2009 10th International Conference on Document Analysis and Recognition

# Logo Matching for Document Image Retrieval

Guangyu Zhu and David Doermann

University of Maryland, College Park, MD 20742, USA
{zhugy, doermann}@umiacs.umd.edu

## Abstract

*Graphics detection and recognition are fundamental research problems in document image analysis and retrieval. As one of the most pervasive graphical elements in business and government documents, logos may enable immediate identification of organizational entities and serve extensively as a declaration of a document's source and ownership. In this work, we developed an automatic logo-based document image retrieval system that handles: 1) Logo detection and segmentation by boosting a cascade of classifiers across multiple image scales; and 2) Logo matching using translation, scale, and rotation invariant shape descriptors and matching algorithms. Our approach is segmentation free and layout independent and we address logo retrieval in an unconstrained setting of 2-D feature point matching. Finally, we quantitatively evaluate the effectiveness of our approach using large collections of real-world complex document images.*

## 1. Introduction

Logos are often used pervasively as declaration of document source and ownership in business and government documents. The problem of logo detection and recognition is of great interest to the document image analysis and retrieval communities because it enables immediate identification of the source of documents based on the originating organization. Facing continually increasing volumes of documents, detecting and recognizing unique, evidentiary graphical symbols, such as logos [15] and signatures [18], is a practical and reliable supplement to the recognition of printed text using OCR and analysis of text by natural language processing. In the context of document image retrieval, logos provide an important form of indexing that enables effective exploration of data.

In the following sections, we first motivate the problems of logo detection, segmentation, and matching for document image retrieval. We then present our approach to graphics recognition based on translation, scale, and rotation-invariant shape descriptors and matching algorithms for generic 2-D feature points, with a focus on the logo matching problem.

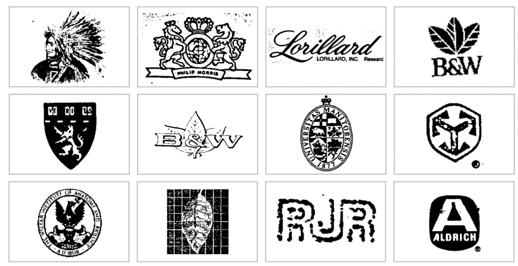


**Figure 1:** Examples of detected and segmented logos from the Tobacco-800 document image database [1, 16].

## 2. Related Work

Prior literature has focused almost exclusively on logo recognition [5, 8, 9, 12, 13]. These studies assume that an effective logo detection and segmentation approach is available. Recognition results are largely reported on the University of Maryland (UMD) Logo Database [7], which contains 105 distinct grayscale logo images. The UMD logo database, however, is far from a perfect recognition benchmark, because it contains only one logo instance per class. Some approaches were evaluated based on the task of group membership recognition (*e.g.* 6 classes in [13]) or subsets of the database (*e.g.* 20 logo classes in [5]), while others included their own logo collections [9, 12]. Furthermore, these approaches generated rotated, noise corrupted, or manually edited logos as test sets using different schemes, making direct comparison difficult.

A fundamental problem in the recognition of graphical symbols is the lack of a general representation based on generic, geometrically invariant features. Doermann *et al*. [8] extracts text and primitive shapes (lines, circles, and rectangles) from logos using many specific feature detectors, and use global and local geometric invariants for matching. Neumann *et al*. [12] uses projection profiles, normalized centroid distance, eccentricity, and various density features for logo recognition. These approaches have limitations. First, it is difficult to robustly extract high-level features (*e.g.* graphical, inverse, or circular text) in a geometrically invariant manner under diverse image qualities and degradations. Second, these methods are hard to extend because they are based on a collection of handpicked and trainable features and a variety of decision rules.

978-0-7695-3725-2/09 $25.00 © 2009 IEEE
DOI 10.1109/ICDAR.2009.60

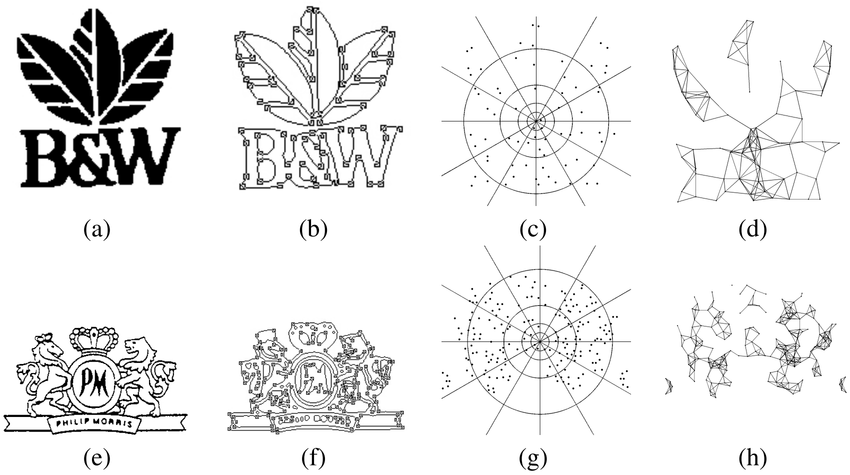

**Figure 2:** Shape contexts [2] and neighborhood graphs [14] constructed from corner feature points. First column: Examples of logos. Second column: detected corners marked on edge images. Third column: Shape contexts descriptors constructed at a point, which provides a large-scale shape description. Fourth column: Neighborhood graphs capture local structures for non-rigid shape matching.

# 3. Logo Detection and Segmentation

Detecting and segmenting free-form graphical patterns such as logos is challenging. Large variations in logo style (see Fig. 1) and low quality images can make detection difficult. Complicating matters, the foreground content of documents generally includes a mixture of machine printed text, diagrams, tables and other elements. From the application perspective, accurate localization is needed for logo recognition. Logo detector must consistently detect and extract complete logos while attempting to minimize the false alarm rate.

We extend our previous logo detection and segmentation approach [15], by incorporating a two-step, partially supervised learning framework that effectively deals with large variations. We learn the base detector—a Fisher classifier at a coarse image scale, from a small set of segmented images and test on a larger pool of unlabeled training images. We then bootstrap these detections to boost a cascade of classifiers at finer image scales, which allows false alarms to be quickly rejected and the detected logo to be more precisely localized. Our logo detection approach is segmentation free and layout independent. Interested readers can refer to [15] for details. Fig. 1 shows detected and segmented logos by our approach from the Tobacco-800 document image database [1, 16].

# 4. Matching and Retrieval

## 4.1 Overview

Given a query logo instance and a database of detected logos, our goal of logo matching is to compute an effective ranked list for logos in the database. By constructing the list of best matching logos, we effectively retrieve the set of documents from the same organizational entities.

We treat a logo as a non-rigid shape, and represent it by a discrete set of 2-D feature points extracted from the object. 2-D point features offer several advantages compared to other compact geometrical entities used in shape representation, because it relaxes the strong assumption that the topology and the temporal order of features are well preserved under image transformations and degradations. For instance, the same portion of contours in one logo sample may overlap, while appearing separated in other cases. Represented by a 2-D point distribution, a shape is more robust under image degradations and noise, while carrying discriminative shape information. As shown in Fig. 2, the shape of a logo is well captured by a finite set $\mathcal{P} = \{P_1, \ldots, P_n\}, P_i \in \mathbb{R}^2$, of $n$ corner feature points computed from the edge image.

We use two state-of-the-art shape matching algorithms for logo matching. The first method is based on the representation of shape contexts, introduced by Belongie *et al*. [2]. In this approach, a spatial histogram defined as *shape context* is computed for each point, which describes the distribution of the relative positions of all remaining points (see column 3 in Fig. 2). Prior to matching, the correspondences between points are solved first through weighted bipartite graph matching. Our second method uses the neighborhood graph matching algorithm by Zheng and Doermann [14], which formulates shape matching as an optimization problem that preserves local structures (see column 4 in Fig. 2). This approach has an intuitive graph matching interpretation, where each point represents a vertex and two vertices are considered connected in the graph if they are neighbors. The problem of finding the optimal match between shapes is thus equivalent to maximizing the number of matched edges between their corresponding graphs under a one-to-one matching constraint. Computationally, neighborhood graphs employ an iterative framework for estimating the correspondences and the transformation. In each iteration, graph matching is initialized us-

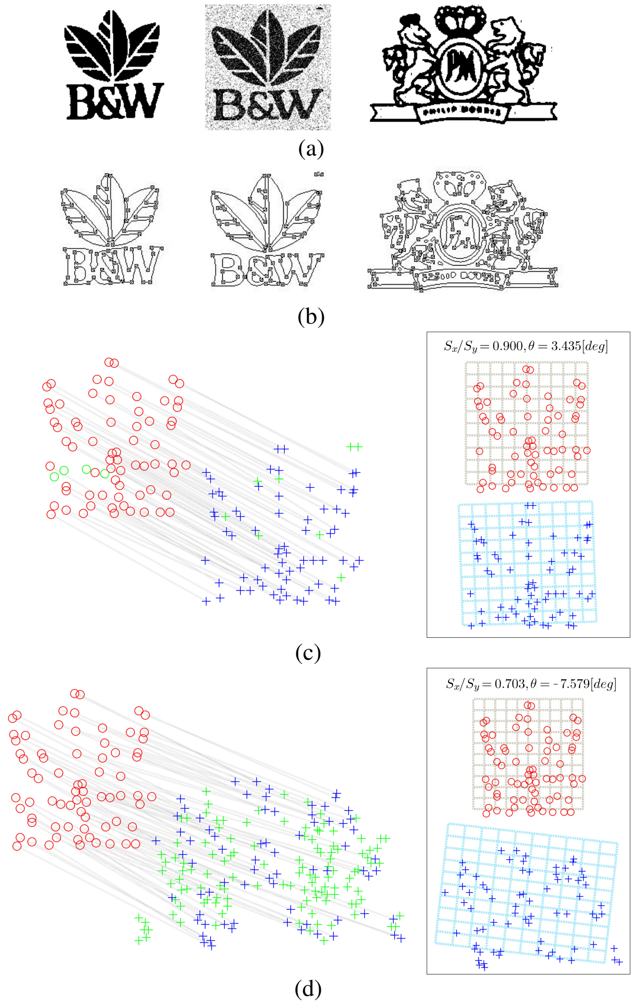


**Figure 3:** Anisotropic scaling and registration quality effectively capture shape differences. (a) Detected logos. (b) Extracted corners. (c) Matching results of first two logos using shape contexts. (d) Matching results of first and third logos using shape contexts. Corresponding points identified by shape matching are linked and unmatched points are shown in green. The computed affine maps are shown in figure legends.

ing the shape context distance [2], and subsequently updated through relaxation labeling for more globally consistent results.

Treating graphics and symbols as 2-D point distributions broadens the space of dissimilarity metrics and enables effective shape matching based on the correspondences and the underlying transformations [19]. We introduce shape dissimilarity metrics that quantitatively measure anisotropic scaling and registration residual error, and present a supervised training framework for effectively combining complementary shape information from different dissimilarity measures by linear discriminant analysis (LDA).

### 4.2 Feature Selection and Extraction

Extracting robust and generic features that can be detected reliably is essential for matching as logos often appear as complex mixtures of graphics and formatted text. We extract corner features from detected logos as follows. We first extract the object contours from the edge image computed by the Canny edge detector [4] and fill in the gaps along the contours. We then use the corner detector of He and Yung [10]. It has shown excellent performance in applications involving real-world scenes compared to other popular feature detectors. It identifies an initial set of corner candidates from local curvature maxima and uses adaptive local thresholds and dynamic support regions to eliminate false corners. Fig. 3(b) shows extracted corners from detected and segmented logos in real document images.

### 4.3 Measures of Shape Dissimilarity

Several measures of shape dissimilarity have demonstrated success in object recognition and retrieval. One is the thin-plate spline bending energy $D_{be}$, and another is the shape context distance $D_{sc}$.

As a conventional tool for interpolating coordinate mappings from $\mathbb{R}^2$ to $\mathbb{R}^2$ based on point constraints, the thin-plate spline (TPS) is commonly used as a generic representation of non-rigid transformation [3]. The TPS bending energy $D_{be}$ [6] measures the amount of *non-linear* deformation to best warp the shapes into alignment. However, $D_{be}$ only measures the deformation beyond an affine transformation, and its functional is zero if the undergoing transformation is purely affine.

The shape context distance $D_{sc}$ between a template shape $\mathcal{T}$ composed of $m$ points and a deformed shape $\mathcal{D}$ of $n$ points is defined in [2] as

$$D_{sc}(\mathcal{T},\mathcal{D}) = \frac{1}{m}\sum_{t\in\mathcal{T}} \arg\min_{d\in\mathcal{D}} C(T(t),d) + \frac{1}{n}\sum_{d\in\mathcal{D}} \arg\min_{t\in\mathcal{T}} C(T(t),d), \tag{1}$$

where $T(.)$ denotes the estimated TPS transformation and $C(.,.)$ is the cost function for assigning correspondence between any two points. Given two points, $t$ in shape $\mathcal{T}$ and $d$ in shape $\mathcal{D}$, with associated shape contexts $h_t(k)$ and $h_d(k)$, for $k = 1, 2, \ldots, K$, respectively, $C(t,d)$ is defined using the $\chi^2$ statistic as

$$C(t,d) \equiv \frac{1}{2}\sum_{k=1}^{K} \frac{[h_t(k) - h_d(k)]^2}{h_t(k) - h_d(k)}. \tag{2}$$

We introduce two new measures of shape dissimilarity and use them as signals for computing ranked list in retrieval. Each dissimilarity measure captures certain shape information from estimated correspondences and transformation. We describe how to effectively combine these measures with limited supervised training in the next subsection.

Our first new measure of dissimilarity $D_{as}$ characterizes the amount of anisotropic scaling between two shapes. Anisotropic scaling is a form of affine transformation that involves change to the relative directional scaling [19]. As illustrated in Fig. 3, the stretching or squeezing of the scale in the computed affine map captures global mismatch in shape dimensions among all registered points, even in the presence of large intra-class variation.

We compute the amount of anisotropic scaling between two shapes by estimating the ratio of the two scaling factors $S_x$ and $S_y$ in the $x$ and $y$ directions, respectively. A TPS transformation can be decomposed into a linear part corresponding to a global affine alignment, together with the superposition of independent, affine-free deformations (or principal warps) of progressively smaller scales [3]. We ignore the non-affine terms in the TPS interpolant when estimating $S_x$ and $S_y$. The 2-D affine transformation is represented as a $2 \times 2$ linear transformation matrix $\mathbf{A}$ and a $2 \times 1$ translation vector $\mathbf{T}$

$$\begin{pmatrix} u \\ v \end{pmatrix} = \mathbf{A} \begin{pmatrix} x \\ y \end{pmatrix} + \mathbf{T}. \tag{3}$$

We can compute $S_x$ and $S_y$ by singular value decomposition on matrix $\mathbf{A}$.

We define $D_{as}$ as

$$D_{as} = \log \frac{\max{(S_x, S_y)}}{\min{(S_x, S_y)}}. \tag{4}$$

Note that we have $D_{as}$ = 0 when only isotropic scaling is involved (*i.e.*, $S_x = S_y$).

We propose another distance measure $D_{re}$ based on the registration residual errors under the estimated non-rigid transformation. To minimize the effect of outliers, we compute the registration residual error from the subset of points that have been assigned correspondence during matching, and ignore points matched to the dummy point $nil$. Let function $M: \mathbb{Z}^+ \to \mathbb{Z}^+$ define the matching between two point sets of size $n$ representing the template shape $\mathcal{T}$ and the deformed shape $\mathcal{D}$. Suppose $t_i$ and $d_{M(i)}$ for $i = 1, 2, \ldots, n$ denote pairs of matched points in shape $\mathcal{T}$ and shape $\mathcal{D}$, respectively. We define $D_{re}$ as

$$D_{re} = \frac{\sum_{i:M(i)\neq nil} ||T(t_i) - d_{M(i)}||}{\sum_{i:M(i)\neq nil} 1}, \tag{5}$$

where $T(.)$ denotes the estimated TPS transformation and $||.||$ is the Euclidean norm.

## 4.4 Shape Distance

After matching, we compute the overall shape distance as the weighted sum of individual distances given by all the measures [17]: shape context distance, TPS bending energy, anisotropic scaling, registration residual errors, and the number of unmatched points.

$$D = w_{sc}D_{sc} + w_{be}D_{be} + w_{as}D_{as} + w_{re}D_{re} + w_{um}D_{um}. \tag{6}$$

The weights in (6) are optimized by linear discriminant analysis using only a small amount of training data.

# 5. Experiments

## 5.1 Baseline Technique

For comparison, we developed a baseline matching approach by computing normalized 2-D cross-correlation between two logos after dimension scaling and rotation correction. The cross-correlation $D_{cc}$ of a query logo $\mathcal{Q}$ with a search logo $\mathcal{P}$ is

$$D_{cc}(\mathcal{Q}, \mathcal{P}) = \frac{1}{n-1} \sum_{x,y} \frac{(q_{x,y} - \bar{q})(p_{x,y} - \bar{p})}{\sigma_q \sigma_p}, \tag{7}$$

where $n$ is the number of pixels.

## 5.2 Evaluation Metrics

We use two most commonly cited measures, average precision and R-precision, to evaluate the performance of each ranked retrieval. Average precision (AP) rewards retrieval systems that rank relevant documents higher, and at the same time penalizes those that rank irrelevant ones higher. R-precision (RP) de-emphasizes the exact ranking among the retrieved relevant documents and is more useful when there are a large number of relevant documents. The overall system performance across all queries are computed quantitatively in mean average precision (MAP) and mean R-precision (MRP), respectively.

## 5.3 Dataset

We demonstrate performance using the $1,290$-image Tobacco-800 database [1, 16]. Tobacco-800 is a public subset of the IIT CDIP Test Collection and has been used in TREC 2006 and 2007 evaluations [1]. It is a realistic, complex dataset for document analysis and retrieval, because these documents were collected and scanned using a wide variety of equipment over time [11]. The image resolutions range from 150 to 300 DPIs and their qualities vary considerably. The Tobacco-800 collection and its associated groundtruth is available in XML format at [16]. We tested our system using a total of 386 logos across 35 classes detected from the Tobacco-800 dataset, among which the number of logos per class varies in the range from 3 to 52.

**Table 1:** Quantitative comparison of retrieval performances.

| Approach (Measure of Dissimilarity) | MAP | MRP |
|---|---|---|
| Correlation with scale and rotation corrections ($D_{cc}$) | $42.5\%$ | $38.2\%$ |
| Neighborhood graphs ($D_{sc} + D_{be}$) | $63.1\%$ | $59.3\%$ |
| Neighborhood graphs ($D_{sc} + D_{be} + D_{as} + D_{re} + D_{um}$) | $75.5\%$ | $70.8\%$ |
| Shape contexts ($D_{sc} + D_{be}$) | $69.7\%$ | $65.3\%$ |
| Shape contexts ($D_{sc} + D_{be} + D_{as} + D_{re} + D_{um}$) | $82.6\%$ | $78.5\%$ |

### 5.4 Results and Discussion

Table 1 summaries the performances of different matching algorithms in combination with different measures of shape dissimilarity. Both neighborhood graphs and shape contexts significantly outperform the correlation method. This demonstrates the competitive advantages of approaches based on 2-D feature matching in the recognition of graphics and symbols. First, their shape descriptors are built from generic 2-D point distribution, which can be robustly extracted in practice. Second, these approaches solve the underlying transformations (affine for linear and TPS for non-linear transformation), which improves shape matching and discrimination.

Shape contexts method gives the best logo matching performance as shown in Table 1. By incorporating rich global shape information, shape contexts descriptors are more robust under significant image degradations than neighborhood graphs, which capture local structures.

Shape dissimilarity measures computed from anisotropic scaling, registration residual error, and the number of unmatched points significantly improve the retrieval performance, demonstrating that we can improve the retrieval quality considerably by combining complementary measures of shape dissimilarity. In addition, this experiment shows the effectiveness of learning the optimal weight associated with different dissimilarity metrics using LDA under limited supervised training.

## 6. Conclusion

In this paper, we have presented an approach to automatically detecting, segmenting, and matching logos from documents with unconstrained layouts and complex background for document retrieval. To robustly handle variety of image qualities and degradations, we treated the logo in the unconstrained setting of a non-rigid shape and demonstrated document image retrieval using state-of-the-art shape representations, measures of shape dissimilarity, and shape matching algorithms. We quantitatively evaluated the effectiveness of our approach in challenging retrieval tests using public, real-world document image collections involving a large number of classes but relatively small numbers of logo instances per class.


## Acknowledgements

The partial support of this research by DARPA through BBN/DARPA award HR001108C0004 and the US Government through NSF Award 1150713501 is gratefully acknowledged.


## References


[1] G. Agam, S. Argamon, O. Frieder, D. Grossman, and D. Lewis. The Complex Document Image Processing Test Collection. Online, 2006. `http://ir.iit.edu/projects/CDIP.html`.

[2] S. Belongie, J. Malik, and J. Puzicha. Shape matching and object recognition using shape contexts. *IEEE Trans. Pattern Anal. and Machine Intell.*, 24(4):509–522, 2002.

[3] F. Bookstein. Principle warps: Thin-plate splines and the decomposition of deformations. *IEEE Trans. Pattern Anal. and Machine Intell.*, 11(6):567–585, 1989.

[4] J. Canny. A computational approach to edge detection. *IEEE Trans. Pattern Anal. and Machine Intell.*, 8(6):679–697, 1986.

[5] J. Chen, M. K. Leung, and Y. Gao. Noisy logo recognition using line segment Hausdorff distance. *Pattern Recognition*, 36(4):943–955, 2003.

[6] H. Chui and A. Rangarajan. A new point matching algorithm for non-rigid registration. *Computer Vision and Image Understanding*, 89(2-3):114–141, 2003.

[7] D. Doermann. The University of Maryland Logo Database. Online, 2008. `http://lampsrv01.umiacs.umd.edu/projdb/project.php?id=47`.

[8] D. Doermann, E. Rivlin, and I. Weiss. Applying algebraic and differential invariants for logo recognition. *Machine Vision and Application*, 9(2):73–86, 1996.

[9] M. Gori, M. Maggini, S. Marinai, J. Q. Sheng, and G. Soda. Edge-backpropagation for noisy logo recognition. *Pattern Recognition*, 36(1):103–110, 2003.

[10] X. C. He and N. H. C. Yung. Corner detector based on global and local curvature properties. *Optical Engineering*, 47(5):057008–1–12, 2008.

[11] D. Lewis, G. Agam, S. Argamon, O. Frieder, D. Grossman, and J. Heard. Building a test collection for complex document information processing. In *Proc. ACM SIGIR Conf.*, pages 665–666, 2006.

[12] J. Neumann, H. Samet, and A. Soffer. Integration of local and global shape analysis for logo classification. *Pattern Recognition Letters*, 23(12):1449–1457, 2002.

[13] T. D. Pham. Variogram-based feature extraction for neural-network recognition of logos. In *Proc. Applications of Artificial Neural Networks in Image Processing*, pages 22–29, 2003.

[14] Y. Zheng and D. Doermann. Robust point matching for non-rigid shapes by preserving local neighborhood structures. *IEEE Trans. Pattern Anal. and Machine Intell.*, 28(4):643–649, 2006.

[15] G. Zhu and D. Doermann. Automatic document logo detection. In *Proc. Int'l Conf. Document Analysis and Recognition*, pages 864–868, 2007.

[16] G. Zhu and D. Doermann. Tobacco-800 Complex Document Image Database and Groundtruth. Online, 2008. `http://lampsrv01.umiacs.umd.edu/projdb/edit/project.php?id=52`.

[17] G. Zhu, Y. Zheng, and D. Doermann. Signature-based document image retrieval. In *Proc. European Conf. Computer Vision*, volume 3, pages 752–765, 2008.

[18] G. Zhu, Y. Zheng, D. Doermann, and S. Jaeger. Multi-scale structural saliency for signature detection. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pages 1–8, 2007.

[19] G. Zhu, Y. Zheng, D. Doermann, and S. Jaeger. Signature detection and matching for document image retrieval. *IEEE Trans. Pattern Anal. and Machine Intell.*, 2009. Preprint Online, `http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=4633365&isnumber=4359286`.