MULTIPLE CLASSIFIER COMBINATION FOR TARGET IDENTIFICATION FROM HIGH RESOLUTION REMOTE SENSING IMAGE

Peijun Du*, Hao Sun, Wei Zhang

Department of Remote Sensing and Geographical Information Science,
China University of Mining and Technology, Xuzhou City Jiangsu Province 221116, P.R.China
dupjrs@cumt.edu.cn, sunhaocumt2000@163.com

Commission I, IV, VII, WG I/2, I/4, IV/2, IV/3, VII/2

KEY WORDS: Multiple Classifier Combination, High resolution remote sensing, Target identification, Boosting, Bagging, Classifier selection, Non-linear combination

ABSTRACT:

Target identification from high resolution remote sensing image is a common task for many applications. In order to improve the performance of target identification, multiple classifier combination is used to QuickBird high resolution image, and some key techniques including selection and design of member classifiers, classifier combination algorithm and target identification methods are investigated. A classifier ensemble is constructed at first, consisting of seven member classifiers: Decision Tree Classifier (DTC) and NaiveBayes classifier, J4.8 decision tree classifier, simple classifier OneR, IBK classifier, feed-forward Neural Network (NN) and Support Vector Machine (SVM). Weighted Count of Errors and Correct results (WCEC) measure is used to select five classifiers for further combination. DTC, J4.8, NN, SVM and IBK are selected and their independence and diversity are evaluated. Some standard MCS methods, such as Boosting, Bagging, linear combination and non-linear combination are experimented to extract road from QuickBird image. The results show that multiple classifier combination can improve the performance of image classification and target identification.

1. INTRODUCTION

Target identification and extraction from remote sensing imagery is one of the most important problems in the integration of Remote Sensing and Geographic Information System (GIS). Traditional low and medium resolution images can't be used to GIS database updating effectively because their resolution can't match the requirement of data precision and details in GIS, but the occurrence of high spatial resolution remote sensing images provides a new way for solving this problem. Despite their high resolution and fine description to ground objects, target identification methods from high resolution remote sensing image are still faced with such difficulties as vast data size, strong impacts of background and noises, and uncertainty in extraction process (Shi et al, 2001).

Recently, multiple classifier system (MCS) has been widely used in a variety of fields as a hot topic of pattern recognition, and multiple classifier combination or classifier ensemble has been introduced to remote sensing information processing(Kittler et al, 2000; Bo et al, 2005; Briem et al, 2002; Mathieu et al, 2006; Benediktsson et al, 2007;).

In this paper, multiple classifier combination is used to target identification from high resolution remote sensing images in order to reduce the non-object noises and enhance the accuracy and confidence of target identification.

2. MULTIPLE CLASSIFIER COMBINATION

2.1 Basic Concept

Multiple classifier combination can be explained briefly as: to derive the final classification decision by integrating the output of multiple learning machines according to a certain combination approach (Xie et al, 2006). In pattern recognition and classification, the algorithm that is effective for one feature set may be unsuitable to other feature sets, and multiple classifiers can provide the complementary information about the classified pattern on hand, so multiple classifier combination may outperform any individual classifier by integrating the advantages of various classifiers. Usually multiple classifiers are organized by two schemes: parallel and concatenation connection (Lv et al, 2000). According to the output information of member classifier, classifier combination can be categorized into three levels: abstract level, rank level and measurement level (Xu et al, 1992). For the target identification from high resolution remote sensing images, the scheme of parallel combination based on abstract level is used.

2.2 Selection of Member Classifier

The performance of multiple classifier system is closely related with member classifiers and their combination strategy, so it is important to decide how to select classifiers from classifier ensemble and how to combine them (Kang et al, 2005). In order to simplify the process, we assume that the number of classifiers selected is a fixed odd number, which is useful for majority vote combination. Here the number is assumed as 5, so there are 21

* Corresponding author: Peijun Du, dupjrs@cumt.edu.cn, dupjrs@gmail.com.
The authors gratefully acknowledge the support of K.C.WONG Education Foundation, Hong Kong.

schemes when 5 classifiers are selected from a set with 7 classifiers. The Weighted Count of Errors and Correct results (WCEC) measure is used to assess the performance of classifier combination.

The classification result based on any two classifiers can be divided into four parts:

1. Samples correctly classified by both classifiers (C1 and C2), marked as a;
2. Samples correctly classified by the first (C1) and incorrectly by the second (C2), marked as b;
3. Samples incorrectly classified by the first (C1) and correctly by the second (C2), marked as c;
4. Samples incorrectly classified by both classifiers (C1 and C2), marked as d.

The results are illustrated in Table 1(Matti et al, 2006).

	C2 right	C2 wrong
C1 right	a	b
C1 wrong	c	d

Table 1. Notation used in the dichotomous outcome for two classifiers

Weighted Count of Errors and Correct results (WCEC) take both correct and incorrect results into consideration and gives suitable weights on them. If there are two classifiers, the Equation is:

$$wcec=a+\frac{1}{2}(b+c)-d_{different}-5d_{same} \tag{1}$$

Where, $d_{different}$ stands for the number of samples incorrectly classified by both classifiers with different errors, and d_{same} represents the number of samples incorrectly classified by both classifiers, but with the same classification results.

This indicator was pairwise measure, and for the whole combination the averaged value over all pairs of classifiers is computed.

2.3 Linear combination methods

Assume M classes exist on a remote sensing image, and $C_1 \cup C_2 \cup \cdots C_i \cdots \cup C_M$, $i \in \{1,2,\ldots,M\}$. For K independent classifiers e_k(k=1,2…K), the output of each classifier was assumed as j_k(k=1,2…K). For abstract level output, the output of member classifier is a class label, which means each classifier provides a label and the labels of multiple classifiers are combined further, usually majority vote is used. If the output of a classifier is the rank of one pixel belonging to every class, the combination is named as rank level combination. If the output of the member classifier can depict the quantitative degree or probability of one pixel belonging to a certain class, for example, posterior probability, and the quantitative index is used to combining multiple classifiers, it is measurement level combination.

For any input feature set x, E(x) = j was assumed as the classification result by MCS (multiple classifiers system). So the linear combination methods can be described as:

$$E(x)= a_1 \bullet j_1(x)+ a_2 \bullet j_2(x)+ a_3 \bullet j_3(x)+\cdots+ a_k \bullet j_k(x) \tag{2}$$

Where, $j_k(x)$ means the output of the *k*th classifier based on the input feature set *x*. a_k(k=1…K) is the weight of the output $j_k(x)$ (Zhou et al, 2006).

There are many linear combination methods such as voting combination, weighted summation combination, consensus theory combination.

2.4 Non-linear combination methods

Recently some non-linear combination methods were investigated and the experiment shown that some non-linear combination methods also got quite good performance (Sun et al, 2001), such as D-S evidence theory and fuzzy integral.

D-S evidence theory assigns probability to sets and can handle the uncertainty caused by unknown factors. D-S evidence theory uses discrimination framework, confidence function, likelihood function and probability allocation function to represent and process knowledge. Suppose that $\Theta=\{C_1,C_2,\cdots C_i,\cdots C_M\}$ is discrimination framework and M is the number of classes, therefore basic probability allocation function m is a function from 2^{Θ} to [0, 1] and it meets the requirement of:

$$\begin{cases} m(\phi)=0 \\ \sum\limits_{A\subseteq\Theta} m(A)=1 \end{cases} \tag{3}$$

If there are two or more different evidences, orthogonal sum can be used to combine those evidences. Assume that Z_1, $Z_2 \ldots Z_n$ are those probability allocation functions corresponding to evidence F_1, F_2, and F_n, and their orthogonal sum $Z=Z_1 \oplus Z_2 \ldots\oplus \ldots \oplus Z_n$ is:

$$Z(\phi)=0 \tag{4}$$

$$Z(A)=K^{-1}\times \sum_{\cap A_i} \prod_{1\le i\le n} Z_i(A_i) \tag{5}$$

$$K=\sum_{\cap A_i\ne\phi} \prod_{1\le i\le n} Z_i(A_i) \tag{6}$$

When various evidences are inconsistent or contradictory each other, the combined result of D-S evidence may be unreasonable (Liu et al, 2003). A modified evidence combination algorithms was proposed and experimented by Sun et al, and it proved that the modified method was superior to traditional method while processing those evidences with high contradiction and inconsistency (Sun et al, 2000). For remote sensing image, different classifier may generate different classified labels, which result in the generation of evidence with high contradiction, so the modified evidence combination is applied to classification integration of high resolution remote sensing images. The detailed equations are as follows (Sun et al, 2000):

$$k_{ij} = \sum_{\substack{A_i \cap A_j = \phi \\ A_i \in F_i, A_j \in F_j}} Z_i(A_i) Z_j(A_j) \tag{7}$$

$$\tilde{k} = \frac{1}{n(n-1)/2} \sum_{i<j} k_{ij} \tag{8}$$

$$\varepsilon = e^{-\tilde{k}} \tag{9}$$

$$Z(A) = p(A) + K^* \varepsilon^* q(A), A \neq \Phi, \Theta \tag{10}$$

$$Z(\Theta) = p(\Theta) + K^* \varepsilon^* q(\Theta) + k(1-\varepsilon) \tag{11}$$

$$p(A) = \sum_{\substack{A_i \in F_i \\ \cap_{i=1}^{n} A_i = A}} Z_1(A_1) Z_2(A_2) \cdots Z_n(A_n) \tag{12}$$

$$q(A) = \frac{1}{n} \sum_{i=1}^{n} Z_i(A) \tag{13}$$

Where, ε is the confidence of evidence, $\tilde{k}$ is the average of contradiction level between two evidences, and K is the total contradiction level of all evidences. This evidence combination method can reduce the limitations caused by high evidence inconsistency.

For multiple classifier combination of remote sensing, the classifier result of each classifier can be viewed as a piece of evidence. Probability allocation function can be represented by the classification accuracy of specific class. For example, if a pixel is classifier to the ith class, the basic probability is: $m(C_i)=P_i$, $m(\Theta)=1- P_i$, where P_i is the accuracy of the ith class by the specific class. After evidence combination being completed, the class with maximum evidence is selected as the final result.

2.5 Boosting and Bagging

Bagging is the abbreviation of Bootstrap Aggregating. In this algorithm, n samples are selected at random from a set with k samples, and instructive iteration is exerted to create some different bags, and every bag is classified by vote to predict its class.

The steps are:

1. For the kth ($k=1,\cdots,k_{max}$) iteration, random sampling is conducted to training sample set and a certain number or proportion of samples are selected and then classified, by which the classification result C_k is stored.
2. For all classification results $C_1, \cdots C_k$, voting is used to integrate the results and predict the final result.

Similar to Bagging, Boosting is also based on the manipulation to training samples. Boosting can process data with weight, so the weights of misclassified samples are increased to concentrate the learning algorithm on specific samples. The detailed steps are:

1. Initialization to assign identical weight to all samples.
2. For the kth ($k=1, \cdots ,k_{max}$) iteration, samples are selected based on weights to generate the kth training sample set that is used to train the kth classifier C_k. The error e is derived by those weighted samples, and the classifier is terminated if e equals to 0 or greater than 0.5 and then turn to step 5.
3. The weights of misclassified samples are increased and those of correct samples are decreased based on the results of classifier C_k.
4. If k is smaller than the biggest iteration number k_{max}, turn to 2, otherwise turn to 5.
5. Stopping iteration.
6. The results of all individual classifiers are summarized based on their weights to generate the final classification result.

3. EXPERIMENTS

3.1 Experiment data

In this experiment, the multi-spectral QuickBird image (spatial resolution is 2.44m) of Xuzhou City, Jiangsu Province, China, is used as the case study image. Training and test samples are selected by ocular interpretation and field investigation (Table 2).

Multiple features are proposed to be used in high resolution image processing owing to the mutual complementation of different features (Lin et al, 2005; Mou et al, 2004). In our experiment, multiple features including gray and spectral vector, vegetation index and texture are used in order to describe and extract objects effectively.

class	training samples	test samples
water	306	83
vegetation	325	165
road	609	143
building	396	146
bare land	126	97
shadow	173	76

Table 2. Information about training samples and test samples

3.2 Experiment flow

Firstly, multi-spectral image is preprocessed and useful features are extracted.

Then the identical training sample set is used to train those member classifiers including decision tree classifier (DTC) and NaiveBayes classifier, J4.8 tree classifier, OneR classifier, IBk classifier, layered feed-forward neural network(NN) (Activation: Logistic, Training Threshold Contribution: 0.9, Training Rate: 0.2, Training Momentum: 0.9, Training RMS Exit Criteria: 0.1, Number of Hidden Layers: 1, Number of Training Iterations: 500)and support vector machine (SVM)(Kernel type: Radial Basis Function, Gamma in Kernel Function: 0.03, Penalty Parameter: 100.00, Pyramid levels: 0).

Thirdly WCEC evaluation criterion is used to select the optimal classifier combination, and this optimal combination is then used to multiple classifier system to extract the object of interest.

Finally targets are recognized based on geometric feature and knowledge.

3.3 Selection of Member Classifier

According to the land cover of study area and task of target identification, the class category consists of six classes: water, vegetation, road, building, bare land and shadow. The same training samples and test samples are used to individual classifier evaluation, and Table 3 is the accuracy of all classifiers.

index	Total accuracy	Kappa	Accuracy rank
DTC	89.5775%	0.8728	1
NaiveBayes	81.8310%	0.7781	6
J4.8	88.7324%	0.8624	2
OneR	74.2254%	0.6841	7
IBk	81.9718%	0.7793	5
NN	85.0704%	0.8179	3
SVM	83.8028%	0.8021	4

Table 3. The performance of individual classifiers

The WCEC measure is used to assess the performance of classifier combination and the results are listed in Table 4.

Member classifiers	WCEC values	Member classifiers	WCEC values
1-2-3-4-5	0.8932	1-3-4-5-7	0.8093
1-2-3-4-6	0.7881	1-3-4-6-7	0.6873
1-2-3-4-7	0.8011	1-3-5-6-7	0.7788
1-2-3-5-6	0.8666	1-4-5-6-7	0.8042
1-2-3-5-7	0.8711	2-3-4-5-6	0.8455
1-2-3-6-7	0.7704	2-3-4-5-7	0.8516
1-2-4-5-6	0.8853	2-3-4-6-7	0.7444
1-2-4-5-7	0.8868	2-3-5-6-7	0.8255
1-2-4-6-7	0.7951	2-4-5-6-7	0.8458
1-2-5-6-7	0.8649	3-4-5-6-7	0.7533
1-3-4-5-6	0.7980		

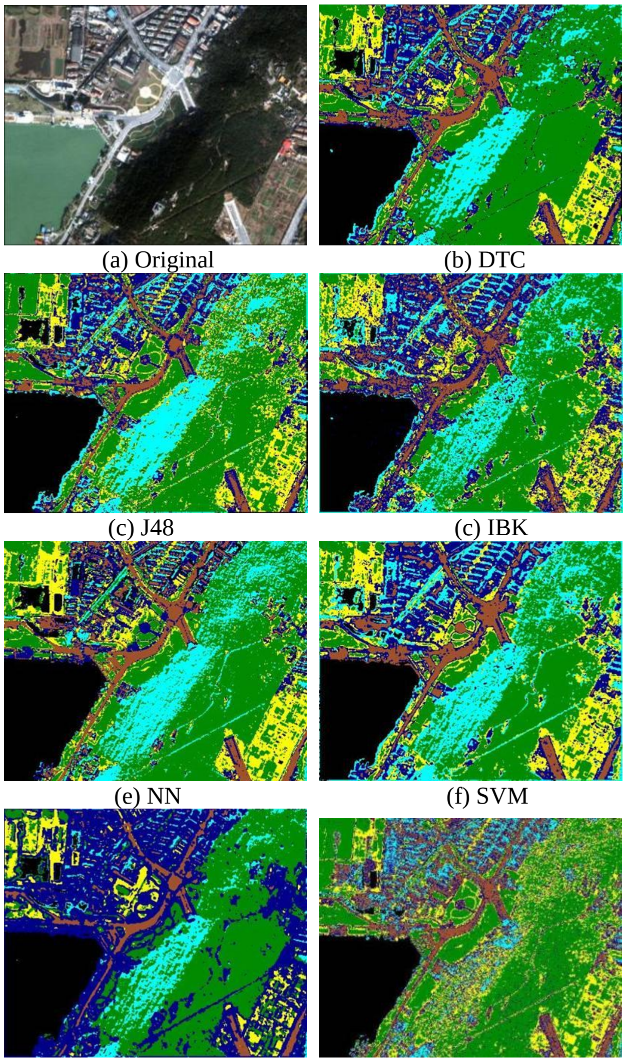
Table 4. WCEC measures of different classifier combination schemes

Note: 1 denotes to DTC, 2 denotes J4.8, 3 denotes to NN, 4 denotes to SVM and 5 denotes to IBk, 6 denotes to NaiveBayes, 7 denotes OneR.

From Table 4, it is easy to found that the combination of DTC, J4.8, NN, SVM and IBK has the biggest WCEC value, so their combination is the best one and used to further target identification.

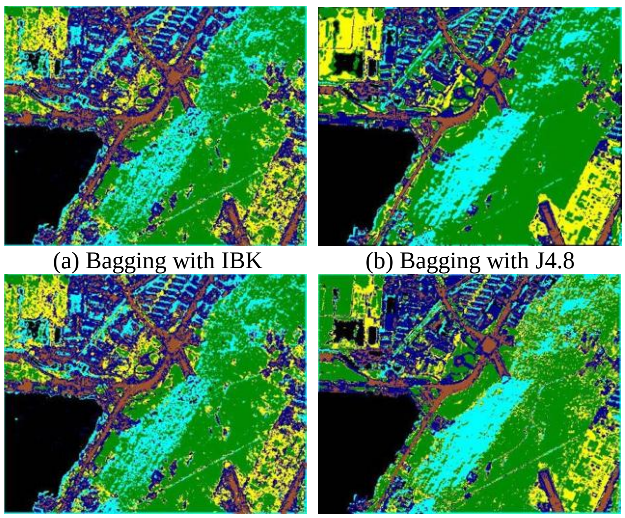
3.4 Target Identification

The experiment results are illustrated in Figure 1~4. Table 5 is the total accuracy and kappa coefficient of individual classifier and multiple classifier system. It can be found that the combination of multiple classifiers can enhance the classification and identification accuracy to a great extent.



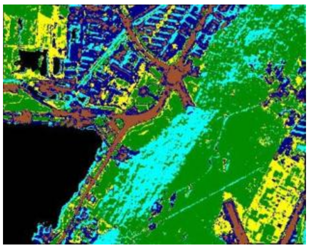
(a) Original (b) DTC
(c) J48 (c) IBK
(e) NN (f) SVM
(g) NaiveBayes (f) OneR

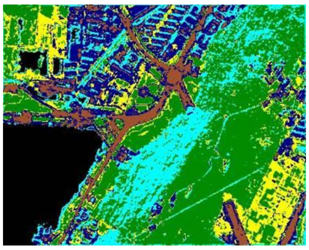
Figure 1. Experiment result of Member Classifiers



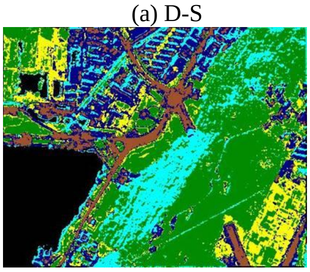
(a) Bagging with IBK (b) Bagging with J4.8
(c) Boosting with IBK (d) Boosting with J48

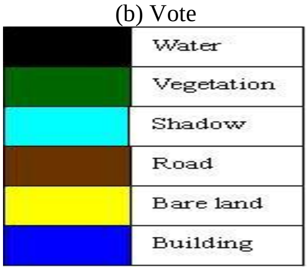
Figure 2. Experiment of Boosting and Bagging





(a) D-S (b) Vote





(c) Weighted summation (d) Legend

Figure 3. Experiment of Linear/Non-Linear Combination

	Classifier	Overall accuracy	Kappa coefficient
Member classifier	DTC	89.5775%	0.8728
	J4.8	88.7324%	0.8624
	NN	85.0704%	0.8179
	SVM	83.8028%	0.8021
	IBK	81.9718%	0.7793
Linear combination	voting combination	90.1408%	0.8797
	Weighted summation combination	90.1408%	0.8797
Non-linear combination	D-S evidence theory	90.1408%	0.8797
bagging	Bagging with J4.8	90.5634%	0.8847
	Bagging with IBK	82.1127%	0.7811
boosting	Boosting with J4.8	89.7183%	0.8744
	Boosting with IBK	81.9718%	0.7793

Table 5. The classification accuracy by different combination method

In order to extract the target of interest, the classification results should be changed to binary image at first, and then edge tracing is conducted to the binary image, and geometric rules and prior knowledge are used to identify the targets. For example, if the target is circular building, the regions with low circular degree should be rejected. Some other geometric features include area, perimeter, rectangle degree, circle degree, central moment, centroid and so on (Inglada, 2007).

For water and vegetation, there are not special geometric features, so the classification results are used directly. For roads, the shape index is used to the classification results, and Figure 4 is the result of road identification.

Figure 4. The result of road extraction

4. CONCLUSION

Multiple combination system is introduced to target identification from high resolution remote sensing image in this paper, and QuickBrid multi-spectral image is used to conduct a case study in Xuzhou City, China. The whole process, including training and test sample selection, member classifier design, feature extraction, classifier selection and combination strategy determination, is investigated to classify the high resolution image and extract interested targets. Diversity of member classifiers is important to multiple classifier system, and weighted count of errors and correct results (WCEC) is used in this paper. Linear combination, non-linear combination, boosting and bagging combination methods were conducted to identify interested target.

Based on the experiments and discussions in this paper, it can be concluded that multiple classifier combination can play important roles in high resolution remote sensing image classification and target identification by making full use of the abundant and detailed information in high resolution image and integrating the benefits of different classifiers. But there are still many issues for further study, for example, selection of member classifier, optimization of feature sets and determination of combination strategy, which will be emphasized in our future research.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the support of K.C.WONG Education Foundation, Hong Kong. The authors thank the support to this research from National High-Tech 863 Program of China (2007AA12Z162), the Natural Science Foundation of China (40871195), the Program for New Century Excellent Talents in University (NCET-06-0476) and the opening fund of Key Laboratory of Advanced Engineering Surveying of SBSM (ES-SBSM-(07)-01).

REFERENCES

Bo Y. C., Wang J. F., 2005. Combining Multiple Classifiers for Thematic Classification of Remotely Sensed Data. Journal of Remote Sensing, 9 (5), pp.555-564.

Briem G. J., Benediktsson J. A., Sveinsson J.R., 2002. Multiple Classifiers Applied to Multisource Remote Sensing Data. IEEE Transactions on Geoscience and Remote Sensing, 40(10), pp.2291-2300.

Inglada J., 2007. Automatic recognition of man-made objects in high resolution optical remote sensing images by SVM classification of geometric image features. ISPRS journal of photogrammetry and remote sensing, 62, pp.236-248.

J.A. Benediktsson, J. Chanussot and M. Fauvel, 2007. Multiple classifiers in remote sensing: from basics to recent developments, 7th International Workshop on Multiple Classifier Systems (MCS), Czech Republic, pp.501-512.

Kittler J., Roli F., 2000. Multiple Classifier Systems. Springer, pp.1-10.

Kang H. J., Doermann D., 2005. Selection of Classifiers for the Construction of Multiple Classifier Systems. Proceedings of the Eight International Conference on Document Analysis and Recognition, pp.263-268.

Lv Y., Shi P. F., Zhao Y. M., 2000. Voting Principle for Combination of Multiple Classifiers. Journal of Shanghai Jiao Tong University, 34(5), pp.680-684.

Liu C. P., Dai J. F., Zhong W. et al, 2003, Multi-source Remote Sensing Information Based on Fuzzy Evidence Theory. Pattern Recognition and Artificial Intelligence, 6, pp.213-218.

Lin H., Li J. P., Mo D. K., 2005. Information Identification on QuickBird Image. Journal of Image and Graphics, 10(10), pp.1504-1511.

Mathieu F., Chanussot J., Benediktsson J. A., 2006. Decision Fusion for the Classification of Urban Remote Sensing Images. IEEE Transactions on Geoscience and Remote Sensing, 44(10), pp.2828-2838.

Matti A., Jorma L., 2006. Using diversity of errors for selecting members of a committee classifier. Pattern Recognition, 39, pp.608–623.

Mou F. Y., Zhu B. Q., He H. Z., 2004. Multiple Features Based Analysis of High-resolution Remotely Sensed Imager. Bulletin of Surveying and Mapping, 10, pp.4-7.

Shi W. Z., Zhu C. Q., 2001. Road Feature Extraction from Remotely Sensed Image: Review and Prospects. ACTA GEODAETICA ET CARTOGRAPHICA SINICA, 30(3), pp.257-361.

Sun Huai-Jiang, Hu Zhong-Shan, Yang Jing-Yu, 2001. A Study on Combining Multiple Classifiers Based on Evidence Theory, CHINESE J.COMPUTERS, 24(3), pp.231-236.

Sun Q., Ye X.Q., Gu W.K., 2000. A New Combination Rule of Evidence Theory. Acta Electronica Sinica, (8), pp.117-119.

Xie Z. X., Yu D. R., Hu Q. H., 2006. Study on sensor fault tolerance with multi-classifiers of SVM fusion. Journal of Harbin Engineering University, 27(Supp), pp.389-293.

Xu L., Krzyzak A., Suen C. Y., 1992. Methods of Combining Multiple Classifiers and Their Applications to Handwriting Recognition. IEEE Transaction on Systems, Man, and Cybernetics, 22(3), pp.418-435.

ZHOU Wei, MA Hong-cha, 2006. Combining methods and diversity measures in multiple classifier systems for thematic classification of landsat TM image. Journal of Heilongjiang Institute of Technology, 20(4), pp.14-21.