2014 IEEE Congress on Evolutionary Computation (CEC)
July 6-11, 2014, Beijing, China

# Neural Network Ensembles For Image Identification Using Pareto-optimal Features

Wissam A. Albukhanajer, Yaochu Jin and Johann A. Briffa
Department of Computing
Faculty of Engineering and Physical Sciences
University of Surrey
Guildford, Surrey, GU2 7XH, United Kingdom
Email: w.albukhanajer@surrey.ac.uk; yaochu.jin@surrey.ac.uk; j.briffa@surrey.ac.uk

***Abstract*—In this paper, an ensemble classifier is constructed for invariant image identification, where the inputs to the ensemble members are a set of Pareto-optimal image features extracted by an evolutionary multi-objective Trace transform algorithm. The Pareto-optimal feature set, called Triple features, gains various degrees of trade-off between sensitivity and invariance. Multilayer perceptron neural networks are adopted as ensemble members due to their simplicity and capability for pattern classification. The diversity of the ensemble is mainly achieved by the Pareto-optimal features extracted by the multi-objective evolutionary Trace transform. Empirical results show that the general performance of proposed ensemble classifiers is more robust to geometric deformations and noise in images compared to single neural network classifiers using one image feature.**

## I. INTRODUCTION

IMAGE identification techniques are widely used in many applications where accuracy and speed are highly important, such as copyright protection [1], face authentication [2] and object recognition [3]. In these applications, effective feature extraction and classification techniques become essential, especially when noise and image deformation such as rotation, scale and translation (RST) are present in the images.

Over the years, a variety of feature extraction techniques have been developed. The Trace transform [4] has been of interest as a feature extraction algorithm due to its attractive ability to represent images in an invariant way. Recent work aimed at improving the feature extraction process using Trace transform was reported in [5]–[8]. Following effective feature extraction, a powerful classification algorithm is also required to accomplish the identification task.

Much research experience has been shown that classification techniques that combine more than one classifier are an effective way to enhance classification performance. These techniques are widely known as *ensemble classifiers*, *mixture of experts* or *committee of learners* [9]. A typical ensemble classifier combines outputs of multiple base classifiers, called *members* or *individual learners* to make a final decision. Over the past decades, an increasing attention has been paid to ensemble classifiers due to their success in improving robustness and accuracy for solving complex machine learning tasks [10]–[16]. Hansen and Salamon [10] introduced an ensemble of identical neural networks trained on subsets of the same database. In that case, generalization is maintained by using different training sets for each neural network, which has shown to produce better results than using the same training set for all nets. On the other hand, Bazell and Aha [11] designed an ensemble of classifiers using a combination of neural networks, naive Bayes and decision trees classifiers for automated morphological galaxy classification using single training set as well as several bootstrap datasets with 10-fold cross-validation. In addition, an ensemble of several neural networks has been built for handwritten digital number recognition [15], where the network outputs are combined using majority voting. To achieve generalization and fault-tolerance, Yu *et al.* [16] proposed an effective Radial Basis Function (RBF) neural network ensemble to improve foreign exchange asset management and investment decision-making.

Generally, members (base classifiers) in an ensemble can use the same type or different types of models. Therefore, ensemble techniques can be largely divided into two categories, *homogeneous* and *heterogeneous* ensembles. Homogeneous ensembles consist of members with the same type of models generated using the same or similar learning algorithms. By contrast, heterogeneous ensembles have members of different types of models [17], [18]. Usually, an ensemble performs no worse than a single member learner under certain conditions [19]. However, the performance of ensembles is heavily dependent on the accuracy and diversity of their members [10], [11]. Having both accurate and diverse members is key to the success of ensembles [20]. Bhowan *et al.* [21] developed a Pareto-based multi-objective genetic programming framework to evolve diverse and accurate classifiers for imbalanced data problem. The accuracy of the minority and majority class was used as learning objectives where some classes represented by a fewer training samples while other classes have more training samples.

In this paper, an ensemble consisting of multilayer perceptron (MLP) classifiers is generated for invariant image identification. The main idea is to utilize diverse features created using an evolutionary multi-objective Trace transform algorithm. Three Pareto-optimal features around the knee point in the Pareto front are employed as inputs to three base classifiers, each of which is a neural network classifier. The Pareto-

This work is supported by EPSRC (CASE) award studentship.

optimal features are all invariant to geometric deformations and robust to additive noise, yet they are different by virtue of the randomness in the selection process. Thus, even if no particular measures are taken to obtain diversity in the base neural network classifiers, diversity in the output of the ensemble members can be guaranteed.

The remainder of this paper is organized as follows. The Pareto-based evolutionary multi-objective Trace transform is described in Section II. Section III presents details of the proposed ensemble classifier. Section IV gives the evaluation of the proposed method and the experimental results. Conclusions and future work are given in Section V.

## II. Evolutionary Multi-Objective Trace Transform

Given a two-dimensional image $I(x, y)$ with $N \times N$, real-valued pixels, a new representation of $I$ can be determined by applying a specific functional $T$, called *Trace* functional, along lines crossing the image $I$ characterized along different orientations $\theta$ and distance $\rho$. The new representation of $I$ is a two-dimensional array that is a function of two variables $\rho$ and $\theta$. This process of converting the image $I$ from the spatial domain $(x, y)$ into the transformed domain $(\rho, \theta)$ is called the *Trace transform* [4]. Several Trace functionals can be used to produce different transforms. If an appropriate functional is used, very similar features may be extracted from the original image regardless geometric transformations applied to the original image. When the $T$ functional is represented by the line integral function, the Trace transform is equivalent to the Radon transform [22], which is a popular tool in curve detection in digital image processing and in reconstruction of tomography images [23]. When applying the *Diametric* functional along the columns of the array produced in the previous step (i.e along $\rho$ dimension), a sequence of numbers will be generated with a length of $n_\theta$ elements, where $n_\theta$ is the number of orientations considered for tracing lines. Finally, applying an additional functional $C$, called *Circus* functional on the sequence of numbers obtained in the previous step will produce a real number, called a *Triple feature*, which can characterize the original image $I$ in an invariant way [4].

Kadyrov and Petrou [4] demonstrated that the traditional Trace transform (TT) functionals can produce Triple features invariant to different geometric transformations applied to the image. Therefore, choosing the correct Trace functionals and its combinations with the Diametric and Circus functionals is an important task to produce Triple features capable of discriminating an image among many other images. Following this idea, we used an evolutionary multi-objective algorithm to search for the optimal combinations of the functionals to produce invariant Triple features [5], which is termed Evolutionary Trace Transform (ETT). To achieve robustness to noise, an improved version of the ETT, called ETTN, has been proposed in [8]. The main difference between ETT and ETTN is that in ETTN method, two extra samples for each class are distorted by Gaussian noise and added during the process of finding the nondominated solutions. In the evolutionary optimization of the Trace transform, the first three images (three classes only) from Fish-94 database [4] (see Fig. 1) are chosen at random and five samples in each class are generated (i.e. 15 samples in total). These image samples are then resized to a lower resolution of $64 \times 64$ to save computational time in evolutionary optimization.

In the ETTN, the following two objective functions are minimized:

$$\begin{aligned} &\min\{f_1, f_2\}, \\ &f_1 = S_w, \\ &f_2 = \frac{1}{(S_b + \epsilon)}, \end{aligned} \tag{1}$$

where $\epsilon$ is a small constant to avoid division by zero. $S_w$ and $S_b$ are within-class variance and between-class variance, respectively, which are defined as follows:

$$\begin{aligned} S_w &= \sum_{k=1}^{K} \sum_{j=1}^{C_k} (\Xi_{jk} - \mu_k^{\Xi})^2, \\ S_b &= \sum_{k=1}^{K} (\mu_k^{\Xi} - \mu^{\Xi})^2, \end{aligned} \tag{2}$$

where

$$\mu_k^{\Xi} = \frac{1}{C_k} \sum_{j=1}^{C_k} \Xi_{jk}, \ \mu^{\Xi} = \frac{1}{K} \sum_{k=1}^{K} \mu_k^{\Xi}$$

and $K$ is the number of classes and $C_k$ the number of samples in class $k$. $\mu_k^{\Xi}$ is the mean of class $k$ of $\Xi$ Triple features, $\Xi_{jk}$ is the $j^{th}$ Triple features of class $k$, and $\mu^{\Xi}$ is the mean of all classes of $\Xi$ Triple features. It is worth mentioning that this problem could be reduced to a single-objective optimization (by multiplying the two objectives, $f_1$ and $f_2$). However, it has been shown [24] that more than one objective should be considered in most optimization and learning problems.

The elitist non-dominated sorting genetic algorithm, NSGA-II for short [25], has been adopted in ETTN. In NSGA-II, a two-step tournament selection is performed on solutions depending on its dominance. At the first step, solutions are ranked according to their fitness values and arranged in an ascending order, i.e. solutions with a lower rank is preferred. If two solutions have the same rank, then it is arranged according to a crowding distance and solutions with a larger crowding distance are preferred. Table I lists the parameter setup for the evolutionary algorithm.

After running the evolutionary algorithm for 200 generations, a Pareto front consisting of six Pareto-optimal solutions is obtained, see Fig. 2. Each solution on the Pareto front is a distinct combination of the $T$, $D$, and $C$ functionals, which

TABLE I: NSGA-II Parameters Set-up

| Parameter | Value |
|---|---|
| Population size $N_p$ | 150 |
| Mutation probability $P_m$ | 0.125 |
| Crossover probability $P_c$ | 0.9 |
| Number of generations | 200 |
| $\epsilon$ | $10^{-5}$ |

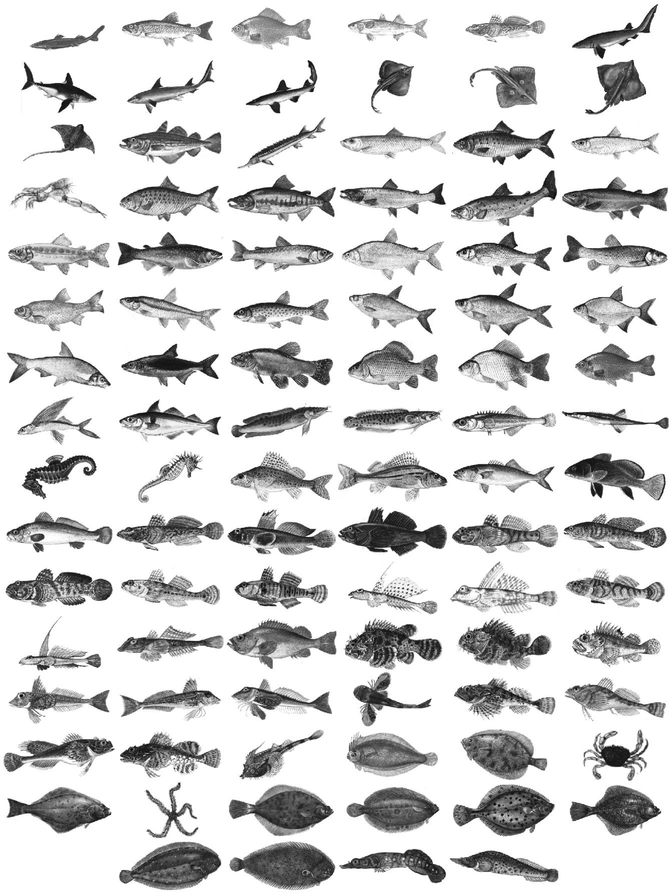


Fig. 1: Fish-94 database.

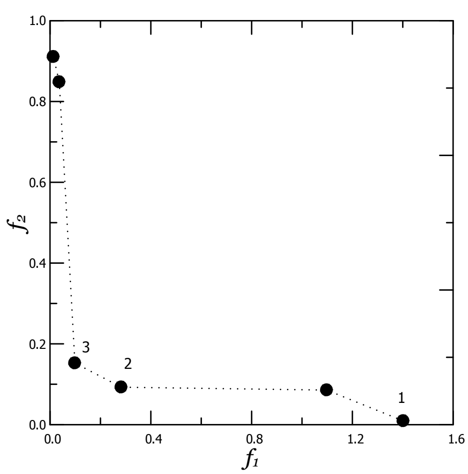



Fig. 2: Pareto front composed of the corresponding solutions representing functionals combinations of ETTN.

are to be used to extract Triple features from the images. To this end, some solutions need to be selected and evaluated on the entire Fish-94 database. We performed several runs of the multi-Objective optimization method and found out that the final Pareto-front contains approximately the same solutions. Solutions around the knee point of the Pareto front such as solutions 2 and 3 in Fig. 2 are selected because, as suggested in [24], these solutions can achieve the best trade-off between two objectives. Intuitively, solutions having a larger between-class variance (minimum $f_2$) are preferred if their within-class variance is adequately small to construct features for efficient classification. Therefore, we selected an additional solution with a minimum $f_2$, i.e., solution 1. In general, it is the user's preference to choose solutions from the Pareto front, which is one of the main advantages of using multi-objective optimization over a single objective optimization. The Pareto-optimal solutions used in this work are listed in Table II, and the corresponding functionals are described in Table III. A full list of $T$, $D$ and $C$ functionals that are used in the entire evolutionary stage can be found in [8]. From these solutions, we create three pairs of Triple features, termed ETTN1, ETTN2 and ETTN3, respectively. The three pairs of features are shown in Table IV, each of which will be used to extract image features to be inputted to a member classifier in the ensemble.

TABLE II: Triple Features Combinations From Evolutionary Trace Transform Algorithm (ETTN) in Fig. 2

| Solution No. | Triple Features |
|---|---|
| 1 | $T_0D_5C_5$ |
| 2 | $T_0D_3C_2$ |
| 3 | $T_0D_1C_2$ |

TABLE III: Functionals Description.

| Functional | Description[1] |
|---|---|
| $T_0$ | $\sum_{i=0}^{n_t-1} \tau_i$ |
| $D_1$ | $\max_{i=0}^{n_\rho-1} \delta_i$ |
| $D_3$ | $\left(\sum_{i=0}^{n_\rho-1} \lvert\delta_i\rvert^4\right)^{\frac{1}{4}}$ |
| $D_5$ | $\max_{i=0}^{n_\rho-1} \delta_i - \min_{i=0}^{n_\rho-1} \delta_i$ |
| $C_2$ | $\sqrt{\frac{1}{n_\theta} \sum_{x=0}^{n_\theta-1} (\xi_i - M)^2}, M = \frac{1}{n_\theta} \sum_{i=0}^{n_\theta-1} \xi_i$ |
| $C_5$ | $\max_{i=0}^{n_\theta-1} \xi_i - \min_{i=0}^{n_\theta-1} \xi_i$ |

[1] $n_t$ is the total number of samples along the tracing line, $\tau_i$ is the value of the $i^{\text{th}}$ pixel along the tracing line, $n_\rho$ is the total number of elements along the columns of Trace matrix, $\delta_i$ is the value of the $i^{\text{th}}$ sample along the columns of the Trace matrix, $n_\theta$ is the total number of elements in the row direction of Trace matrix (the number of orientations considered), and $\xi_i$ is the value of the $i^{\text{th}}$ sample along the row direction of Trace matrix.

## III. An Ensemble Classifier

Traditionally, an ensemble classifier consists of several base learners (similar or different) combined using majority voting, averaging, bagging or stacking [18]. In our work, the base

TABLE IV: Pairs of Triple Features Combinations From ETTN Algorithm

| ETTN feature no. | Triple features pairs |
|---|---|
| ETTN1 | $T_0D_5C_5$, $T_0D_3C_2$ |
| ETTN2 | $T_0D_5C_5$, $T_0D_1C_2$ |
| ETTN3 | $T_0D_3C_2$, $T_0D_1C_2$ |

learners are feed-forward multilayer perceptron (MLP) neural networks trained with the backpropagation algorithm. The reason behind choosing identical MLP neural networks is that this work focuses on the effectiveness of using Pareto-optimal features to create diversity required by ensemble classifiers. Each MLP neural network consists of an input layer, one hidden layer and one output layer. The input layer consists of two neurons receiving Triple features extracted from the images using a corresponding pair of Trace functionals found by the ETTN as discussed in the previous section. The number of neurons in the hidden layer is often determined heuristically, while the number of output neurons equals to the number of classes in the database. For the fish database used in this work, there are 94 classes.

Each base classifier is then trained with its own features extracted from training images. The learning rate for each MLP is 0.3 with training time of 5000 epochs. Each class contains 44 training samples, which include the original image, the rotated, scaled and translated (RST) images. In the training, a 5-fold cross-validation is applied. Fig. 1 shows the 94 original images of the Fish-94 database, which has 94 image classes with size of $200 \times 400$ pixels as in [4] and we converted it to a standard size of $256 \times 256$ pixels. Note that the training samples contain no combined transformations, i.e., the 44 samples of each image used for training the base classifiers are either a rotated, scaled or translated version of the original image. Nevertheless, the test images may be created with a combination of RST transformations and two types of additive noise, namely Gaussian noise and salt & pepper noise. After training, the classifiers are combined together and a final decision is made using majority voting. The test images contain a wide range of geometric transformations and additive noise. Fig. 3 illustrates the structure of the ensemble classifier proposed in this work.

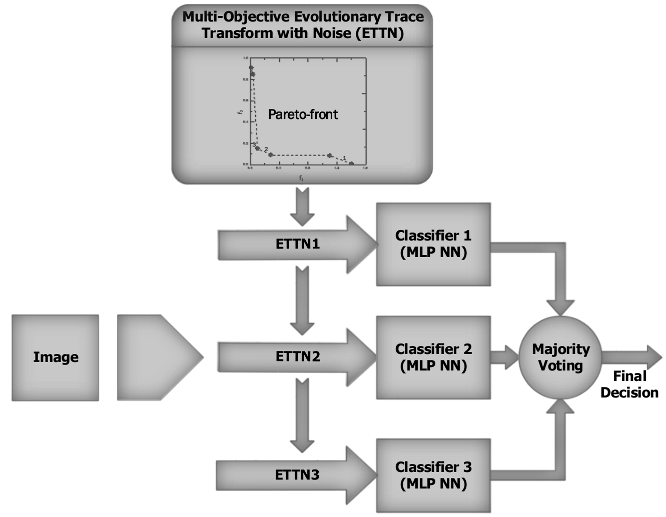



Fig. 3: The structute of the ensemble classifier using a Pareto-optimal feature set.

## IV. Experimental Study

In the experimental study we compare the performance of each base classifier and the ensemble classifier using image features extracted by ETTN1, ETTN2 and ETTN3, as well as the performance of the traditional Trace Transform (TT) described in [4]. In the experiments we present test images generated from the Fish-94 database with random object rotation and translation with a given scale factor and additive noise (single image per class). The experiments are performed on a PC with Intel Core 2 Duo E8500 3.1GHz processor and 3GB of RAM.

First, we test the performance of the classifiers against robustness to scale deformation only (i.e., no rotation, translation or noise). Fig. 4 shows the classification accuracy of the compared classifiers when the test images are scaled with a scaling factor from 0.9 down to 0.1. We can clearly see that the average classification accuracy of the ensemble is much better than that of the member classifiers and the traditional Trace transform. Moreover, the individual members exhibit a competitive performance compared to the traditional Trace transform, despite the fact that the traditional Trace transform uses many features, whereas each member classifier uses a pair of Triple features only. This suggests that compared to the traditional Trace algorithm, the single classifiers using the features extracted by the evolutionary Trace transform can perform comparably well at a much lower computational cost.

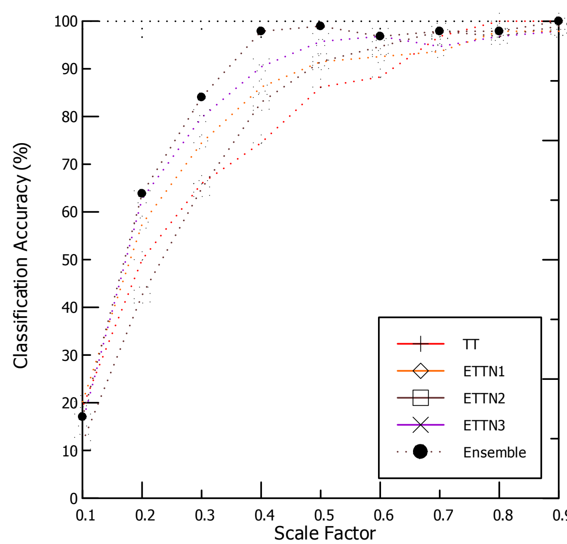
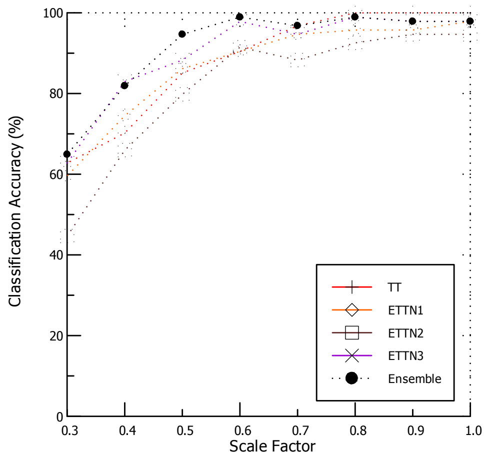
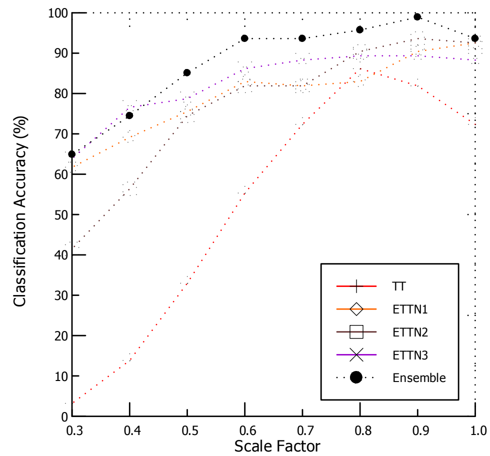


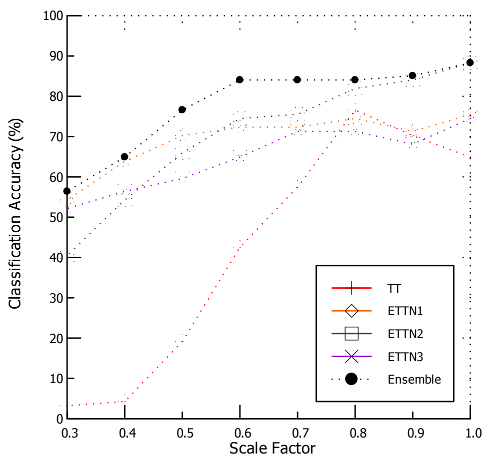

Fig. 4: Robustness to scale only (one test sample per class for each scale factor).



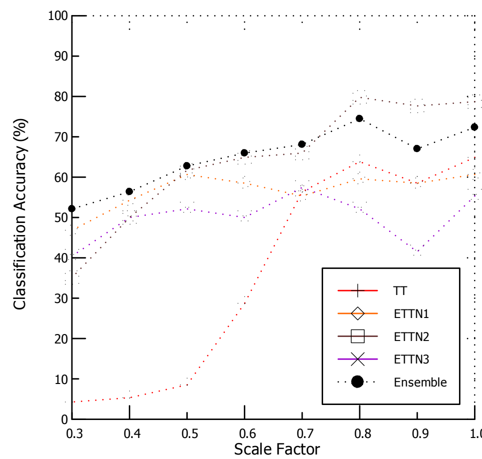

(a) $\sigma^2 = 0$.



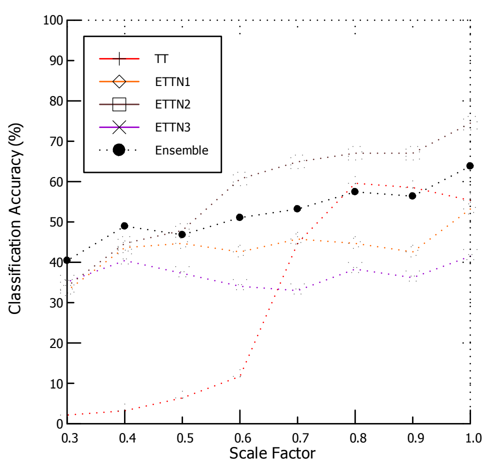

(b) $\sigma^2 = 2$.



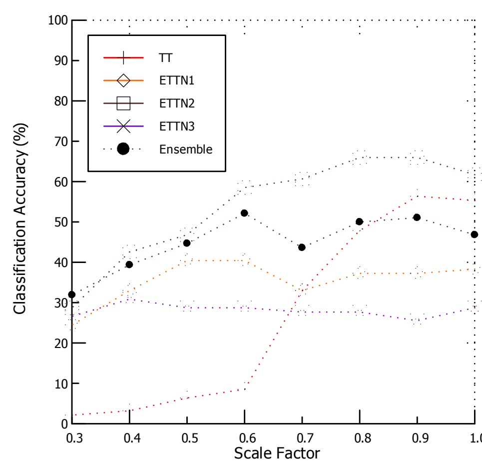

(c) $\sigma^2 = 4$.




(d) $\sigma^2 = 6$.

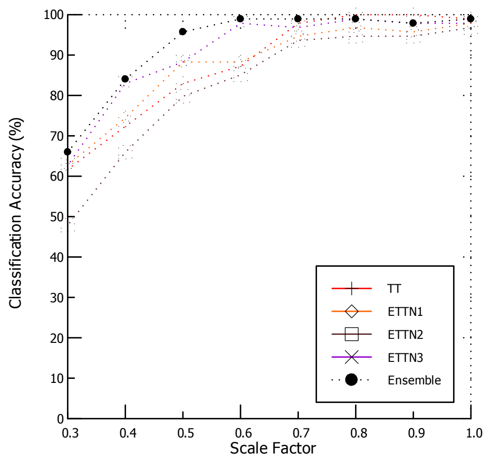



(e) $\sigma^2 = 8$.

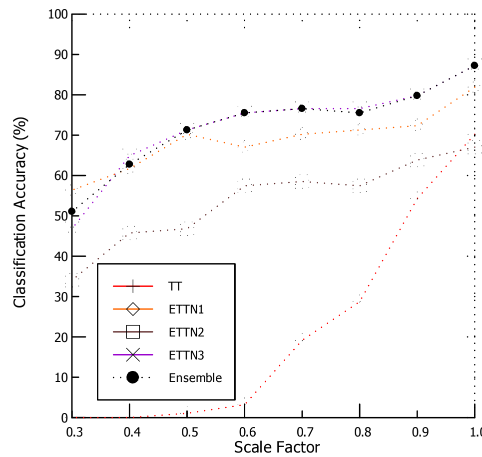



(f) $\sigma^2 = 10$.

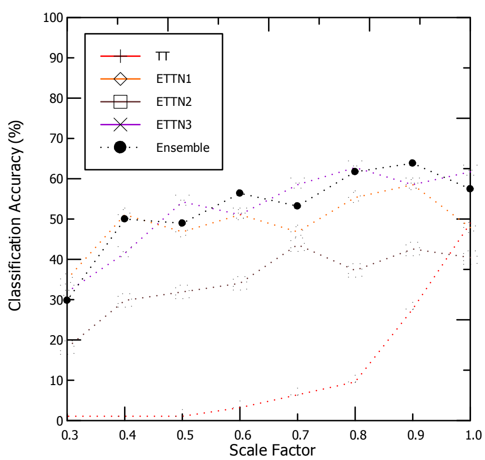
Fig. 5: Robustness to Gaussian noise of zero mean and standard deviation $\sigma^2 = 0$, 2, 4, 6, 8 and 10, of each approach, (Fish-94 database). Performance are shown when the object is scaled from 1 to 0.3, rotated and translated in a random way, and Gaussian noise has been added to the whole image with standard deviation values corresponds to each figure (one test sample per class for each scale factor).



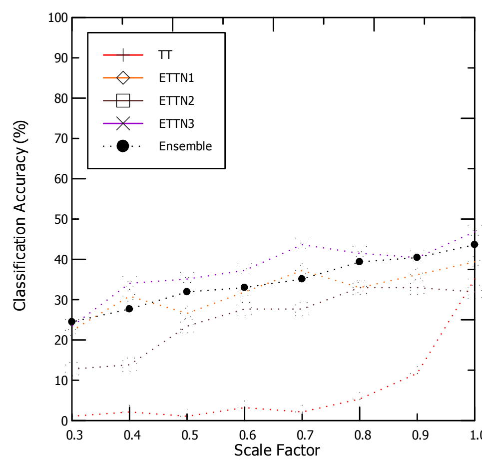

(a) 0% noise.



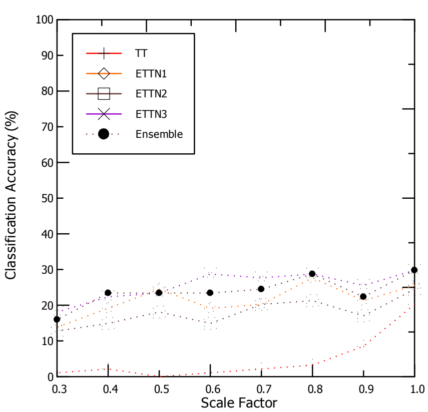

(b) 1% noise.



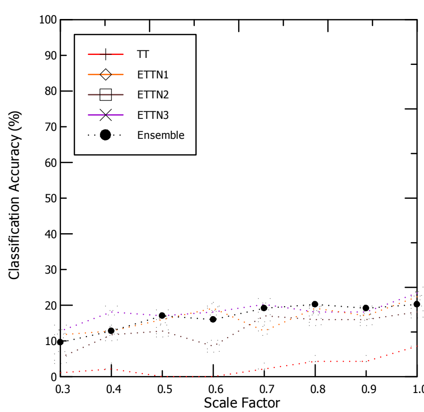

(c) 2% noise.



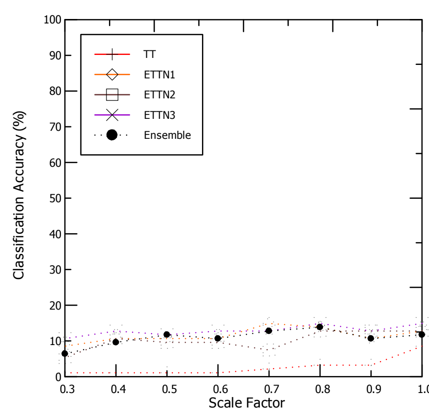

(d) 3% noise.




(e) 4% noise.




(f) 5% noise.




(g) 6% noise.

Fig. 6: Robustness to additive salt & pepper noise with percentage of altered pixels % = 0, 1, 2,...,6 of each approach, (Fish-94 database). Performance are shown when the object is scaled from 1 to 0.3, rotated and translated in a random way, and noise has been added to the whole image with noise levels corresponds to each figure (one test sample per class for each scale factor).

Second, we examine the performance of the compared classifiers using test images having more severe deformations and noise. In these test samples, the images are subject to random rotation and translation with a range of Gaussian noise and a given scaling factor. The Gaussian noise has zero mean and standard deviation ranging from 0 to 10 with an increment of 2. Noise is added to the whole image (i.e. to the object and the background) after RST transformation. Fig. 5 presents the classification accuracy when the test images are subject to Gaussian noise for different scaling factors. In Fig. 5(a), the ensemble classifier is assessed with noiseless test samples deformed by a combination of random rotations and translation. We can see that there is a slight drop in classification accuracy compared to the results on test images transformed with scale only. These results give an idea on how the system holds against RST deformation. However, ETTN2 seems to be less sensitive to the increase in noise (c.f. Figs. 5(b)-(f)). Generally, classifiers ensemble exhibits better average performance than individual (base) classifiers.

Finally, we test the system with image samples polluted with salt & pepper noise. Similar to Gaussian noise, salt & pepper noise is also added to the whole image with a percentage of altered pixels from 0% to 6%. Fig. 6 shows the performance of the compared classifiers when the test images are polluted with salt & pepper noise in addition to random rotation and translation with a given scaling factor. From these figures, the classification accuracy of the traditional Trace transform (TT) deteriorate rapidly compared to the classifiers using features extracted by ETTN1, ETTN2 and ETTN3. ETTN1 appears to be more robust to salt & pepper noise than to Gaussian noise. However, the performance of ETTN2 drops more rapidly with an increasing noise level. In most cases, the ensemble classifier shows better performance throughout the tests with and without additive noise. Although the member classifiers are of a similar MLP structure which has a small diversity, the input features to each member are extracted using Pareto-optimal pairs of functionals chosen by the ETTN. This enhances diversity of the ensemble classifier and thereby the classification performance.

## V. Conclusions

This paper has presented an ensemble classifier using a set of Pareto-optimal Trace transform features. Compared to the traditional Trace transform that uses thousands of features, the single or ensemble classifiers using the features extracted by the evolutionary multi-objective Trace transform are able to accurately classify noisy RST deformed images with a much lower computational cost. Our results indicate that diversity in the Pareto-optimal features can introduce diversity in the ensemble classifier. As a result, no particular effort is needed to generate diverse base classifiers. Future work will consider using different types of base classifiers to further enhance diversity in the ensemble. It is hoped that compared to the ensemble classifier proposed in this work, where the diversity is mainly imposed by the Pareto-optimal features, heterogeneous ensembles using Pareto-optimal features will create an additional level of diversity, thereby increasing the overall accuracy of the ensemble classifiers.

## References


[1] H. Zhu, M. Liu, and Y. Li, "The RST invariant digital image watermarking using Radon transforms and complex moments," *Digital Signal Processing*, vol. 20, no. 6, pp. 1612–1628, 12 2010.

[2] S. Srisuk, "Detection and Recognition Strategies for a Face Authentication System," Ph.D. dissertation, Mahanakorn University of Technology, March 2004.

[3] H. Pan and L.-Z. Xia, "Efficient object recognition using boundary representation and wavelet neural network," *Neural Networks, IEEE Transactions on*, vol. 19, no. 12, pp. 2132–2149, 2008.

[4] A. Kadyrov and M. Petrou, "The trace transform and its applications," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 8, pp. 811–828, 2001.

[5] W. A. Albukhanajer, Y. Jin, J. A. Briffa, and G. Williams, "Evolutionary Multi-Objective Optimization of Trace Transform for Invariant Feature Extraction," in *2012 IEEE Congress on Evolutionary Computation (CEC)*, Brisbane, Australia, 10-15, June 2012, pp. 401–408. [Online]. Available: http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=6256160

[6] A. Frias-Velazquez, C. Ortiz, A. Pizurica, W. Philips, and G. Cerda, "Object identification by using orthonormal circus functions from the trace transform," in *Image Processing (ICIP), 2012 19th IEEE International Conference on*, 2012, pp. 2153–2156.

[7] W. A. Albukhanajer, Y. Jin, J. A. Briffa, and G. Williams, "A comparative study of multi-objective evolutionary trace transform methods for robust feature extraction," in *Evolutionary Multi-Criterion Optimization*, ser. Lecture Notes in Computer Science, R. Purshouse, P. Fleming, C. Fonseca, S. Greco, and J. Shaw, Eds. Springer Berlin Heidelberg, 2013, vol. 7811, pp. 573–586. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-37140-0_43

[8] W. A. Albukhanajer, J. A. Briffa, and Y. Jin, "Evolutionary Multi-objective Feature Extraction in the Presence of Noise," *Supbmitted for publication*, January 2014.

[9] L. I. Kuncheva and C. J. Whitaker, "Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy," *Machine learning*, vol. 51, no. 2, pp. 181–207, 2003.

[10] L. Hansen and P. Salamon, "Neural network ensembles," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 12, no. 10, pp. 993–1001, 1990.

[11] D. Bazell and D. W. Aha, "Ensembles of classifiers for morphological galaxy classification," *The Astrophysical Journal*, vol. 548, no. 1, pp. 219–223, February 2001.

[12] J. Kodovsky, J. Fridrich, and V. Holub, "Ensemble classifiers for steganalysis of digital media," *Information Forensics and Security, IEEE Transactions on*, vol. 7, no. 2, pp. 432–444, 2012.

[13] B. Cyganek and M. Wozniak, "A framework for image analysis and object recognition in industrial applications with the ensemble of classifiers," in *Emerging Technologies Factory Automation (ETFA), 2013 IEEE 18th Conference on*, 2013, pp. 1–4.

[14] M. Roy, D. Routaray, S. Ghosh, and A. Ghosh, "Ensemble of multilayer perceptrons for change detection in remotely sensed images," *Geoscience and Remote Sensing Letters, IEEE*, vol. 11, no. 1, pp. 49–53, 2014.

[15] J. Wang, J. Yang, S. Li, Q. Dai, and J. Xie, "Number image recognition based on neural network ensemble," in *Natural Computation, 2007. ICNC 2007. Third International Conference on*, vol. 1, 2007, pp. 237–240.

[16] L. Yu, K. K. Lai, and S. Wang, "Multistage RBF neural network ensemble learning for exchange rates forecasting," *Neurocomputing*, vol. 71, no. 16-18, pp. 3295 – 3302, 2008, advances in Neural Information Processing (ICONIP 2006) / Brazilian Symposium on Neural Networks (SBRN 2006).

[17] S. Bian and W. Wang, "Investigation on diversity in homogeneous and heterogeneous ensembles," in *Neural Networks, 2006. IJCNN '06. International Joint Conference on*, 2006, pp. 3078–3085.

[18] M. Gashler, C. Giraud-Carrier, and T. Martinez, "Decision tree ensemble: Small heterogeneous is better than large homogeneous," in *Machine Learning and Applications, 2008. ICMLA '08. Seventh International Conference on*, 2008, pp. 900–905.

[19] S. Wang and X. Yao, "Relationships between diversity of classification ensembles and single-class performance measures," *Knowledge and Data Engineering, IEEE Transactions on*, vol. 25, no. 1, pp. 206–219, 2013.

[20] X. Yao and Y. Liu, "Making use of population information in evolutionary artificial neural networks," *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, vol. 28, no. 3, pp. 417–425, 1998.

[21] U. Bhowan, M. Johnston, M. Zhang, and X. Yao, "Evolving diverse ensembles using genetic programming for classification with unbalanced data," *Evolutionary Computation, IEEE Transactions on*, vol. 17, no. 3, pp. 368–386, June 2013.

[22] S. R. Deans, *The Radon Transform and Some of Its Applications*. Krieger Publishing Company, 1983.

[23] P. Toft, "Radon transform: Theory and implementation," Ph.D. dissertation, Technical University of Denmark, Denmark, 1996.

[24] Y. Jin and B. Sendhoff, "Pareto-based multiobjective machine learning: An overview and case studies," *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, vol. 38, no. 3, pp. 397–415, 2008.

[25] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective Genetic Algorithm: NSGA-II," *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 2, pp. 182–197, 2002.