Noname manuscript No.
(will be inserted by the editor)

Shape Recognition through Multi-level Fusion of Features and Classifiers

Xinming Wang[1] · Weili Ding[1*] · Han Liu[2] · Xiangsheng Huang[3]

Received: date / Accepted: date

Abstract Shape recognition is a fundamental problem and a special type of image classification, where each shape is considered as a class. Current approaches to shape recognition mainly focus on designing low-level shape descriptors, and classify them using some machine learning approaches. In order to achieve effective learning of shape features, it is essential to ensure that a comprehensive set of high quality features can be extracted from the original shape data. Thus we have been motivated to develop methods of fusion of features and classifiers for advancing the classification performance. In this paper, we propose a multi-level framework for fusion of features and classifiers in the setting of granular computing. The proposed framework involves creation of diversity among classifiers, through adopting feature selection and fusion to create diverse feature sets and to train diverse classifiers using different learning algorithms. The experimental results show that the proposed multi-level framework can effectively create diversity among classifiers leading to considerable advances in the classification performance.

Keywords Machine Learning · Ensemble Learning · Image Classification · Shape Recognition · Feature Extraction · Granular Computing

Xinming Wang
E-mail: wang8297161@stumail.ysu.edu.cn

✉[*] Weili Ding
E-mail: weiye51@ysu.edu.cn

Han Liu
E-mail: LiuH48@cardiff.ac.uk

Xiangsheng Huang
E-mail: xiangsheng.huang@ia.ac.cn

[1] Department of Automation, Institute of Electrical Engineering, Yanshan University, 438 West of Hebei Avenue, Haigang District, Qinghuangdao 066004, China
[2] School of Computer Science and Informatics, Cardiff University, Queen's Buildings, 5 The Parade, Cardiff CF24 3AA, United Kingdom
[3] Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China
[*] Corresponding author

1 Introduction

Shape recognition is a critical part of pattern recognition due to its wide applications in image retrieval, object detection surveillance systems and any related areas. In the recent years, machine learning gains its popularity due to its modeling ability. In shape recognition tasks, the same kind of shapes is assigned a specific label, which consists of data samples, and effective feature descriptors extracted from these data combined with powerful machine learning algorithms usually lead to good recognition results.

Feature extraction and classification are two significant steps in shape recognition, which can directly affect the recognition results. In the past few years, effective shape features and classification methods have been studied. In general, a high dimensional feature set usually contains redundant information which has negative effects on the recognition result. Specifically, the redundant information exists in local features and global features. In addition, classification performance can be varied due to the diversity among different classifiers, i.e., classifiers trained using different learning algorithms show different classification performance on the same feature set, and the classifiers trained using the same algorithm show different performance on different feature sets. Therefore, it becomes our motivation to select effective features without destroying local and global relationships and advance the classification performance through fusing different classifiers trained on distinct features.

In this paper, first, we propose a multi-level framework for shape recognition, which involves creation of diversity among feature subsets by adopting feature selection and fusion, and training diverse classifiers on them using seven learning algorithms which are decision tree, k nearest neighbour, support vector machine, fuzzy rule, probabilistic neural network, random forests and gradient boosted trees. Second, we discuss how to improve the recognition rate by multi-classifier fusion from the perspective of granular computing. With these contributions we are able to create diverse classifiers to advance the performance.

The rest of this paper is organized as follows: Section 2 presents related work on shape features and machine learning related algorithms. In Section 3, we give the details of the extracted features and the design of feature fusion and classifier fusion. Experimental configurations and discussion of the results are given in Section 4. In Section 5, we summarize the contributions of this paper and suggest some future directions.

2 Related Work

In this section, we provide an overview of feature extraction in the context of shape recognition and a review of machine learning techniques that have been popularly used leading to effective recognition of shapes.

2.1 Overview of features used for shape recognition

Shape recognition (Kurnianggoro et al, 2018; Zhang and Lu, 2004) has been widely studied in computer vision in the past decades. Shape representation descriptors with hierarchical structure have better performance on shape classification due to the fact that the coarse grained characteristics can distinguish the obvious differences between two shapes, while the differences in details can be further recognized through fine-grained features. In general, the existing shape features can be roughly divided into global approaches and local approaches.

Global approaches extract features on the whole shape contour or shape contents which include essential parameters, stochastic methods, scale space descriptors, spectral domains, moment-based descriptors and grid based methods. Essential parameters like the center of gravity, convexity and solidity are simple global descriptors with low complexity of computation and implementation. These simple descriptors only perform well on drastic perceptually shapes. However, a meaningful shape descriptor can be constructed by a combination of these simple descriptors. Stochastic methods like Auto-regressive (AR) (Sekita et al, 1992) model describe shape boundaries through its parameters. Scale space descriptors like curvature scale space (CSS) (Mokhtarian and Bober, 2003) describe the change of curvature of a shape in different smoothing values which are less sensitive to noise and boundaries variations. Fourier descriptor (FD) and wavelet descriptor (WD) can handle the issue of noise sensitivity in a spectral domain. Classic moment-based descriptors include Hu-moments and Zernike moments. The former have relatively low complexity of computation and characteristic of being invariant to translation, scaling and rotation, whereas the latter are more robust. Grid based methods (Lu and Sajjanhar, 1999) encode the shape context as a binary feature vector. Before the encoding, the normalization such as scaling the shape into a fixed size needs to be achieved to cop with the issue of translation, rotation and scaling. Recently, Multi-scale angular features were extracted from contour points by Arjun and Mirnalinee (2018). Due to the fact that a multi-scale scheme will lead to complex computation, sequential backward selection (SBS) was employed against the expensive computation. The experiments on the MPEG-7 shape data set show that this descriptor is invariant to scaling and rotation transformation.

Rather than describing the shape from the whole shape contour or region, local approaches broke the whole shape contour or region into parts. Local approaches include chain codes, histogram based descriptors, decomposition based methods and medial axis based methods. Chain codes (Malo and Freeman, 2009) can encode a shape contour to a set of vectors with 4 or 8 directions. Since it focuses on the local information of the given shape contour, variations might impact a lot on the descriptors. Histogram based descriptors like shape context (Belongie et al, 2002) can reflect the distribution of contour points nearby each point which is less sensitive to noise data. The decomposition based methods like polygon decomposition divides the contour into small primitives, each of which includes internal angle, distance from the next vertex, and its x and y coordinates. Medial axis based methods like region skeleton describe the shape using its topological structure. Recently, by Giangreco Maidana et al (2018), a new local descriptor named Contour-Point Signature (CPS) contains information of arbitrary contour points which are proven to be invariant to translation, scaling and rotation. By Lin et al (2015), region area descriptor (RAD), region skeleton descriptor (RSD) and simplified shape signature (SSS) were proposed. RAD and RSD exploit information from shape contour and skeleton while SSS is only contour based. The final descriptor is concatenated by these novel descriptors. Also, contour and skeleton information were also considered as complementary, Skeleton-associated Shape Context (SSC) was proposed by Shen et al (2016). And the Bag of features scheme was applied to encode the SSC part to a meaningful feature as Bag of Skeleton-associated Contour Parts (BSCP). By Priyanka and Sudhakar (2018), a descriptor using hybrid geometrical concepts to exploit local information named Triangulated Feature Descriptor (TFD) was proposed.

2.2 Review of machine learning methods

While many shape features have been proven to be effective, little attention has been paid to classification approaches in shape recognition. Machine learning has been an important tool for pattern recognition, and in which, feature fusion and classifier fusion are two important ways to improve the performance.

As mentioned in 2.1, many shape features combined with learning algorithms can achieve good performance. By Priyanka and Sudhakar (2018), the combination of Triangulated Feature Descriptor (TFD) and kNN achieve the accuracy of 95.35%. SVM is used by Shen et al (2016) to verify the proposed feature in terms of accuracy. Comparing with SVM, k-ELM is more efficient for high-dimensional data. In Lin et al (2015), kernel extreme learning machine (k-ELM) is used to classify the shape.

Generally, feature fusion can be separated as two stages: feature selection and feature combination. Feature selection (Jovi et al, 2015) plays an important role in classification tasks and its main purpose is to make the classifier achieve better performance by reducing the redundant features and selecting the discriminative features without transformation. Commonly used methods are filter methods, wrapper methods, embedded methods and hybrid methods. In order to achieve the goal of enhancing the efficacy of the learning algorithm, a wrapper based method directly validates the candidate feature subsets through training classifiers by using the algorithms such as support vector machine (SVM), K nearest neighbour (KNN) and naive bayes (NB). The optimal feature subset is selected according to the performance of the classifiers trained on the candidate feature subsets. In addition, the performance might be varied a lot due to the use of different classifiers. Unlike the wrapper method, the filter method validates each feature subset according to predefined performance measures, such as information gain and chi-square, instead of using an algorithm for training classifiers on the candidate feature subsets. In general, wrapper methods have better performance than filter methods but show higher computational complexity. Embedded methods are literally embedded in the classifier training algorithm to select the features which can reduce the cost of computation without loss of classification performance such as Wang et al (2015); Bermejo et al (2014). After feature selection, features are combined through a parallel strategy or a serial strategy (Yang et al, 2003). Many studies used the feature fusion strategy to advance the performance. A fusion method inspired by Canonical correlation analysis (CCA) to find the discriminant features was proposed by Sun et al (2005), and experiments on a handwritten Arabic numerals data set and a face data set showed that the recognition rate using this method could be improved considerably. By Guan et al (2010), good performance on an x-ray image data set was achieved using a learning-based feature selection. In Lin et al (2013), three shape contour features including shape context(SC), inner distance shape context (IDSC) and contour points distribution histogram (CPDH) were fused, then KNN was used for classification on the MPEG-7 shape dataset. The results show that the fusion features achieve the excellent performance.

According to the study of Mohandes (Mohandes et al, 2018), fusion can be conducted in the sensor level, the feature level or the decision level, depending on the stage at which the fusion method operates. Here, we focus on the problem of classifiers fusion methods in the decision level according to the outputs of the classifiers, the fusion rules and the ensemble creation methods. Comparative studies relevant to the combination rules can be referred in Kuncheva (2002). Ensemble learning is a popular way to improve the overall performance in pattern recognition. Bagging and Boosting are two classic approaches of ensemble learning. Bagging aims to generate multiple samples from the original data set to enable more diverse classifiers trained on the samples, such as random forests (RF). Boosting makes the classifier perform better through iterative training such as Gradient boosted trees (GBT) and Adaboost. Bagging and Boosting employ majority voting and weighted voting, respectively, for classifiers fusion. Inspired by granular computing, a probabilistic voting method was proposed by Liu and Cocea (2017b), in which the experimental results show that it is effective to improve the overall performance. Recently, classifiers fusion methods have been used in many fields. A local and global classifier fusion framework was proposed by Ding et al (2017) to enhance the performance on digital chest x-ray images analysis. A whole image is divided into sub-

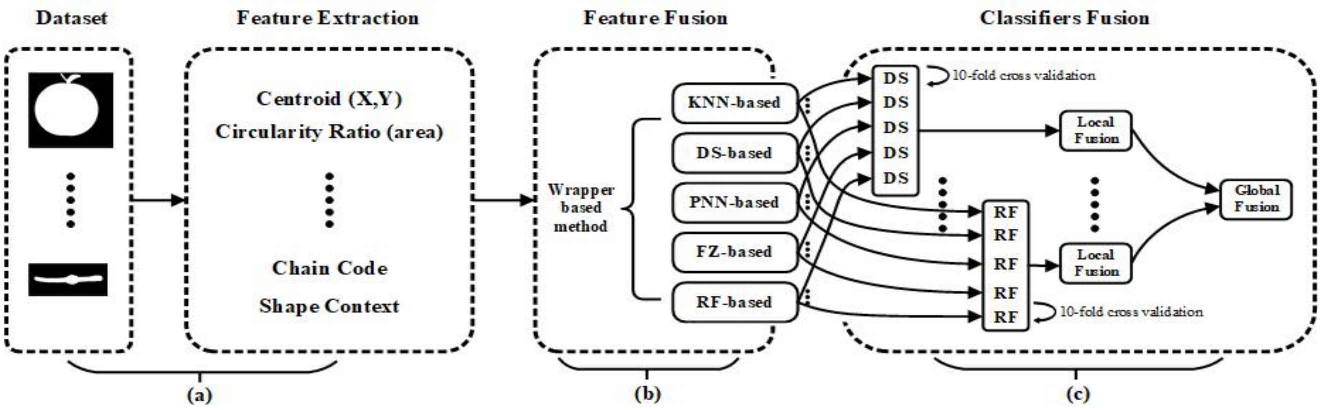

Fig. 1 Overview of the proposed approach: (a).Global and local features are extracted from each shape. (b).Five wrapper based feature selection methods are used to select 5 best feature sets. (c).Each base classifier in each local ensemble is trained on one of the 5 best feature sets and predictions of them are fused. Finally, results of each local ensemble are fused into results of the global ensemble.

regions for features extraction to train local classifiers. Global classifiers are trained on the global features extracted from the whole image. The final decision is obtained using the linear fusion of local and global classifiers. A classifier fusion system named Droid Fusion for Android malware detection was proposed by Yerima and Sezer (2018). In Wan et al (2018), the KNN and NB classifiers are fused in the decision level for tourist route recommendations. A previous study (Ding et al, 2018) has made a simple attempt with fusion of classifiers.

3 Proposed Approach

The proposed approach of shape recognition can be divided into three steps as shown in Fig. 1. In the first step, the global features and the local features are extracted from the shape data set. Then, the distinct feature subsets are obtained by different wrapper-based approaches of feature selection and serial combination. Furthermore, in the local fusion stage, 5 classifiers are trained on the selected feature subsets by using one of pre-selected learning algorithms to create a primary (local) ensemble. Finally, the local ensembles, which are created using the learning algorithms for training of base classifiers, are fused to achieve global fusion of classifiers.

3.1 Feature extraction

Considering the hierarchical shape descriptors have better ability of representation, we combine 17 simple shape descriptors into a robust and accurate shape decriptor, which contains local and global features, reflecting the strong complementarity of hierarchical descriptors, that is, the global features can distinguish the drastic differences between two shapes, while the differences in details can be further recognized through local features.

The first feature is circle variance f_1 which represents the ratio of standard deviation σ to average value μ of radial distance. Radial distance is a vector in which each element ρ_i represents the distance from each contour point p_i to centroid g. The definitions are (1).

$$f_1 = \frac{\sigma}{\mu}; \rho_i = \|p_i - g\|_2 \tag{1}$$

The second feature is circularity ratio (area) f_2 which is the ratio of the area of the object A_{shape} to its area of tangential circle A_{circle} while circularity ratio (perimeter), the third feature f_3 is the ratio of the area of the shape A_{shape} to the power 2 of the perimeter of its tangential circle P^2_{circle}. Both of the above features are shown as follows (2).

$$f_2 = \frac{A_{shape}}{A_{circle}}; f_3 = \frac{A_{shape}}{P^2_{circle}} \tag{2}$$

The other five ratio involved features are solidity (f_4), eccentricity (f_5), convexity (f_6), hole area ratio (f_7) and rectangularity (f_8) which are defined in Equation (3). Given an object and its convex hull, convexity represents the ratio of the perimeter of the convex hull P_{hull} to the perimeter of the object P_{shape}. Solidity reflects the ratio of the area of the object A_{shape} to the area of the convex hull A_{hull}. Eccentricity is the ratio of the length of major axis λ_1 to the length of minor axis λ_2. In addition, λ_1 (f_9) and λ_2 (f_{10}) are also chosen as features. Hole ratio is the ratio of the area of hole A_{hole} to A_{shape}. Rectangularity represents the ratio of A_{shape}

to the area of the bounding box A_{box}.

$$f_4 = \frac{A_{shape}}{A_{hull}}; f_5 = \frac{\lambda_1}{\lambda_2};$$
$$f_6 = \frac{P_{hull}}{P_{shape}}; f_7 = \frac{A_{hole}}{A_{shape}}; \tag{3}$$
$$f_8 = \frac{A_{shape}}{A_{box}}; f_9 = \lambda_1; f_{10} = \lambda_2;$$

Another feature f_{11} is centroid which is defined as the average values of x-axis coordinates and y-axis coordinates in Equation (4) where x and y jointly represent the coordinate position of the object and n represents the number of pixels of the object.

$$f_{11} = (\frac{1}{n}\sum\nolimits_{i=1}^{n} x_i, \frac{1}{n}\sum\nolimits_{i=1}^{n} y_i) \tag{4}$$

Hu moments (f_{12}) and Zernik moments (f_{13}) which have already been proven to be compact and effective are also used here. More details can be found in Hu (1962); Reed Teague (1980).

Smoothness (f_{14})(Ding et al, 2015), defined as Equation (5), represents the curve bending degree. We use this feature to measure the smoothness of the contour.

$$f_{14} = \sum_{n=1}^{m} (k_n - k_{mean}) \tag{5}$$

where k_n is the curvature of each contour point and k_{mean} is the mean of the curvature of all points on the shape contour.

Corner number (f_{15}) of the shape contour is the final global feature chosen in our study. It is extracted by using the CSS algorithm (Mokhtarian and Bober, 2003).

When dealing with the issue of partial occlusion and noise, local features would be more effective. Chain codes (f_{16}) (Malo and Freeman, 2009) and shape contexts (f_{17}) (Belongie et al, 2002) were extracted to describe the local contour of each shape in details. For each contour point, 8-orientation encoding method was used in chain codes and 36 bins (6 for radial direction and also for circumferential direction) were set in shape context to capture distribution of adjacent contour points. Due to the fact that the length normalization of the input data is necessary for many machine learning methods, we normalize the chain codes features into a vector of length 8.

$$f_{16} = \{C_1, C_2, \cdots C_8\} \tag{6}$$

where each element represents the frequency of each direction. We also normalize the shape contexts features into a vector of length 36, namely,

$$f_{17} = \{S_1, S_2, \cdots S_n, n = 36\} \tag{7}$$

where each element represents the statistical values of all contour points in each bin.

Finally, a feature vector F_i with a dimension of 66 which contains global and local shape features was constructed for each sample, shown in Table 1.

$$F_i = \{f_1, f_2, ...f_{17}\} \tag{8}$$

Table 1 Compositions of the feature vector with a dimension of 66.

dimension	8	36	2	7	13
feature	f_{16}	f_{17}	f_{11}	f_{12}	f_1, f_2, f_3, f_4, f_5 $f_6, f_7, f_8, f_9, f_{10}$ f_{13}, f_{14}, f_{15}

3.2 Feature selection

Wrapper based methods of feature selection can achieve good performance and are also convenient to implement. Instead of using only one wrapper-based approach of feature selection, multiple wrapper-based approaches of feature selection were adopted for the original feature set. In other words, different learning algorithms are used for classifiers training in order to achieve different ways of feature subset selection, leading to diverse feature subsets being obtained. From this point of view, more diverse models can be trained on different feature subsets, which leads to advances in the classification performance in the classifiers fusion stage. After selection, the dimensions of global and local features depend on the wrapper used, for example, the dimensions of global and local features decreased from 22 to 13 and from 44 to 30, respectively, after using PNN-based feature selection. Furthermore, the dimensions of feature subsets selected using different wrappers are different from each other. For instance, the dimensions of features decreased to 43 by using PNN-based feature selection whereas the dimensions decreased to 38 by using RF-based feature selection. The combination of global and local features is similar to the concept "parallel structure" as shown in Fig. 2, which can roughly discriminate objects with distinct shapes using global features and identify the specific object by using local features. Since wrapper based feature selection does not consider the internal relationships between global and local features, it should be used separately on the global and local feature sets to avoid the inexplicability of the selected feature subset. Furthermore, for each local feature, chain code and shape context also should be selected respectively.

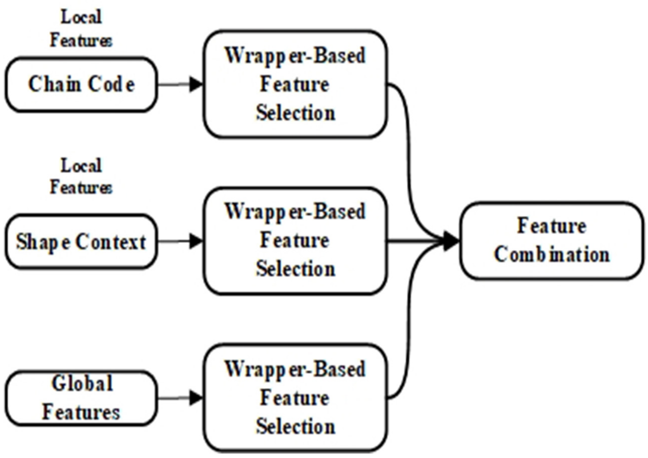

Fig. 2 Parallel structure for feature selection: Chain code, shape context and global features are selected individually in the same wrapper based feature selection. The selected features are then combined in series connection.

In particular, five base classifiers, which are trained by using decision tree (DT), KNN, PNN, Fuzzy Rule (FR) and RF, were adopted to select the optimal feature subset depending on their performance. The forward feature selection strategy was employed here to decide the start feature, which considers adding features to an empty set to test the performance. After feature selection, five feature subsets were obtained corresponding to the trained classifiers.

3.3 Classifiers fusion

After feature selection, five feature subsets are obtained through different wrapper based approaches of feature selection. Inspired by random feature subset selection in random forests, performance are likely to be improved by training multiple classifiers using distinct feature subsets obtained in the feature selection stage. A hierarchical structure is used in the classifiers fusion stage, which involves local fusion and global fusion. There are seven local fusion parts corresponding to seven learning algorithms which are DT, KNN, SVM, PNN, FR, RF and GBT. For each local fusion part, five feature subsets are used here to train five distinct classifiers using the same learning algorithm, and the predictions of these classifiers are fused by using the mean rule. In this setting, the performance is expected to be improved in comparison with each of the individual classifiers. Finally, global fusion is undertaken by fusing in the mean rule the predictions obtained from the local fusion parts.

3.4 Applications of granular computing concepts

The proposed framework of fusion of features and classifiers is essentially designed in the setting of granular computing, which is an information processing paradigm. In general, granular computing is aimed at structural thinking at the philosophical level but also at structural problem solving at the practical level (Yao, 2005b).

In theory, granular computing concepts mainly involve granules and granularity (Pedrycz, 2011; Pedrycz and Chen, 2011, 2015b,a). A granule is essentially a collection of smaller particles that can form a larger unit. Due to the different sizes of granules, it is highly necessary to have different levels of granularity for structural information processing.

In practice, the applications of granular computing concepts are usually achieved through two operations, namely, granulation and organization (Yao, 2005a). The former operation aims to decompose a whole (a larger granule in a higher level of granularity) into parts (smaller granules in a lower level of granularity), whereas the latter operation aims to integrate parts into a whole (Yao, 2005a). The two operations have been popularly taken to implement the top-down and bottom-up approaches, respectively (Liu and Cocea, 2017a; Liu et al, 2018).

In our proposed framework, granulation is operated through decomposing the information of original images into two parts, namely, chain code and shape context, where both parts involve local features. Organization is operated through fusion of features selected from three feature sets, namely, chain code, shape context and global features, as shown in Fig. 2. In the above context, each feature set F_i is viewed as a granule g_i in a higher level of granularity and each subset F_{ij} of selected features is viewed as a sub-granule of g_i in a lower level of granularity.

On the other hand, in the setting of classifiers fusion, a primary (local) ensemble, which consists of 5 base classifiers trained on 5 different feature subsets by using the same learning algorithm, is viewed as a basic granule in the bottom level of granularity. The secondary (global) ensemble, which consists of 5 local ensembles created using 5 different learning algorithms, is viewed as a larger granule (in a higher level of granularity) that is made up of 5 basic granules. Moreover, the fusion of base classifiers for creation of a local ensemble and the fusion of local ensembles for creation of a global ensemble are both viewed as a kind of organization.

Table 2 Accuracy of each classifier trained on different feature sets in local classifier fusion

Accuracy	DT-based	FR-based	RF-based	PNN-based	KNN-based	Local Classifier Fusion
DT (Quinlan, 1993)	77%	75.9%	74.9%	74.7%	76.5%	84.9%
KNN (Aha et al, 1991)	81.6%	80.9%	81.2%	81.9%	81.9%	81.4%
SVM (Platt, 1999)	73.4%	73.6%	74%	72.6%	73.1%	74.1%
PNN (Berthold and Diamond, 1998)	78.4%	77.8%	79.1%	77%	78.1%	79.1%
FR (Berthold, 2003)	80.4%	79.2%	79.1%	78.1%	82.1%	83.6%
RF (Cutler et al, 2012)	94.4%	95.1%	95.1%	93.9%	94.6%	95.3%
GBT (Friedman, 2000)	86.2%	86.4%	86.6%	86.6%	86.3%	87.8%

4 Experimental Results

Our experiment is conducted by using the MPEG-7 CE Shape-1 Part B data set which contains 1400 images, 70 shape categories and 20 images per category (Thakoor et al, 2007). Several instances of the data set are shown in Fig. 3. This experiment was built on the KNIME Analysis Platform, which has abundant nodes for applying machine learning algorithms on Intel Core i7-6700K.

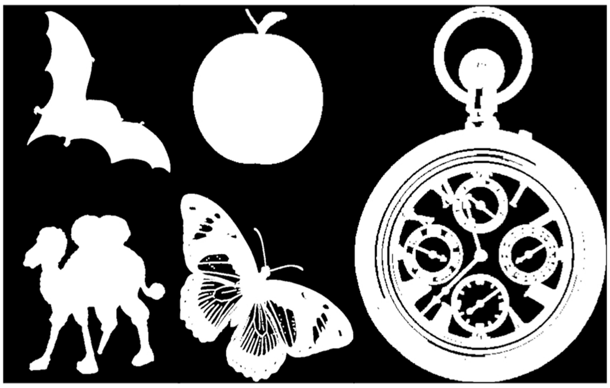

Fig. 3 Examples in MPEG-7 CE Shape-1 Part B dataset

Here we give the settings of each algorithm in each stage. In the feature selection stage, diverse feature subsets were obtained through wrapper based feature selection (Dash and Liu, 1997) approaches which are driven by 5 learning algorithms, namely, KNN (Aha et al, 1991), DT (Quinlan, 1993), PNN (Berthold and Diamond, 1998), FR (Berthold, 2003) and RF (Cutler et al, 2012). In the classifier fusion stage, two additional learning algorithms, GBT (Friedman, 2000) and SVM (Platt, 1999), alongside the above five ones are used. The setting of the learning algorithms in the feature selection stage is the same as the one in the fusion stage.

The only parameter for the KNN algorithm is the value of K which is set to 7 in this experiment. The RBF kernel for SVM with sigma =13 and overlapping penalty = 1 is used. For the DT learner, Gini index is used for attribute selection and the Reduced Error Pruning (REF) method is used to simplify decision trees to avoid overfitting. The Min number records per node is set as 2, and the average split point in general options is chosen. Root split and Binary nominal splits options are unchecked. For the FR node, activation across all rules is computed according to the Min/Max norm, and the volume border based shrink function is chosen to reduce the rules to avoid conflicts. As for the PNN learner, theta minus and theta plus are set to the default values, which are 0.2 and 0.4, respectively. As for the RF learner, information gain ratio is chosen in tree options for split criterion, and the ensemble size is set to 100, which means that a forest consists of 100 trees. Another ensemble learning method used in this experiment is the GBT learner. For this learner, the tree depth is set as 10, the number of models as 20 and the learning rate as 0.1. When dealing with the issue of an instance belonging to the none class, XGboost is used to handle the missing value. Data sampling and attributes sampling and selection are the key to the ensemble learner. Bootstrapping is used in random forests, which means that the data sampling mode should be set as random with replacement. The number of instances before and after sampling is the same. For GBT, no data sampling method is used in this experiment, i.e. Bootstrapping is not used here, which means that each tree is trained on the same sample. For attributes sampling and selection, each tree in the RF learner and the GBT learner uses a different feature subset, the size of each feature subset is the square root of the total number of attributes. Other configurations include mid-point splits and binary splits for nominal columns in tree options and static random seed are chosen for both ensemble learners. The Mean rule is chosen as the fusion rule and the weight of each classifier is set as 1.

The accuracy obtained through the global fusion of classifiers is 91.4%. No matter whether the RF predictor is involved or not in the global fusion, the accuracy is not changed. The results on the accuracy of the local fusion of classifiers and each of the base classifiers are shown in Table 2. Each row represents the accuracy of each base classifier trained on each of the obtained feature subsets and the accuracy of local fusion of classifiers. For instance, the first row indicates the accuracy of each DT classifier trained through five feature subsets selected by using DT, FR, RF, PNN and KNN for feature evaluation and the accuracy of fusion of them. By using distinct feature subsets to train classifiers, the adoption of local fusion generally leads to an improvement of the performance in comparison with the performance of each individual base classifier, except for the case that KNN is used for training base classifiers on the feature subsets. When DT is used for training base classifiers, the improvement of the performance is much more obvious (by 10.2%), in comparison with using the other algorithms for training base classifiers. For all of the 7 selected learning algorithms, RF shows the best performance, which also indicates the relevance of adopting ensemble learning. Even though local fusion leads to an effective improvement of the classification performance, it is still worse than the one obtained using the RF learner only. The above phenomenon is likely due to the case that some instances are only correctly classified by RF but the summed confidence of the other classifiers is higher than the one of RF, leading to incorrect classifications of these instances through fusion of the classifiers. In the above case, the total number of incorrectly classified instances would be increased, if some instances that are incorrectly classified by RF can not be correctly classified after the fusion of the classifiers, leading to the drop in the overall classification accuracy, in comparison with using RF.

In order to show in-depth analysis of why and how classifiers fusion can lead to advances in the performance, we provide some analysis of the diversity among classifiers in an ensemble. In particular, for local fusion of classifiers, Pearson correlation coefficient for each pair of base classifiers is shown in Tables 3-5,7-10. Each base classifier is trained on the feature subset selected by using the learning algorithm specified in a bracket.

Table 3 shows that the diversity among the DT classifiers is generally higher (correlation coefficient between 0.722 and 0.778). In this case, the local fusion of

Table 3 Pearson correlation coefficient of each DT classifier trained through 5 selected feature sets

Correlation	DT (DT)	DT (FR)	DT (RF)	DT (PNN)	DT (KNN)
DT(DT)	1	0.756	0.751	0.751	0.754
DT(FR)	0.756	1	0.778	0.741	0.756
DT(RF)	0.751	0.778	1	0.722	0.738
DT(PNN)	0.751	0.741	0.722	1	0.76
DT(KNN)	0.754	0.756	0.738	0.760	1

the DT classifiers leads to a considerable improvement of the classification performance (0.849), in comparison with the best performing base classifier (0.77), as shown in Table 2.

Table 4 Pearson correlation coefficient of each KNN classifier trained through 5 selected feature sets

Correlation	KNN (DT)	KNN (FR)	KNN (RF)	KNN (PNN)	KNN (KNN)
KNN(DT)	1	0.924	0.951	0.886	0.9
KNN(FR)	0.924	1	0.925	0.884	0.906
KNN(RF)	0.951	0.925	1	0.883	0.89
KNN(PNN)	0.886	0.884	0.883	1	0.905
KNN(KNN)	0.9	0.906	0.89	0.905	1

Table 4 shows that the diversity among the KNN classifiers is generally lower (correlation coefficient between 0.883 and 0.925). In this case, the local fusion of the KNN classifiers leads to a marginal drop in the classification performance (0.814), in comparison with the best performing base classifier (0.819).

Table 5 Pearson correlation coefficient of each SVM classifier trained through 5 selected feature sets

Correlation	SVM (DT)	SVM (FR)	SVM (RF)	SVM (PNN)	SVM (KNN)
SVM(DT)	1	0.829	0.835	0.767	0.786
SVM(FR)	0.829	1	0.832	0.773	0.800
SVM(RF)	0.835	0.832	1	0.770	0.796
SVM(PNN)	0.767	0.773	0.770	1	0.796
SVM(KNN)	0.786	0.800	0.796	0.796	1

Table 5 shows that the diversity among the SVM classifiers is generally not high enough (correlation coefficient between 0.767 and 0.835). In this case, the local fusion of the SVM classifiers leads to a marginal improvement in the classification performance (0.741), in comparison with the best performing base classifier (0.74).

Table 7 shows that the diversity among the PNN classifiers is generally not high enough (correlation coefficient between 0.820 and 0.920). In this case, the local fusion of the PNN classifiers leads to the unchanged

Table 6 Accuracy of each classifier without feature selection in local classifier fusion

Accuracy	All Features	All Features	All Features	All Features	All Features	Local Classifier Fusion
DT (Quinlan, 1993)	74.4%	74.4%	74.6%	75.4%	75.6%	77.2%
KNN (Aha et al, 1991)	81.3%	81.5%	81.4%	82.2%	81.6%	79.4%
SVM (Platt, 1999)	74.7%	75.5%	75.1%	74.2%	73.9%	73.8%
PNN (Berthold and Diamond, 1998)	78%	77.5%	77.3%	77.4%	77.6%	77.8%
FR (Berthold, 2003)	78.4%	80.3%	79.1%	78.4%	80%	79.1%
RF (Cutler et al, 2012)	94.2%	94.5%	94.6%	94.7%	94.5%	94.9%
GBT (Friedman, 2000)	85.5%	86.6%	85.8%	85.6%	87.5%	85.1%

Table 7 Pearson correlation coefficient of each PNN classifier trained through 5 selected feature sets

Correlation	PNN (DT)	PNN (FR)	PNN (RF)	PNN (PNN)	PNN (KNN)
PNN(DT)	1	0.912	0.920	0.820	0.846
PNN(FR)	0.912	1	0.919	0.842	0.852
PNN(RF)	0.920	0.919	1	0.834	0.853
PNN(PNN)	0.820	0.842	0.834	1	0.85
PNN(KNN)	0.846	0.852	0.853	0.85	1

classification performance (0.791), in comparison with the best performing base classifier (0.791).

Table 8 Pearson correlation coefficient of each FR classifier trained through 5 selected feature sets

Correlation	FR (DT)	FR (FR)	FR (RF)	FR (PNN)	FR (KNN)
FR(DT)	1	0.845	0.857	0.814	0.859
FR(FR)	0.845	1	0.867	0.804	0.846
FR(RF)	0.857	0.867	1	0.808	0.829
FR(PNN)	0.814	0.804	0.808	1	0.818
FR(KNN)	0.859	0.846	0.829	0.818	1

Table 8 shows that the diversity among the FR classifiers is generally not too low (correlation coefficient between 0.804 and 0.867). In this case, the local fusion of the FR classifiers leads to a slight improvement of the classification performance (0.836), in comparison with the best performing base classifier (0.821).

Table 9 shows that the diversity among the RF classifiers is generally not high (correlation coefficient between 0.953 and 0.967). In this case, the local fusion of the RF classifiers only leads to a marginal improvement in the classification performance (0.953), in comparison with the best performing RF classifier (0.951).

Table 10 shows that the diversity among the GBT classifiers is generally not too low (correlation coefficient between 0.825 and 0.838). In this case, the local fusion of the GBT classifiers leads to advances in the classification performance (0.878), in comparison with the best performing GBT classifier (0.866).

Table 9 Pearson correlation coefficient of each RF classifier trained through 5 selected feature sets

Correlation	RF (DT)	RF (FR)	RF (RF)	RF (PNN)	RF (KNN)
RF(DT)	1	0.967	0.963	0.958	0.963
RF(FR)	0.967	1	0.963	0.956	0.965
RF(RF)	0.963	0.963	1	0.953	0.957
RF(PNN)	0.958	0.956	0.953	1	0.955
RF(KNN)	0.963	0.965	0.957	0.955	1

Table 10 Pearson correlation coefficient of each GBT classifier trained through 5 selected feature sets

Correlation	GBT (DT)	GBT (FR)	GBT (RF)	GBT (PNN)	GBT (KNN)
GBT(DT)	1	0.827	0.832	0.828	0.838
GBT(FR)	0.827	1	0.832	0.836	0.830
GBT(RF)	0.832	0.832	1	0.833	0.842
GBT(PNN)	0.828	0.836	0.833	1	0.825
GBT(KNN)	0.838	0.83	0.842	0.825	1

The accuracy obtained through the global fusion of classifiers without feature selection is 89.6%, which decreased by 1.8%, comparing to the accuracy of global fusion with feature selection. Table 6 shows the results on the accuracy of the local fusion of classifiers and the one of each base classifier trained on the same feature set. While the accuracy of local fusion of DT and that of RF increased slightly by 1.6% and 0.2%, respectively, the accuracy of the other local fusion all decreased, and KNN dropped most with 2.8%. Comparing with Table 2, performances of classifiers in the same local ensemble are more similar due to the case that classifiers are trained on the same feature set. For instance, the standard deviation of accuracies of 5 DT classifiers trained on distinct feature sets is 0.89, which is higher than the standard deviation obtained on the same feature set, which is 0.52. Similarly, comparing to the worst performing one of the DT classifier, the improvement of accuracy through local fusion of the DT classifiers trained on different feature sets (10.2%) is higher than the one obtained through local fusion of the DT classifiers trained on the same feature set (2.8%). Moreover, through comparing the results shown in Tables 3 and 8, we can see that the performance achieved through local fusion of classifiers is improved, when using feature selection in comparison with using the original feature set, no matter which one of the seven learning algorithm is used for training base classifier. In other words, the performance of local fusion could be improved by training classifiers on distinct feature sets.

Table 11 Accuracy evaluation of our approach compared to the state-of-the-art methods

Algorithm	Accuracy
Class segment set (Sun and Super, 2005)	90.9%
CS+SP (Bai et al, 2009)	96.6%
BCF+SVM (Wang et al, 2014)	97.16%
BSCP (Shen et al, 2016)	98.41±0.44%
Our Approach	84.1%

Comparisons of accuracy derived from various methods on MPEG-7 shape dataset are shown in Table 11. In order to make a fair comparison with other algorithms, half training validation, which uses 50% data for training and the remaining 50% for testing, is adopted in our experiment. In terms of fusion, two methods were introduced in Boln-Canedo and Alonso-Betanzos (2018). The first one is by training base classifiers separately on distinct feature sets selected by using different learning algorithms, and then adopting fusion of these classifiers, which is the main strategy of our proposed framework. The second one is adopting the fusion of distinct feature subsets after feature selection and then training classifiers on the finally fused feature subset. Our approach focuses on creation of the diversity in distinct feature subsets obtained by using various wrapper based feature selection methods, From a granular computing perspective, the experimental results demonstrate that the proposed framework achieved effectively an improvement of the classification accuracy by creating diversity among classifiers trained using a same learning algorithm on distinct feature subsets. However, the significance level of the performance improvement can be affected by the quality of features extracted. Some algorithms extract high-level features to describe shapes such as bag of contour fragment (BCF) which are capable of better ability to describe shapes leading to advanced performance. The relatively undesirable performance in our approach is probably due to the fact that features used in this experiment can not represent shapes completely, which limit the space for improvements of performance. Accuracy of our approach could be further improved if deeper feature extraction and fusion are investigated in future.

Overall, the results shown in Tables 2-10 indicate that the adoption of wrapper based feature selection driven by different learning algorithms leads to creation of diverse feature subsets and provides the potential of training diverse classifiers on the selected feature subsets using the same learning algorithm. However, this depends on the characteristics of learning algorithms, e.g., some algorithms may be insensitive to the changes to the feature sets leading to very similar classifiers trained on different feature subsets.

Moreover, it is the key to create diversity among classifiers so that the fusion of classifiers is more likely to lead to an improvement of the classification performance. For example, the diversity among the base classifiers trained using DT is the much higher than the diversity among the base classifiers trained using any other algorithm, which results in the most significant improvement of the performance through fusion of the DT classifiers.

5 Conclusion

In this paper, we have proposed a multi-level framework of fusion of features and classifiers in the setting of granular computing. In particular, we have designed to extract features in different ways and adopt the wrapper based feature selection driven by different learning algorithms for obtaining diverse subsets of fused features, in order to enable creation of diversity among classifiers trained on these feature subsets. We have also adopted different learning algorithms for investigating local fusion of classifiers trained using each of the selected learning algorithms and exploring the potential of diversity creation based on different learning strategies of these algorithms.

The experimental results have shown that the local fusion of classifiers trained using the same learning algorithm generally leads to advances in the classification performance. The diversity analysis also indicates that fusion of classifiers is likely to lead to an improvement of the classification performance as long as the classifiers show high diversity to each other and none of the classifiers shows very different performance from the others. The same claim also applies to the case of global fusion of all the primary (local) ensembles.

In future, it is worth to investigate in more depth the diversity creation through deep feature extraction and selection in the setting of multi-granularity learning (Liu and Cocea, 2017a, 2018). Moreover, it is also worth to investigate the effectiveness of adopting the proposed framework of ensemble learning in the context of multi-attribute decision making (Xu and Wang, 2016; Liu and You, 2017; Chatterjee and Kar, 2017; Lee and Chen, 2008; Zulueta-Veliz and Garca-Cabrera, 2018), and incorporate fuzzy set theory related techniques (Zadeh, 1965; Wang and Chen, 2008; Chen et al, 2012, 2009; Chen and Chen, 2011; Chen and Tanuwijaya, 2011; Chen and Chen, 2001; Chen and Chang, 2011; Chen et al, 2013) into the proposed framework to achieve fuzzy ensemble learning (Nakai et al, 2003).

Acknowledgements This work was supported by the National Key R&D Program of China (No.2018YFB1308302), the National Natural Science Foundation of China (61573356) and the Natural Science Foundation of Hebei Province (No. F2016203211).

References

Aha DW, Kibler D, Albert MK (1991) Instance-based learning algorithms. Machine Learning 6(1):37–66

Arjun P, Mirnalinee TT (2018) An efficient image retrieval system based on multi-scale shape features. Journal of Circuits Systems and Computers 27(11):1850,174

Bai X, Liu W, Tu Z (2009) Integrating contour and skeleton for shape classification. In: 2009 IEEE 12th International Conference on Computer Vision Workshops, ICCV Workshops, pp 360–367

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

Bermejo P, Gmez J, Puerta J (2014) Speeding up incremental wrapper feature subset selection with naive bayes classifier. Knowledge-Based Systems 55:140–147

Berthold MR (2003) Mixed fuzzy rule formation. International Journal of Approximate Reasoning 32(2):67 – 84

Berthold MR, Diamond J (1998) Constructive training of probabilistic neural networks. Neurocomputing 19(1):167 – 183

Boln-Canedo V, Alonso-Betanzos A (2018) Ensembles for feature selection: A review and future trends. Information Fusion 52

Chatterjee K, Kar S (2017) Unified granular-number-based ahp-vikor multi-criteria decision framework. Granular Computing 2(3):199–221

Chen SJ, Chen SM (2001) A new method to measure the similarity between fuzzy numbers. In: IEEE International Conference on Fuzzy Systems, Melbourne, Australia, pp 1123–1126

Chen SM, Chang YC (2011) Weighted fuzzy rule interpolation based on ga-based weight-learning techniques. IEEE Transactions on Fuzzy Systems 19(4):729–744

Chen SM, Chen CD (2011) Handling forecasting problems based on high-order fuzzy logical relationships. Expert Systems with Applications 38(4):3857–3864

Chen SM, Tanuwijaya K (2011) Fuzzy forecasting based on high-order fuzzy logical relationships and automatic clustering techniques. Expert Systems with Applications 38(12):15,425–15,437

Chen SM, Wang NY, Pan JS (2009) Forecasting enrollments using automatic clustering techniques and fuzzy logical relationships. Expert Systems with Applications 36(8):11,070–11,076

Chen SM, Munif A, Chen GS, Liu HC, Kuo BC (2012) Fuzzy risk analysis based on ranking generalized fuzzy numbers with different left heights and right heights. Expert Systems with Applications 39(7):6320–6334

Chen SM, Chang YC, Pan JS (2013) Fuzzy rules interpolation for sparse fuzzy rule-based systems based on interval type-2 gaussian fuzzy sets and genetic algorithms. IEEE Transactions on Fuzzy Systems 21(3):412–425

Cutler A, Cutler DR, Stevens JR (2012) Random Forests, Springer US, Boston, MA, pp 157–175

Dash M, Liu H (1997) Feature selection for classification. Intell Data Anal 1(3):131–156

Ding M, Antani S, Jaeger S, Xue Z, Candemir S, Kohli M, Thoma G (2017) Local-global classifier fusion for screening chest radiographs. In: SPIE Medical Imaging

Ding W, Gao X, Su Y, Li X (2015) A movement smoothness measuring algorithm based on curvature estimation. Chinese High Technology Letters 25(8):835–842

Ding W, Wang X, Liu H, Hu B (2018) An empirical study of shape recognition in ensemble learning context. In: 2018 International Conference on Wavelet Analysis and Pattern Recognition (ICWAPR), Chengdu, China, pp 256–261

Friedman J (2000) Greedy function approximation: A gradient boosting machine. The Annals of Statistics 29

Giangreco Maidana A, Legal-Ayala H, Schaerer C, Villamayor-Venialbo W (2018) Contour-point signature shape descriptor for point correspondence. International Journal of Image and Graphics 18:1850,007

Guan H, Antani SK, Long LR, Thoma GR (2010) Comparative study of shape retrieval using feature fusion approaches. Proceedings of the IEEE Symposium on Computer-Based Medical Systems pp 226–231

Hu MK (1962) Visual pattern recognition by moment invariants. Information Theory Ire Transactions on 8(2):179–187

Jovi A, Brki K, Bogunovi N (2015) A review of feature selection methods with applications. In: 2015 38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), Opatija, Croatia, pp 1200–1205

Kuncheva L (2002) A theoretical study on six classifier fusion strategies. Pattern Analysis and Machine Intelligence, IEEE Transactions on 24:281–286

Kurnianggoro L, Wahyono, Jo KH (2018) A survey of 2d shape representation: Methods, evaluations, and future research directions. Neurocomputing 300:1 – 16

Lee LW, Chen SM (2008) Fuzzy multiple attributes group decision-making based on the extension of topsis method and interval type-2 fuzzy sets. In: Proceedings of the 2008 International Conference on Machine Learning and Cybernetics, Kunming, China, vol 6, pp 3260–3265

Lin C, Pun CM, Vong CM, Adjeroh D (2015) Efficient shape classification using region descriptors. Multimedia Tools and Applications 76

Lin G, Zhu H, Kang X, Fan C, Zhang E (2013) Multifeature structure fusion of contours for unsupervised shape classification. Pattern Recognition Letters 34(11):1286–1290

Liu H, Cocea M (2017a) Fuzzy information granulation towards interpretable sentiment analysis. Granular Computing 2(4):289–302

Liu H, Cocea M (2017b) Granular computing based approach for classification towards reduction of bias in ensemble learning. Granular Computing 2(3):131–139

Liu H, Cocea M (2018) Granular Computing Based Machine Learning: A Big Data Processing Approach. Springer, Berlin

Liu H, Cocea M, Ding W (2018) Multi-task learning for intelligent data processing in granular computing context. Granular Computing 3(3):257–273

Liu P, You X (2017) Probabilistic linguistic todim approach for multiple attribute decision-making. Granular Computing 2(4):332–342

Lu G, Sajjanhar A (1999) Region-based shape representation and similarity measure suitable for content-based image retrieval. Multimedia Systems 7(2):165–174

Malo S, Freeman H (2009) On the encoding of arbitrary geometric configurations. Electronic Computers Ire Transactions on EC-10(2):260–268

Mohandes M, Deriche M, Aliyu SO (2018) Classifiers combination techniques: A comprehensive review. IEEE Access 6:19,626–19,639

Mokhtarian F, Bober M (2003) Robust Image Corner Detection Through Curvature Scale Space. Springer Netherlands, Dordrecht

Nakai G, Nakashima T, Ishibuchi H (2003) A fuzzy ensemble learning method for pattern classification. Journal of Japan Society for Fuzzy Theory and Intelligent Informatics 15(6):671–681

Pedrycz W (2011) Information granules and their use in schemes of knowledge management. Scientia Iranica 18(3):602–610

Pedrycz W, Chen SM (2011) Granular Computing and Intelligent Systems: Design with Information Granules of Higher Order and Higher Type. Springer, Heidelberg

Pedrycz W, Chen SM (2015a) Granular Computing and Decision-Making: Interactive and Iterative Approaches. Springer, Heidelberg

Pedrycz W, Chen SM (2015b) Information Granularity, Big Data, and Computational Intelligence. Springer, Heidelberg

Platt J (1999) Fast training of support vector machines using sequential minimal optimization. Advances in Kernel Methods: Support Vector Learning pp 185–208

Priyanka S, Sudhakar M (2018) Tetrakis square tiling-based triangulated feature descriptor aiding shape retrieval. Digital Signal Processing 79:125 – 135

Quinlan R (1993) C4.5: Programs for Machine Learning, vol 1

Reed Teague M (1980) Image-analysis via the general-theory of moments. Journal of the Optical Society of America (1917-1983) 69:1468

Sekita I, Kurita T, Otsu N (1992) Complex autoregressive model for shape recognition. IEEE Transpattanalmachintell 14(4):489–496

Shen W, Jiang Y, Gao W, Zeng D, Wang X (2016) Shape recognition by bag of skeleton-associated contour parts. Pattern Recognition Letters 83

Sun KB, Super BJ (2005) Classification of contour shapes using class segment sets. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), vol 2, pp 727–733 vol. 2

Sun QS, Zeng SG, Liu Y, Heng PA, Xia DS (2005) A new method of feature fusion and its application in image recognition. Pattern Recognition 38(12):2437 – 2448

Thakoor N, Gao J, Jung S (2007) Hidden markov model-based weighted likelihood discriminant for 2-d shape classification. Trans Img Proc 16(11):2707–2719

Wan L, Hong Y, Huang Z, Peng X, Li R (2018) A hybrid ensemble learning method for tourist route recommendations based on geo-tagged social networks. International Journal of Geographical Information Science pp 1–22

Wang A, An N, Chen G, Li L, Alterovitz G (2015) Accelerating wrapper-based feature selection with k-nearest-neighbor. Knowledge-Based Systems 83

Wang HY, Chen SM (2008) Evaluating students' answerscripts using fuzzy numbers associated with degrees of confidence. IEEE Transactions on Fuzzy Systems 16(2):403–415

Wang X, Feng B, Bai X, Liu W, Latecki LJ (2014) Bag of contour fragments for robust shape classification. Pattern Recognition 47(6):2116 – 2125

Xu Z, Wang H (2016) Managing multi-granularity linguistic information in qualitative group decision making: an overview. Granular Computing 1(1):21–35

Yang J, yu Yang J, Zhang D, feng Lu J (2003) Feature fusion: parallel strategy vs. serial strategy. Pattern Recognition 36(6):1369 – 1381

Yao J (2005a) Information granulation and granular relationships. In: IEEE International Conference on Granular Computing, Beijing, China, pp 326–329

Yao Y (2005b) Perspectives of granular computing. In: Proceedings of 2005 IEEE International Conference on Granular Computing, Beijing, China, pp 85–90

Yerima S, Sezer S (2018) Droidfusion: A novel multilevel classifier fusion approach for android malware detection. IEEE Transactions on Cybernetics PP:1–14

Zadeh L (1965) Fuzzy sets. Information and Control 8(3):338–353

Zhang D, Lu G (2004) Review of shape representation and description techniques. Pattern Recognition 37(1):1 – 19

Zulueta-Veliz Y, Garca-Cabrera L (2018) A choquet integral-based approach to multiattribute decision-making with correlated periods. Granular Computing 3(3):245–256