of it. For a convex hull, its orientation θ using the 2^{th} moments, the square root of the variance of the main axis with respect to the orientation $\sigma_{\max}$, the square root of the variance on its orthogonal direction $\sigma_{\min}$, the maximum length $l_{\max}$, minimum edge $l_{\min}$, and the square root of its size l_{size} are also calculated. Then, a connected component descriptor CComp is constructed as:

$$
\begin{aligned}
\mathrm{CComp} = \{&\kappa_r, \kappa_\sigma, \kappa_s, \kappa_l, \kappa_b, l_{\text{size}}, \mathrm{IDs}, \vec{\mathrm{ac}}\} \\
\text{with}\quad & \kappa_r = l_{\max}/l_{\min}; \\
& \kappa_\sigma = \sigma_{\max}/\sigma_{\min}; \\
& \kappa_s = l_{\text{size}}/\sigma_{\max}; \\
& \kappa_l = l_{\text{size}}/l_{\max}.
\end{aligned}
$$

In it, κ_b means blackness, a division between the number of pixels of the connected component and the area of the convex hull. The descriptor elements $\kappa_r, \kappa_\sigma, \kappa_s, \kappa_l, \kappa_b$ are designed to be invariant to image scaling and rotation. The

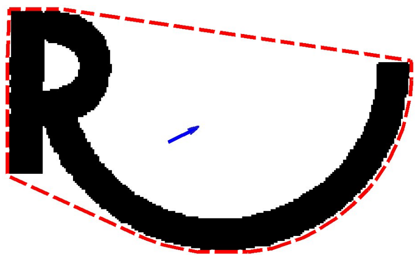

Figure 2. The convex hull (dashed line) and anchor line (arrow) of a connected component

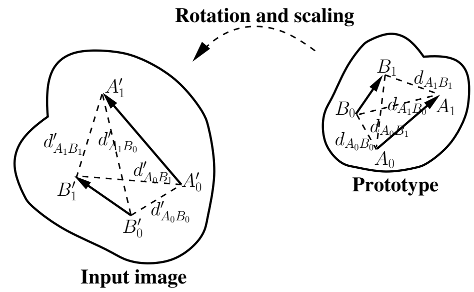

Figure 3. Illustration of the geometrical matching using anchor lines

feature level matching using them is only a calculation of Euclidean distance. With these parameters as axis, all features can be indexed in a hyper-cubic structure, which leads to even faster searching and matching. As mentioned before, a component can be shared by several prototypes. The IDs is used to save these references. Each reference includes three indices respectively for prototype, layer and the feature belonging to this prototype. $\vec{ac}$ is the anchor line of the connected component. The elements in the descriptor are used for geometrical reconstruction, which will be explained in the next subsection.

A different feature used are the line profiles, which are two curves describing the upper and lower boundary of a text line. Two lines can be matched by calculating the cross correlation of their smoothed profiles. If text lines are found in a logo, they can be significant features.

B. Geometrical Reconstruction and Verification

An input document image can have many feature-level correspondences in the database. From IDs of the matched prototypes features, the number of feature matches for a prototype can be derived. However, if the matched features are not structured in the same way as the prototypes, they are just invalid votes. In order to find the features within the same geometrical relationship, the anchor line is introduced. For connected components, it is a line from the gravity center of the connected component pointing to the gravity center of the area in its convex hull. In Figure 2, the arrow is an example of it. For text lines, whose bounding boxes are rectangles with large width/length ratio, the anchor line simply connects the left edge with the right edge. An anchor line has always the same relative angle to the main axis of the feature it belongs to and encodes the scaling information.

Figure 3 illustrates how we check the consistency of the relative position of two features in the prototype and the input image. Suppose two features A, B of a logo prototype have anchor lines $\vec{A_0A_1}$ and $\vec{B_0B_1}$. In the input image, A and B found their matches A' and B'. The anchor lines of them are $\vec{A'_0A'_1}$ and $\vec{B'_0B'_1}$. If the four distances $d_{A_0B_0}$, $d_{A_1B_1}$, $d_{A_1B_0}$ and $d_{A_0B_1}$ have the same proportion β to their correspondences $d_{A'_0B'_0}$, $d_{A'_1B'_1}$, $d_{A'_1B'_0}$ and $d_{A'_0B'_1}$, it can be said that that these two matched features have the same geometrical relationship to their prototype. If A' or B' finds another feature which complies with the rule above for the same β, and so on, this pair of features can be extended to a clique. If the number of the features of a clique exceeds a pre-defined threshold, a logo candidate is found. The scaling factor is β. The rotated angle α is the relative angle of two vectors: the sum of the anchor lines in the clique and of their matches in a prototype.

When a logo candidate is selected, the region of the logo candidate in the input image is determined according to the matched anchor lines, α, β and the selected region during logo learning. Then, this region is cut out, normalized and compared to the prototype image in the database. If the percentage of the different pixels is below a pre-defined threshold, this candidate is verified.

III. Experiments

In the experiments, we first tested our approach on the public Tobacco-800 database [11], [12], [13], [4]. It is a realistic database for document analysis and retrieval, as these documents were collected and scanned using a wide variety of equipment over time. The dimensions of documents range from 1200 x 1600 to 2500 x 3200 pixels with resolutions varying from 150 to- 300 DPIs. The ground truth of the logos is also publicly available [4].

The results are shown in Table I. Suppose the number of logos in the ground truth is n_t, the number of detected logo is n_d, and the number of correct detected logo is n_c, the recall and precision are defined as:

$$\begin{aligned} recall &= n_c/n_t, \\ precision &= n_c/n_d. \end{aligned}$$

It can been seen that our approach has a higher precision combined with an improved recall. At the same time, it requires less than half of the CPU time of [4]. Our number

Table I
RESULTS ON THE TOBACCO-800 DATABASE

Approaches	N. of Images	N. of Logos	N. of Training Set	Recall	Precision	Mean Time (ms)
Zhu and Doermann[4]	1290	432	50	84.2%	73.5%	680
Wang and Chen[5]		416	100	80.4%	93.3%	absent
Our approach		**415**	**50**	**86.5%**	**99.4%**	**328**

of logos (415) differs from the ground truth publisher (432) because the logos which appear only once in the documents are excluded. The running time is measured on a single core of an Intel Core Duo 2.00GHz CPU. Rusinol and Llados report their results based on another database [6], which makes a direct comparison very difficult. Generally speaking, they also have a high precision in results. By the use of SIFT features, a better recall could be achieved with a much longer running time (>3s). When shape context features are used, they reduce the CPU time to 2s per image with a precision drop by 10%. Our approach is faster than theirs and has comparable results.

Another experiment was done in a letter sorting task. In 5000 real letters scanned by the post office with resolution 96 DPI, 1191 letters had company logos. With a logo database of 262 logos, a read rate of 98.7% (1175/1191) and an error rate of 0.04% (2/5000) have been achieved. The average running time was only 75ms.

IV. SUMMARY

In this paper, we introduced a system which is aiming for a fast, segmentation-free and layout-independent logo detection and recognition on document images. Regarding a logo as a group of features with restricted geometrical relationships, we presented three major contributions. At first, a generic feature-based recognition scheme was put forward. Secondly, we introduced two kinds of logo features, shape descriptors of the connected components and line profiles which are simple and representative. Thirdly, a new concept named anchor line was introduced, which provides a novel way to reconstruct logo prototypes and conduct the final verification. Several experiments on real documents proved the effectiveness of our approach.

V. ACKNOWLEDGEMENT

This work was partially funded by the German Federal Ministry of Economy and Technology (BMWi) under the THESEUS project.

REFERENCES

[1] D. Doermann, E. Rivlin, and I. Weiss, “Applying algebraic and differential invariants for logo recognition,” *Mach. Vision Appl.*, vol. 9, no. 2, pp. 73–86, 1996.

[2] J. Neumann, H. Samet, and A. Soffer, “Integration of local and global shape analysis for logo classification,” in *IWVF-4: Proc. of the 4th Int. Workshop on Visual Form*, London, UK, 2001, pp. 769–778.

[3] C.-H. Wei, Y. Li, W.-Y. Chau, and C.-T. Li, “Trademark image retrieval using synthetic features for describing global shape and interior structure,” *Pattern Recognition*, vol. 42, no. 3, pp. 386–394, 2009.

[4] G. Zhu and D. Doermann, “Automatic document logo detection,” in *ICDAR ’07: Proc. of Int. Conf. on Document Analysis and Recognition*, Washington, DC, USA, 2007, pp. 864–868.

[5] H. Wang and Y. Chen, “Logo detection in document images based on boundary extension of feature rectangles,” in *ICDAR ’09: Proc. of the Tenth Int. Conf. on Document Analysis and Recognition*, Barcelona, Spain, 2009, pp. 1335–1339.

[6] M. Rusinol and J. Llados, “Logo spotting by a bag-of-words approach for document categorization,” in *ICDAR ’09: Proc. of the Tenth Int. Conf. on Document Analysis and Recognition*, Barcelona, Spain, 2009, pp. 111–115.

[7] G. Zhu and D. Doermann, “Logo matching for document image retrieval,” in *ICDAR ’09: Proc. of the Tenth Int. Conf. on Document Analysis and Recognition*, Barcelona, Spain, 2009, pp. 606–610.

[8] F.-F. Li and P. Perona, “A bayesian hierarchical model for learning natural scene categories,” in *CVPR ’05: Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’05) - Volume 2*, Washington, DC, USA, 2005, pp. 2:524–531.

[9] D. G. Lowe, “Distinctive image features from scale-invariant keypoints,” *Int. J. Comput. Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[10] S. Belongie, J. Malik, and J. Puzicha, “Shape matching and object recognition using shape contexts,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 4, pp. 509–522, 2002.

[11] G. Agam, S. Argamon, O. Frieder, D. Grossman, and D. Lewis, *The IIT Complex Document Image Processing (CDIP) Test Collection Project*, Illinois Institute of Technology, 2006. [Online]. Available: http://ir.iit.edu/projects/CDIP.html

[12] D. Lewis, G. Agam, S. Argamon, O. Frieder, D. Grossman, and J.Heard, “Building a test collection for complex document information processing,” in *Proc. 29th Int. ACM SIGIR Conference*, 2006, pp. 665–666.

[13] *The Legacy Tobacco Document Library (LTDL)*, University of California, San Francisco, 2007. [Online]. Available: http://legacy.library.ucsf.edu/