Logo recognition based on the Dempster-Shafer Fusion of Multiple Classifiers

Mohammad Ali Bagheri[1], Qigang Gao[1], and Sergio Escalera[2]

[1] Faculty of Computer Science, Dalhousie University, Halifax, Canada
[2] Computer Vision Center, Campus UAB, Edifici O, 08193, Bellaterra, Spain
Dept. Matemtica Aplicada i Anlisi, Universitat de Barcelona, Gran Via de les Corts Catalanes 585, 08007, Barcelona, Spain.

Abstract. The performance of different feature extraction and shape description methods in trademark image recognition systems have been studied by several researchers. However, the potential improvement in classification through feature fusion by ensemble-based methods has remained unattended. In this work, we evaluate the performance of an ensemble of three classifiers, each trained on different feature sets. Three promising shape description techniques, including Zernike moments, generic Fourier descriptors, and shape signature are used to extract informative features from logo images, and each set of features is fed into an individual classifier. In order to reduce recognition error, a powerful combination strategy based on the Dempster-Shafer theory is utilized to fuse the three classifiers trained on different sources of information. This combination strategy can effectively make use of diversity of base learners generated with different set of features. The recognition results of the individual classifiers are compared with those obtained from fusing the classifiers' output, showing significant performance improvements of the proposed methodology.

Keywords: Logo recognition, ensemble classification, Dempster-Shafer fusion, Zernike moments, generic Fourier descriptor, shape signature

1 Introduction

The research of document image processing has received great attention in recent years because of its diverse applications, such as digital libraries, online shopping, and office automation systems. An important problem in the field of document image processing is the recognition of graphical items, such as trademarks and company logos. Logos are mainly used by companies and organizations to identify themselves on documents. Given an image segment from a document image and a logo database, the task of logo recognition is to find whether the image segment corresponds to a logo in the database. The successful recognition of logos facilitates automatic classification of source documents, which is considered a key strategy for document image analysis and retrieval.

Logo analysis in document images involves two main steps: (1) detecting the probable logo from a document image; (2) classifying the detected logo candidate segment into one of the learned logos in the database [25]. The first step is referred to as logo detection, while the second is usually called logo recognition. In this work, we focus on the logo recognition phase.

From the machine learning point of view, logo recognition is considered a multi-class classification task since each logo category is considered a separate target class. In this view, the classification system involves two main stages: the selection and/or extraction of informative features and the construction of a classification algorithm. In such a system, a desirable feature set can greatly simplify the construction of a classification algorithm, and a powerful classification algorithm can work well even with a low discriminative feature set.

In the last decade, active research has been conducted on logo recognition. Most of the research work has focused on providing a framework for logo recognition by the extraction of informative features [8] or the analysis of image structures [1]. The classification algorithm is usually used as a black box tool.

In this work, we aim to enhance the recognition efficiency of logo images by augmenting the classification stage. Here, we evaluate the performance of an ensemble of three classifiers, each trained on different feature sets extracted from three shape description techniques. Three promising shape desciptors, including Zernike moments, generic Fourier descriptors (GFD), and shape signature based on centroid distance are used to extract an informative set of features from logo images. Then, each set of features is fed into a base classifier and fused by the Demspter-Shafer based combination method. The classification results of the individual classifiers are compared with those obtained from fusing the classifiers' output. The experimental results show that this strategic combination of shape description techniques can significantly improve the recognition accuracy.

The contribution of this work is two-fold: (1) the application of the ensemble approach to address a challenging image vision classification problem; (2) improving the recognition performance by utilizing a combination strategy that is appropriate for fusing different sources of information. This strategy can effectively make use of diversity of base classifiers trained on different set of features.

The rest of this work is organized as follows: Section 2 first provides a brief review of the ensemble classification approaches and then explains the Dempster-Shafer fusion of ensemble classifiers. In Section 3, the proposed logo classification framework is explained in detail. The experimental results on a well-know logo dataset are reported in Section 4. Finally, Section 5 states the conclusions of the paper.

2 Multiple Classifier Systems

Combining multiple classifiers to achieve higher accuracy is one of the most active research areas in the machine learning community [7]. It is known under various names, such as multiple classifier systems, classifier ensemble, committee of classifiers, mixture of experts, and classifier fusion. Multiple classifier systems can generate more accurate classification results than each of the individual classifiers [22]. In such systems, the classification task can be solved by integrating different classifiers, leading to better performance. However, the ensemble approach depends on the assumption that single classifiers' errors are uncorrelated, which is known as classifier diversity. The intuition is that if each classifier makes different errors, then the total errors can be reduced by an appropriate combination of these classifiers.

The design process of a multiple classifier system generally involves two steps [22]: the collection of an ensemble of classifiers and the design of the combination rule. These steps are explained in detail in the next subsections.

2.1 Creating an ensemble of classifiers

There are three general approaches to creating an ensemble of classifiers in state-of-the-art research, which can be considered as different ways to achieve diversity. The most straightforward approach is using different learning algorithms for the base classifiers or variations of the parameters of the base classifiers e.g. different initial weights or different topologies of a series of neural network classifiers. Another approach, which has been getting more attention in the related literature, is to use different training sets to train base classifiers. Such sets are often obtained from the original training set by resampling techniques, such as the procedures presented in Bagging and AdaBoost [10].

The third approach, which is employed in this work for classification of logo images, is to train the individual classifiers with datasets that consist of different feature subsets, or so-called ensemble feature selection [21]. While traditional feature selection algorithms seek to find an optimal subset of features, the goal of ensemble feature selection is to find different feature subsets to generate accurate and diverse classifiers. The Random subspace method (RMS) proposed by Hu in [12] is one early algorithm that builds an ensemble by randomly choosing the feature subsets. More recently, different techniques based on this approach have been proposed.

2.2 Design of a combination rule

Once a set of classifiers are generated, the next step is to construct a combination function to merge their outputs, which is also called decision optimization. The most straightforward strategy is the simple majority voting, in which each classifier votes on the class it predicts, and the class receiving the largest number of votes is the ensemble decision. Other strategies for combination function include weighted majority voting, sum, product, maximum and minimum, fuzzy integral, decision templates, and the Dempster-Shafer (DS) based combiner [16],[17].

Inspired by the Dempster-Shafer (DS) theory of evidence [6], a combination method is proposed in [24], which is commonly known as the Dempster-Shafer fusion method. By interpreting the output of a classifier as a measure of evidence provided by the source that generated the training data, the DS method fuses an ensemble of classifiers. Here, we skip the details of how this originated from DS theory and will explain the DS fusion algorithm in the following subsection.

Dempster-Shafer fusion method Let $x \in R^n$ be a feature vector and $\Omega = \{\omega_1, \omega_2, \ldots, \omega_c\}$ be the set of class labels. Each classifier h_i in the ensemble $H = \{h_1, h_2, \ldots, h_L\}$ outputs c degrees of support. Without loss of generality, we can assume that all c degrees are in the interval $[0, 1]$. The support that classifier h_i, gives to the hypothesis that $\mathbf{x}$ comes from class ω_j is denoted by $d_{i,j}(x)$. Clearly, the larger the support, the more likely the class label ω_j. The L classifier outputs for a particular instance $\mathbf{x}$ can be organized in a decision profile, $DP(x)$, as the following matrix [17]:

$$DP(x) = \begin{pmatrix} d_{1,1}(x) & \cdots & d_{1,j}(x) & \cdots & d_{1,c}(x) \\ \vdots & & \vdots & & \vdots \\ d_{i,1}(x) & \cdots & d_{i,j}(x) & \cdots & d_{i,c}(x) \\ \vdots & & \vdots & & \vdots \\ d_{L,1}(x) & \cdots & d_{L,j}(x) & \cdots & d_{L,c}(x) \end{pmatrix}$$

The Dempster-Shafer fusion method uses decision profile to find the overall support for each class and subsequently labels the instance $\mathbf{x}$ in the class with the largest support. In order to obtain the ensemble decision based on DS fusion method, first, the c decision templates, $DT_1, \ldots, DT_c$, are built from the training data. Roughly speaking, decision templates are the most typical decision profile for each class ω_j. For each test sample, $\mathbf{x}$, the DS method compare the decision profile, $DP(x)$, with decision templates. The closest match will label $\mathbf{x}$. In order to predict the target class of each test sample, the following steps are performed [17][24]:

1. Build decision templates: For $j = 1, \ldots, c$, calculate the means of the decision profiles for all training samples belonging to ω_j. Call the mean a decision template of class ω_j, DT_j.

$$DT_j = \frac{1}{N_j} \sum_{z_k \in \omega_j} DP(z_k) \tag{1}$$

where N_j in the number of training samples belong to ω_j.

2. Calculate the proximity: Let DT_j^i denote the ith row of the decision template DT_j, and D_i the output of the ith classifier, that is, the ith row of the decision profile $DP(x)$. Instead of similarity, we now calculate proximity Φ, between DT_j^i and the output of classifier D_i for the test sample x:

$$\Phi_{j,i}(x) = \frac{(1 + \|DT_j^i - D_i(x)\|)^{-1}}{\sum_{k=1}^{c}(1 + \|DT_j^i - D_i(x)\|)^{-1}} \tag{2}$$

where $\|.\|$ is a matrix norm.

3. Compute belief degrees: Using Eq. (2), calculate for each class $j = 1, \ldots, c$

and for each classifier $i = 1, \ldots, L$, the following belief degrees, or evidence, that the ith classifier is correctly identifying sample $\mathbf{x}$ into class ω_j:

$$b_j(D_i(x)) \frac{\Phi_{j,i}(x) \prod_{k \neq j}(1 - \Phi_{k,i}(x))}{1 - \Phi_{j,i}(x)[1 - \prod_{k \neq j}(1 - \Phi_{k,i}(x))]} \tag{3}$$

4. Final decision based on class support: Once the belief degrees are achieved for each source (classifier), they can be combined by Dempster's rule of combination, which simply states that the evidences (belief degree) from each source should be multiplied to obtain the final support for each class:

$$\mu_j(x) = K \prod_{i=1} b_j(D_i(x)), \ j = 1, \ldots, c \tag{4}$$

where K is a normalizing constant ensuring that the total support for ω_j from all classifiers is 1. The DS combiner gives a preference to class with largest $\mu_j(x)$.

3 Framework of the proposed logo recognition system

As mentioned earlier, this work focuses on the second step of logo analysis: logo recognition. The problems of image segmentation and logo detection are beyond the scope of this work. Figure 1 shows the framework of our logo recognition system. In the followings, we describe the main phases of the framework.

The logo image database we used is the MPEG7 dataset [3] . This dataset consists of $C = 70$ classes with 20 instances per class, which represents a total of 1400 object images. Figure 2 shows a few of samples for some categories of this dataset. This dataset has been widely used as the benchmark dataset for logo classification and retrieval [9], [19], [23].

3.1 Preprocessing

An effective classification system should be invariant to the translation, rotation, and scaling (TRS) of logo images. Generally, there are two approaches to achieve the invariance property. The first one is to use shape descriptors that are naturally invariant to TRS. The second approach is to employ some preprocessing steps before using shape description techniques in order to provide TRS invariance.

Here, we used three shape description techniques: shape signature based on centroid distance, Zernike moments, and generic Fourier descriptor (GFD). The shape signature descriptor has the desirable properties of translation, rotation, and scaling invariance. However, the Zernike moments are invariant only to the rotation, and are not invariant to scaling and translation. Similarly, generic

[3] MPEG7 Repository dataset: http://www.cis.temple.edu/ latecki/

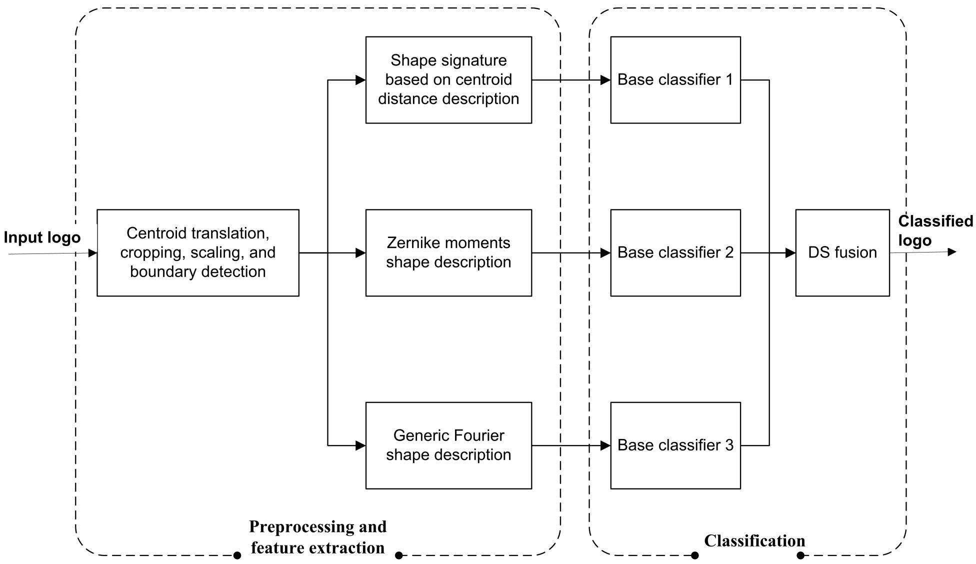

Fig. 1. The framework of the proposed logo classification system based on the Dempster-Shafer fusion of multiple classifiers

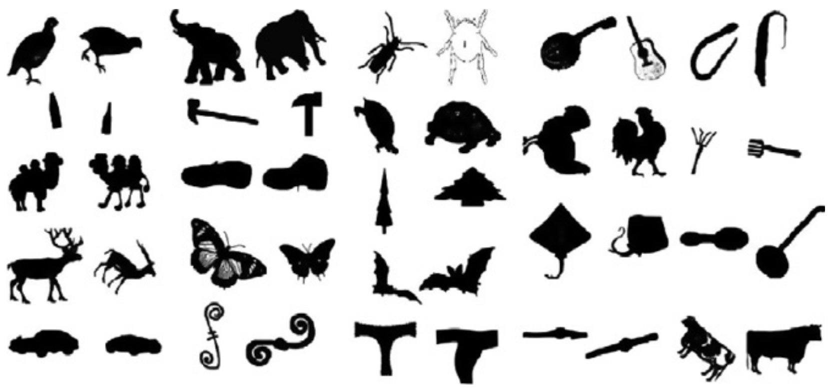

Fig. 2. Some examples of labeled images in the MPEG7 dataset.

Fourier descriptor is not natural translation invariant. Therefore, for effective usage of Zernike moments and generic Fourier descriptor in the logo classification framework, the input images need to be normalized for scale and translation.

In the preprocessing phase, translation invariance is achieved by finding the geometrical centroid,(x_0, y_0), of the image and shifting the origin to the centroid of every image. For scale invariance, we create a circular image by superposing a circle centred at the geometrical centroid, with a radius equal to the distance between the centroid and the outermost pixel of the logo image. Finally, we scale the circular image to a square of size 256×256 pixels.

3.2 Feature extraction of logo images

Some researchers have studied the problem of logo recognition by applying different feature extraction methods such as algebraic and differential invariants [8],[14], edge direction histogram [4],[14], Zernike and pseudo-Zernike moments [15],[18], string-matching techniques [5], template matching [14], and wavelet features [20].

In this work, we employed three different image description techniques:

- **Shape signatures based on centroid distance:** A shape signature, $z(u)$, is a 1-D function representing 2-D areas or boundaries, which can be a unique descriptor of a shape. Shape signatures are mostly used as an input vector to the Fourier Descriptor (FD). Zhang and Lu studied different FD methods for image retrieval and showed that FDs derived from centroid distance perform better than FDs derived from other shape signatures in terms of overall performance [27]. Thus, we used centroid distance-based shape signatures in this work.
- **Zernike moments (ZM):** ZMs are observed to outperform many moment-based shape descriptors, such as geometric moments, Legendre moments, and pseudo-ZMs [28]. The superiority of ZMs is mainly due to the fact that their basis functions are orthogonal. Therefore, Zernike moments can describe an image with no redundancy or overlap of information between the moments [13]. Here, logo images are mapped onto a set of complex Zernike polynomials and the first 4-order Zernike moments are computed. The reader is referred to [13] for a more detailed description of the ZM computation.
- **Generic Fourier descriptors (GFD):** The GFD is extracted from spectral domain by applying the 2D Fourier transform on polar-raster sampled shape images [26]. The process of employing GFD is similar to the conventional FD:

$$GFD(\rho, \phi) = \sum_r \sum_i f(r, \theta_i) exp[-j2\pi(\frac{r}{R}\rho + \frac{2\pi i}{T}\phi)]$$

  where $0 \leq r \leq R$ and $\theta_i = i(2\pi/T)(0 \leq i \leq T)$; $0 \leq \rho \leq R$, $0 \leq \phi \leq T$. R and T are the radial and angular resolutions, respectively and $f(x, y)$ is the binary image function [26].

4 Classification results

In this stage, we aim to classify different logo images based on DS fusion of individual classifiers. The classification performance is obtained by means of stratified 10-fold cross-validation over 10 runs. In order to improve the reliability of the results, the experiments are conducted using different numbers of classes, i.e. different numbers of logo categories and using two classification algorithms, including 1) Support Vector Machine with the Gaussian Kernel and 2) Multilayer Perceptron (MLP) with 10 nodes in the hidden layer. For SVM implementation, we use the LIBSVM package (version 3.1) developed by Chang and Lin [3], tuning Kernel parameters via cross-validation.

The summary of the results are reported in Table 1 and Table 2 for SVM and MLP as the base learners. These tables show the classification accuracy of individual classifiers and the one achieved by the Dempster-Shafer fusion of them.

Table 1. Classification accuracy of single and fused classifiers using SVM as the base learner

# classes	**Single classifier trained only on**			**DS fusion**
	GFD	Zernike moments	Shape signature	
10	98.30	98.50	97.20	99.20
20	95.65	96.10	95.20	98.15
30	94.33	93.23	92.30	96.47
40	93.93	93.73	91.65	96.78
50	92.66	91.94	90.36	95.72
60	92.43	92.50	90.30	96.07
70	91.79	91.57	89.36	95.37

Table 2. Classification accuracy of single and fused classifiers using MLP as the base learner

# classes	**Single classifier trained only on**			**DS fusion**
	GFD	Zernike moments	Shape signature	
10	89.50	94.00	90.90	98.70
20	81.65	84.95	79.30	96.60
30	66.80	67.13	63.43	93.57
40	54.03	55.58	53.85	90.55
50	47.98	46.86	45.16	88.44
60	40.97	40.52	39.72	86.30
70	34.59	34.84	35.64	84.91

It is important to note the outperformance of the fused results in comparison with the individual classifier. This improvement is clearer when the number of classes of the datasets is increased. In that case, the inter-class variability is reduced, and thus, it is easier to confuse patterns from different classes.

As an additional analysis, we compare classification results of merging classifiers by different combination methods. Figure 3 and Figure 4 show the classification accuracy of individual classifiers and ensemble systems by different fusion methods. Considered combination methods include fusion by majority voting, maximum, sum, minimum, average, naive-Bayes, and the Dempster-Shafer fusion method.

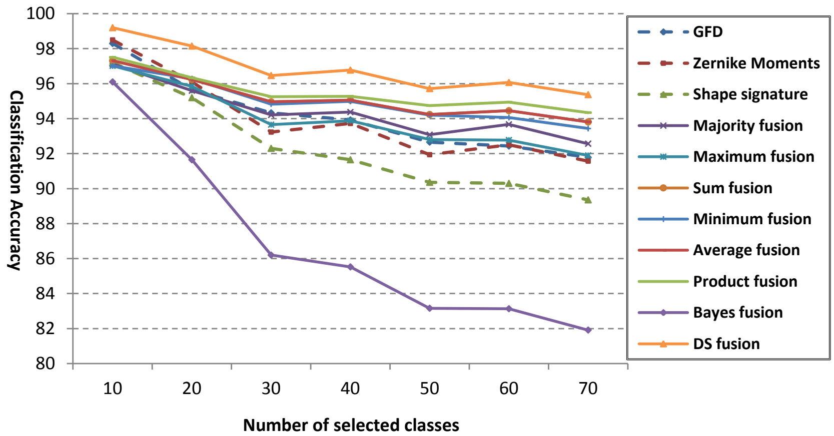

Fig. 3. Classification accuracy of single and fused classifiers by different combination methods using SVM as the base learner

As these figure show, the best classification accuracy is achieved by means of the Dempster-Shafer fusion method. The general results for different numbers of classes are summarized below:

- In these experiments, the three descriptors used show similar performance in terms of classification accuracy on the MPEG-7 dataset.
- The classification results reveal the critical role of the combination method. As Figure 3 and 4 show, only using diverse classifiers is not enough to improve the classification performance of the ensemble system. If the combination method does not properly make use of the ensemble diversity, then no benefit arises from fusing multiple classifiers [2]. For example, the commonly used majority voting combination method does not make significant use of the diversity among ensemble classifiers in these experiments. Therefore, the classification accuracy obtained by fusing the classifiers' outputs can be even worse than the one achieved by single classifiers trained only with one set of

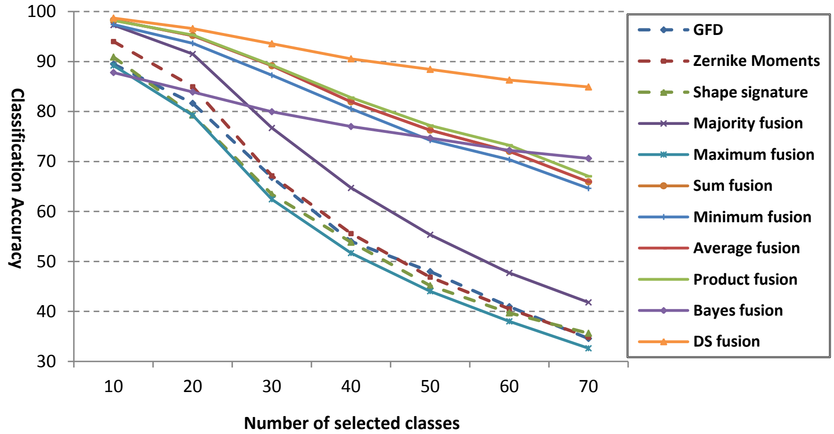

Fig. 4. Classification accuracy of single and fused classifiers by different combination methods using MLP as the base learner

shape descriptors. On the other hand, the Dempster-Shafer fusion method has significantly improved the classification performance.

- The ensemble system, even by using a poor fusion method, generally performs better than the base classifiers. This finding confirms the philosophy of the ensemble systems: combining the outputs of several learners can reduce the risk of an unfortunate selection of a poorly performing learner.
- The MLP classification accuracy of individual classifiers dramatically decreases as the number of classes increase. This finding is mainly due to the fact that MLP classifiers solve the whole multi-class classification problem concurrently. Therefore, it is more difficult to separate a large number of classes. However, the case for SVM is different. The SVM algorithm solves the multi-class problem by decomposing it into several smaller binary problems using the one-versus-one scheme. It has been shown that this approach, known as class binarization, achieves better classification performance compared to the approach that aims to solve the whole multiclass problem at once [11].

5 Conclusions

In this work, we evaluated the performance of an ensemble of three classifiers, each trained on different feature sets. Three efficient shape description methods, including shape signature, Zernike moments, and generic Fourier descriptors, were used to extract informative features from logo images and each set of features was fed into an individual classifier. In order to reduce recognition error, the Dempster-Shafer combination theory was employed to fuse the three classifiers trained on different sources of information. The classification results of the individual classifiers were compared with those obtained from fusing the classifiers by the Dempster-Shafer combination method.

Generally speaking, using ensemble methods for the classification of logo images is effective, though different combination methods would show different performances, and even some combination of base classifiers and ensemble methods would deteriorate the performance of the best single classifier. However, as demonstrated by our experiments, by using the DS fusion method, the classification performance was significantly increased compared with single classifiers trained by a specific set of features.

Acknowledgment

The authors would like to thank Mis. Fatemeh Yazdanpanah for her kind help with conducting experiments.

References

1. Alajlan, N., Kamel, M., Freeman, G.: Multi-object image retrieval based on shape and topology. Signal Processing: Image Communication 21(10), 904–918 (2006)
2. Brodley, A., Lane, T.: Creating and exploiting coverage and diversity. In: Proceedings of AAAI-96 Workshop on Integrating Multiple Learned Models,. pp. 8–14. Portland, OR (1996)
3. Chang, C.C., Lin, C.J.: Libsvm: A library for support vector machines. ACM Transactions on Intelligent Systems and Technology 2(3), 1–27 (2011)
4. Ciocca, G., Schettini, R.: Content-based similarity retrieval of trademarks using relevance feedback. Pattern Recognition 34(8), 1639–1655 (2001)
5. Cortelazzo, G., Mian, G., Vezzi, G., Zamperoni, P.: Trademark shapes description by string-matching techniques. Pattern Recognition 27(8), 1005–1018 (1994)
6. Dempster, A.: Upper and lower probabilities induced by multivalued mappings. Annals of Mathematical Statistics 38(2), 325–339 (1967)
7. Dietterich, T.G.: Machine learning research: Four current directions. Artificial Intell. Mag. 18(4), 97–136 (1997)
8. Doermann, D., Rivlin, E., Weiss, I.: Applying algebraic and differential invariants for logo recognition. Machine Vision and Applications 9(2), 73–86 (1996)
9. Escalera, S., Forns, A., Pujol, O., Llads, J., Radeva, P.: Circular blurred shape model for multiclass symbol recognition. Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on 41(2), 497–506 (2011)
10. Freund, Y., Schapire, R.: A decision-theoretic generalization of on-line learning and an application to boosting. Journal of Computer and System Sciences 55(1), 119–139 (1997)
11. Garca-Pedrajas, N., Ortiz-Boyer, D.: An empirical study of binary classifier fusion methods for multiclass classification. Information Fusion 12(2), 111–130 (2011)
12. Ho, T.K.: The random subspace method for constructing decision forests. IEEE Transactions on Pattem Analysis and Machine Intelligence 20, 832–844 (1998)
13. Hwang, S., Kim, W.: A novel approach to the fast computation of zernike moments. Pattern Recognition 39(11), 2065–2076 (2006)
14. Jain, A., Vailaya, A.: Shape-based retrieval: A case study with trademark image databases. Pattern Recognition 31(9), 1369–1390 (1998)

15. Kim, Y., Kim, W.: Content-based trademark retrieval system using a visually salient feature. Image and Vision Computing 16(12-13), 931–939 (1998)
16. Kittler, J., Hatef, M., Duin, R., Matas, J.: On combining classifiers. IEEE Transactions on Pattern Analysis and Machine Inteligence 20(3), 226–239 (1998)
17. Kuncheva, L.I.: Combining Pattern Classifiers: Methods and Algorithms. Wiley, New York, NY (2004)
18. Li, S., Lee, M., Pun, C.: Complex zernike moments features for shape-based image retrieval. Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on 39(1), 227–237 (2009)
19. Mohd Anuar, F., Setchi, R., Lai, Y.k.: Trademark image retrieval using an integrated shape descriptor. Expert Systems with Applications p. http://dx.doi.org/10.1016/j.eswa.2012.07.031 (2012)
20. Neumann, J., Samet, H., Soffer, A.: Integration of local and global shape analysis for logo classification. Pattern Recognition Letters 23, 1449–1457 (2002)
21. Opitz, D.W.: Feature selection for ensembles (1999)
22. Polikar, R.: Ensemble based systems in decision making. IEEE Circuits and Systems Magazine 6(3), 21–45 (2006)
23. Qi, H., Li, K., Shen, Y., Qu, W.: An effective solution for trademark image retrieval by combining shape description and feature matching. Pattern Recognition 43(6), 2017–2027 (2010)
24. Rogova, G.: Combining the results of several neural network classifiers. Neural Networks 7, 777–781 (1994)
25. Seiden, S., Dillencourt, M., Irani, S., Borrey, R., Murphy, T.: Logo detection in document images. In: International Conference on Imaging Science, Systems, and Technology. pp. 446–449. Las Vegas, Nevada (1997)
26. Zhang, D., Lu, G.: Shape-based image retrieval using generic fourier descriptor. Signal Processing: Image Communication 17(10), 825–848 (2002)
27. Zhang, D., Lu, G.: Study and evaluation of different fourier methods for image retrieval. Image and Vision Computing 23(1), 33–49 (2005)
28. Zhang, D., Lu, G.: Review of shape representation and description techniques. Pattern Recognition 37(1), 1–19 (2004)