

AN EMPIRICAL STUDY OF SHAPE RECOGNITION IN ENSEMBLE LEARNING CONTEXT

WEILI DING¹, XINMING WANG¹, HAN LIU², BO HU¹

¹Laboratory of Pattern Recognition and Intelligent Systems

Key Laboratory of Industrial Computer Control Engineering of Heibei Province, Department of Automation
Institute of Electrical Engineering, Yanshan University, Qinghuangdao, 066004, China

²School of Computer Science and Informatics, Cardiff University

Queen's Buildings, 5 The Parade, Cardiff, CF24 3AA, United Kingdom

E-MAIL: weiy51@ysu.edu.cn, wang8297161@stumail.ysu.edu.cn, liuh48@cardiff.ac.uk, jessical@stumail.ysu.edu.cn

Abstract:

Shape recognition has been a popular application of machine learning, where each shape is defined as a class for training classifiers that recognize the shapes of new instances. Since training of classifiers is essentially achieved through learning from features, it is crucial to extract and select a set of relevant features that can effectively distinguish one class from other classes. However, different instances could present features which are highly dissimilar, even if these instances belong to the same class. The above difference in feature representation can also result in high diversity among classifiers trained by using different algorithms or data samples. In this paper, we investigate the impact of multi-classifier fusion on shape recognition by using six features extracted from a 2D shape data set. In particular, popular single learning algorithms, such as Decision Trees, Support Vector Machine and K Nearest Neighbours, are adopted to train base classifiers on features selected by using a wrapper approach. Furthermore, two popular ensemble learning algorithms (Random Forests and Gradient Boosted Trees) are adopted to train decision tree ensembles on the same feature sets. The outputs of the two ensemble classifiers are finally combined with the outputs of all the other base classifiers. The experimental results show the effectiveness of the above setting of multi-classifier fusion for advancing the performance in comparison with using each single (non-ensemble) learning algorithm.

Keywords:

Shape recognition; Machine learning; Ensemble learning; Multi-classifier fusion; Feature extraction

1. Introduction

Shape recognition, which is a special type of pattern recognition, has become a popular application area of machine learning. In this context, each specific shape is defined as a class for training a classifier that can identify the shape of an object shown in an image. Training of classifiers is essentially achieved through learning from features defined for each specific class. From this point of view, it is important to extract and select a set of features relevant for each class, such that the relationships between features and classes can be identified effectively in the training stage.

Due to the diversity among different objects, instances, which belong to the same class, could even present features that are highly dissimilar. From machine learning perspectives, different features usually show different fitness to different learning algorithms, which would result in high diversity among classifiers trained by using different algorithms or data samples. Therefore, it becomes a motivation to combine different classifiers in the setting of ensemble learning, in order to boost the overall performance of classification.

In this paper, we investigate the impact of multi-classifier fusion on shape recognition. In particular, six features are extracted from a 2D shape data set, and four of the features are selected for training several different classifiers that are fused in the setting of ensemble learning. The experimental results show that our proposed setting of ensemble learning achieves to advance the overall classification performance, despite a very small number of features available for classifiers training.

The rest of this paper is organized as follows: Section 2

presents related work on shape recognition and machine learning algorithms. In section 3, we provide details of feature extraction and describe the design of the multi-classifier fusion framework. In Section 4, we provide details of the experimental setup and discussion of results. In Section 5, the contributions of this paper are summarized and some further directions are suggested for advancing this research area in the future.

2. Related works

Shape recognition [1, 2] has been widely studied in computer vision in the past decades. Methods of combining machine learning with descriptors are increasingly getting popular in this field. In general, the existing shape features can be roughly divided into region based methods and contour based methods.

In region based techniques, moment-based descriptors [3], grid method [5], shape matrix [6], convex hull [2], and media axis [7] features have been proposed. In these features, geometric moments [3] and Zernike moments [4] are widely utilized in region based techniques. The former ones are based on combinations of the lower order moments. Although they have merits, as low computation cost and invariant to rotation, geometric moments show unstable performance on scaled images. The later ones are robust to noise and have less information redundancy with expensive calculation cost. In contour based methods, there are simple shape descriptors, boundary moments, shape signature, stochastic method, generic Fourier descriptor [8], chain code [9], scale space descriptors [10], spectral descriptor [11]. Chain code is sensitive to noise due to the fact that it has high dimensions while Fourier descriptor (FD) and wavelet descriptor (WD) can handle the issue of noise sensitivity in spectral domain. Recently, three novel region shape descriptors, namely, region area descriptor (RAD), region skeleton descriptor (RSD) and simplified shape signature (SSS), which adopted the topological structure of a shape in [12]. A descriptor, which is called Bag-of-Saliency-Points (BoSP) and was also proposed in [13], uses salient points combined with histogram construction using a codebook. Compared to previous works, BoSP is invariance to geometric transformation and efficient to enhance speed of calculation while maintaining the promising performance.

For shape classification, machine learning approaches gain their popularity in this field. Several machine learning algorithms, such as support vector machine (SVM) [14], k nearest neighbours (KNN) [15], naive bayes (NB) [16], decision tree (DT) [17], random forest (RF) [18], and feature pooling have been proven effective in solving classification problems. Gen-

erally, a single machine learning algorithm cannot achieve an ideal classification performance. Thus multi-class linear SVM classifier was trained to classify the data from the pooling function in [19]. Since the computation cost of Multi-layer Perceptron (MLP) and SVM is expensive, k-ELM, a similar method to SVM, was adopted in [12]. [20] applied 10-fold cross validation to improve the accuracy of NB method on leaf shape classification and the accuracy is 74.10%. Aiming to achieve a better performance level, multi-classifiers fusion methods are proposed in recent years. For example, three shape descriptors, Zernike moments, generic Fourier descriptors, and shape signature were extracted in [21], and then fusion of the SVM, MLP and Dempster-Shafer methods was adopted to make multi-classifiers performance optimal. Three different CNN frameworks, including LeNet, MNIST-CNN, BS-CNN, were used to classify binary shape images in [22], and an adorable performance was gained.

The aim of this paper is to present a multi-classifier fusion approach for shape recognition. The key contribution of our work is to describe how to select features and how to achieve multi-classifier fusion from a shape dataset by using DT, SVM, KNN, fuzzy rule (FR), PNN, random forest (RF) and Gradient boosted tree (GBT). We show that the different level fusion of the classifiers would lead to different recognition accuracy.

3. Proposed Approach

The proposed approach of shape recognition can be divided into four steps, as outlined in Fig.1. In the first step, we extract features as center of gravity in dimension X and Y, mean, coarseness, proportion and aspect ratio, and then proportion and aspect ratio were removed in the feature selection step. To prevent the case of over-fitting and improve the level of generalization, in the third step, 10-fold cross validation was adopted for training and validation of each classifier. Finally, we collect and combine the outputs of these classifiers to boost the performance in the fusion process.

3.1. Feature Extraction

Given a binary image $I = f(i, j); i < M, j < N$ of shape X, six features are extracted directly from the input shape image.

The first two features are center of gravity in dimension x

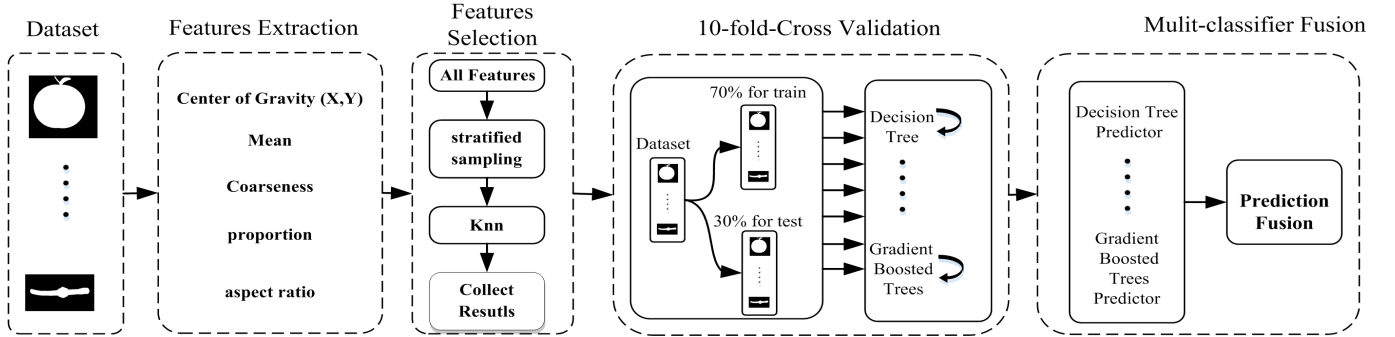


FIGURE 1. Overview of the proposed approach

and y which are defined as follows,

$$m_{pq} = \sum_{j=1}^N \sum_{i=1}^M i^p j^q f(i, j)$$

$$u_{pq} = \sum_{j=1}^N \sum_{i=1}^M (i - i_c)^p (j - j_c)^q f(i, j)$$

$$i_c = m_{10}/m_{00}, j_c = m_{01}/m_{00}$$

where μ_{pq} is centre distance and m_{pq} is m plus n order moment while i, j are the coordinate(x,y) of the image, p and q are non-negative integers. M is the length of image while N is the width of the image, i_c and j_c are the coordinates of center of gravity.

The third feature is formed by mean of the image, which can present distribution of color and is a statistic feature. The value of mean is defined by Eq.(2):

$$u_i = \frac{1}{N} \sum_{j=1}^N f(i, j) \quad (2)$$

where $i, j, f(i, j)$ is same as in(1) while N is the number of the pixels of an image.

Coarseness is the forth feature, which is a texture feature related to the size of the primitive elements.

$$A_k(i, j) = \sum_{x=i-2^{k-1}}^{i+2^{k-1}-1} \sum_{y=j-2^{k-1}}^{j+2^{k-1}-1} f(x, y) / 2^{2k} \quad (3)$$

$A_k(i, j)$ is the six averages of pixels calculated by computing the windows of size $2^k \times 2^k$, $k = 0, 1, \dots, 5$, around the pixel.

$$E_{k,h}(i, j) = |A_k(i + 2^{k-1}, j) - A_k(i - 2^{k-1}, j)|$$

$$E_{k,v}(i, j) = |A_k(i, j + 2^{k-1}) - A_k(i, j - 2^{k-1})| \quad (4)$$

Then we compute absolute differences $E_{k,h}(i, j)$ between the pairs of non-overlapping averages in the horizontal and vertical directions at each pixel(i, j). Find the value k that maximums the difference $E_{k,h}(i, j)$ in either direction and set the best size $S_{best}(i, j) = 2^k$. Finally, F_{crs} is calculated by averaging $S_{best}(i, j)$ over the entire image.

$$F_{crs} = \frac{1}{m \times n} \sum_{i=1}^m \sum_{j=1}^n S_{best}(i, j) \quad (5)$$

The fifth feature, namely proportion, represents the proportion of object to the whole image, which is defined as Eq.(6):

$$P = \frac{\sum f(i, j)}{m \times n}$$

$$f(i, j) = \begin{cases} 1, & \text{if } (i, j) \in D \\ 0, & \text{otherwise.} \end{cases} \quad (6)$$

where D presents the area of the object, m and n mean the width and the length of the image, respectively.

The final feature is aspect ratio:

$$r = \frac{j_{max} - j_{min}}{i_{max} - i_{min}} \quad (7)$$

where y_{max} and y_{min} present the maximum and minimum of coordinate in dimension y of the object, while x_{max} and x_{min} present the maximum and minimum of coordinate in dimension x of the object.

3.2. Feature Selection

As shown in Fig.2, the KNN based wrapper feature selection approach, which aims to find the best feature subset for training a classifier and improves recognition accuracy, is proposed in this section. Firstly, feature vectors extracted in section 3.1

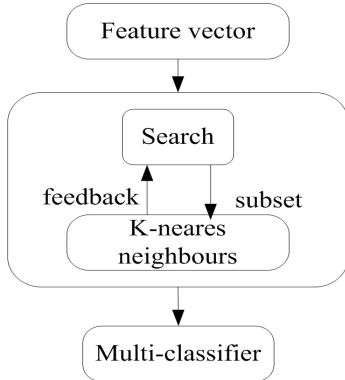


FIGURE 2. Process of feature selection

are divided into a training set and a validation set in the ratio of 70:30, through stratified sampling. We use K-neares neighbours (KNN) to evaluate the features and the k value is set to 7. Two features (proportion, aspect ratio) are removed in this step, and then we applied the 10-fold cross validation approach for training and validating each classifier by using a specific learning algorithm. As a result from the two stages, we have a feature vector x consist of center of gravity in dimension X and Y, mean and coarseness. The goal of cross validation is to measure the prediction accuracy of a classifier trained on the data set. Also, 10-fold cross validation is a useful method to enhance the ability of generalization. In particular, the input feature vectors are divided into 10 parts, 9 of which are taken as the training set in turn, and 1 as the test set. The average accuracy obtained through the 10 times run is used to estimate the overall performance of the classifier.

3.3. Multi-classifier Fusion

The strategy of multi-classifier fusion is based two principles: a) each single classifier must not be bad: b) different classifiers need to be complementary by encouraging the diversity among these classifiers, as indicated in [23].

In the proposed framework, we first train several single classifiers by using popular algorithms (e.g. SVM and KNN) that have different learning strategies, in order to select better performing classifiers and increase the diversity among them for the fusion purpose. Also, two decision tree ensembles (RF and GBT) are trained to get more diverse decision trees for boosting the performance in comparison with using a single decision tree. Finally, the outputs of the two ensemble classifiers are fused with the outputs of the selected single classifier for advancing the performance further.

In terms of how to fuse different classifiers, popular ways include majority voting and weighted voting, which are used in RF and GBT for fusion of different decision trees, respectively. However, in our proposed framework, the fusion of classifiers trained by using different learning algorithms is achieved through algebraic rules, which is essentially based on the mean/median/max of the continuous outputs (posterior probability for each of these classes) of these classifiers.

4. Experimental Setup and Results

Our experiment is conducted by using a MPEG-7 dataset which contains 1400 images, 70 shape categories, and 20 images per categories [24]. This experiment was built on the KN-IME Analysis Platform, which has abundant nodes for applying machine learning algorithms on Intel Core i7-6700K CPU 4GHz. As it was mentioned in Section 3, the KNN based wrapper feature selection approach and 10-fold cross validation were adopted in this experiment. The value of k is set to 7 since an odd number is generally recommended to avoid ties. 10-fold cross validation was undertaken in each classifier training stage. In the multi-classifier fusion stage, 5 base classifiers were trained by using DT, KNN, FR, SVM and PNN, respectively, towards selecting better performing ones for late fusion. Also, two decision tree ensembles are trained by using RF and GBT, respectively, towards increasing the diversity of decision tree classifiers.

In terms of parameters setting for these algorithms, we use gini index for attribute selection in decision tree learner, and the Reduced Error Pruning (REF) method is used to simplify decision trees. The min number records per node is set to 2, and average split point is checked in general options. Root split and Binary nominal splits options were unchecked. The RBF kernel was used in SVM learner and we set the $\sigma = 13$, overlapping penalty = 1. Similar to feature selection stage, we set K to 7 for K nearest neighbor. Min/Max norm and volume border based shrink function was chosen in Fuzzy Rule learner. As for PNN learner, theta minus and theta plus are set to 0.2 and 0.4, respectively. As for random forest learner, we use information gain ratio for split criterion in tree ensemble learner and the ensemble size (the number of decision trees that make up a random forest) is set to 100. We checked fraction of data to learn single model option and set the value as 1. The data sampling mode is set as random with replacement. We choose ‘‘Sample(square root)’’ for attribute sampling and ‘‘Use different set of attributes for each node’’ for attribute selection. Gradient boosted trees learner was used in this experiment by setting the tree depth as 10, number of models as 20 and learning rate as

0.1. We use XGboost for missing value handling. Bagging options were unchecked in advanced options. In terms of attribute sampling and selection, Gradient Boosted Trees learner has the same setting as random forest learner. Advanced options were unchecked in Fuzzy Rule learner and PNN learner. We used mid-point splits and binary splits for nominal columns in tree options of Random Forest learner and Gradient Boosted Trees learner and static random seed as well. In multi-classifier fusion stage, we use mean, median, maximum method respectively to boost the prediction accuracy.

TABLE 1. Table 1 Accuracy and Cohen’s kappa of individual and fusion classifiers

| | Accuracy | Cohen’s kappa |
|----------------------|----------|---------------|
| DT | 0.63 | 0.625 |
| KNN | 0.693 | 0.688 |
| SVM | 0.699 | 0.694 |
| FR | 0.709 | 0.705 |
| PNN | 0.682 | 0.678 |
| RF | 0.759 | 0.756 |
| GBT | 0.699 | 0.694 |
| Mean-based fusion | 0.763 | 0.759 |
| Median-based fusion | 0.766 | 0.762 |
| Maxmium-based fusion | 0.696 | 0.691 |

Table 1 shows the results of our experiment on MPEG-7 dataset. The recognition accuracy rates and Cohen’s kappa for different classifiers are listed in this table. Here, the performance of the DT method is the lowest, which are 63% (accuracy) and 62.5% (Cohen’s kappa) respectively. Both RF and GBT achieves to improve the performance of decision tree learning, and the RF ensemble classifier even outperforms all the other individual classifiers, with the accuracy of 75.9% and Cohen’s kappa of 75.6%. For the three multi-classifier fusion methods, namely, Mean-based fusion, Median-based fusion and Maximum-based fusion methods, the first two show their capability of boosting the overall performance comparing with the RF ensemble. The Median-based fusion method achieves the best performance in terms of both accuracy and Cohen’s kappa.

Overall, the results show very supporting evidence for adopting mutli-classifier fusion, especially given the fact that the number of features is very small leading to lower classification accuracy (no greater than 70%) through using individual classifiers but our setting of ensemble learning achieves to improve the performance up to 76.6%. In other words, the results indicate that it is encouraged to adopt mutli-classifier fusion towards advancing the overall classification performance, espe-

cially when the number of features is not sufficient for training an individual classifier with high performance.

5. Conclusions

In this paper, we proposed a four-step approach of shape recognition. In particular, six features were extracted from a 2D shape data set and four of them were then selected heuristically by using KNN based wrapper approach. Furthermore, five base classifiers were trained on the selected features by using DT, KNN SVM, KNN, FR, PNN, respectively. Also, two decision tree ensembles were trained by using RF and GBT, respectively. Finally, all the above classifiers were fused through three algebraic rules, namely mean, median and maximum.

The experimental results show that the training of two decision tree ensembles achieves to advance the performance of decision tree learning and the RF ensemble classifier even outperforms all the five base classifiers. Moreover, the algebraic fusion through both the mean and median rules achieves to advance the classification performance further, in comparison with the RF classifier.

In the future, we will investigate to define more relevant features to advance the performance of the multi-classifier fusion on shape recognition. Also, due to the large number of classes, it is very likely to have more features that are only relevant for one or some but not all of the classes, so we will look to adopt multi-task feature selection, i.e. it is to select an optimal set of features for each class towards effective recognition of instances of each specific class.

Acknowledgements

This paper is supported by the Natural Science Foundation of Hebei Province (No. F2016203211) and Young Teachers independent research program of Yanshan University (Class A, No.15LGA014).

References

- [1] L. Kurnianggoro et al., “A survey of 2D shape representation: Methods, evaluations, and future research directions”, *Neurocomputing*, pp. 116, March. 2018.
- [2] Dengsheng Zhang, Guojun Lu, “Review of shape representation and description techniques”, *Pattern Recognition*, Vol 37, No. 1, pp. 1-19, Jan. 2004.

- [3] Ming. K. Hu, "Visual pattern recognition by moment invariants", *IRE Transactions on Information Theory*, Vol 8, No. 2, pp. 179-187, Mar. 1962.
- [4] Michael. R. Teague., "Image analysis via the general theory of moments", *Journal of the Optical Society of America*, Vol 70, No. 8, pp. 920-930, Aug. 1980.
- [5] Guojun Lu, A. Sajjanhar, "Region-based shape representation and similarity measure suitable for content-based image retrieval", *Multimedia Systems*, Vol 7, No. 2, pp. 165-174, Mar. 1999.
- [6] A. Goshtasby, "Description and discrimination of planar shapes using shape matrices", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol 7, No. 6, pp. 738-743, Nov. 1985.
- [7] Bryan S. Morse, "Computation of object cores from grey-level images", University of North Carolina at Chapel Hill, Ph.D. Thesis, November 1994.
- [8] Dengsheng Zhang, Guojun Lu, "Generic Fourier descriptor for shape-based image retrieval", *IEEE International Conference on Multimedia and Expo (ICME2002)*, Lausanne, Switzerland, pp. 425-428, Aug 2002.
- [9] H. Freeman, "On the encoding of arbitrary geometric configurations", *IRE Transactions on Electronic Computers*, Vol 10, No. 2, pp. 260-268, Jul. 1961.
- [10] Berrada, F., et al, "Review of 2d shape descriptors based on the curvature scale space approach", *2011 International Conference on Multimedia Computing and Systems*, Ouarzazate, Morocco, pp. 16, Apr 2011.
- [11] Yunwen Chen, Yanqiu Chen, "Invariant description and retrieval of planar shapes using radon composite features", *IEEE Trans Signal Process*, Vol 56, No. 10, pp. 4762-4771, Nov. 2008.
- [12] Cong Lin, et al., "Efficient shape classification using region descriptors", *Multimedia Tools and Applications*, Vol 76, No. 1, pp. 83-102, Jan. 2017.
- [13] Glauco V. Pedrosa, Agma J. M. Traina, Celia A. Z. Barcelos, "Retrieving 2D shapes by similarity based on bag of salience points", *Multimedia Tools and Applications*, Vol 76, No. 20, pp. 20957-20971, Oct. 2017.
- [14] C. Cortes, V. Vapnik, "Support-vector networks", *Machine Learning*, Vol 20, No. 3, pp. 273-297, Sep. 1995.
- [15] N. S. Altman, "An Introduction to Kernel and Nearest-Neighbor Nonparametric Regression", *American Statistician*, Vol 46, No. 3, pp. 175-185, Aug. 1992.
- [16] George H. John, "Estimating Continuous Distributions in Bayesian Classifiers", In *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*, pp. 338-345, Aug. 1995.
- [17] Quinlan. J. Ross, *C4.5: Programs for Machine Learning*, Morgan Kaufmann, San Francisco, 1993.
- [18] L. Breiman, "Random Forests", *Machine Learning*, Vol 45, No. 1, pp. 5-32, Oct. 2001.
- [19] Xinggang Wang, et al, "Bag of contour fragments for robust shape classification", *Pattern Recognition*, Vol 47, No. 6, pp. 2116-2125, Jun. 2014.
- [20] Francis. Rey. F. Padoa, Elmer. A. Maravillas, "Using Nave Bayesian method for plant leaf classification based on shape and texture features", *International Conference on Humanoid, Nanotechnology, Information Technology, communication and Control, Environment and Management*, Cebu City, Philippines, pp. 1-5, Dec 2015.
- [21] Mohammad. A. Bagheri, Qigang Gao, S. Escalera, "Logo Recognition Based on the Dempster-Shafer Fusion of Multiple Classifiers", *Advances in Artificial Intelligence*, Vol 7884, No. 1, pp. 1-12, 2013.
- [22] Habibollah. A. Atabay, "BINARY SHAPE CLASSIFICATION USING CONVOLUTIONAL NEURAL NETWORKS", *IIOAB Journal*, Vol 7, No. 5, pp. 332-336, Oct. 2016.
- [23] Z. H. Zhou. "Ensemble Methods: Foundations and Algorithms", Chapman and Hall/CRC, 2012.
- [24] N. Thakoor, J. Gao, S. Jung, "Hidden Markov Model-Based Weighted Likelihood Discriminant for 2-D Shape Classification", *IEEE TRANSACTIONS ON IMAGE PROCESSING*, Vol 16, No. 11, pp. 222-226, Nov. 2007.