Feature Selection for Finger Knuckle Print-based Multimodal Biometric System

Madasu Hanmandlu

Electrical Engineering Department Indian Institute of Technology, Delhi, India

ABSTRACT

In this paper, feature level fusion of finger knuckle prints (FKP's) is implemented. To overcome the curse of dimensionality, feature selection using the triangular norms is proposed. There has been no effort on feature selection using the t-norms in the literature. In this paper we address the problem of feature selection on the finger knuckle print using the t-norms. An unknown parameter in t-norms is learnt using Reinforced Hybrid evolutionary technique. Feature level fusion is performed by combining the significant features of all FKP's. Results show an improvement in the accuracy when the features are selected by a divergence function derived from the new entropy function using t-norms on two pairs of training features taken at a time. Results of both identification and verification rates show a significant improvement in the performance with feature level fusion.

Keywords

Feature selection, feature level fusion, triangular norms, Finger knuckle print.

1. INTRODUCTION

Feature selection has been a dominant area of research in biometrics, image retrieval, data mining, health care, and text categorization. The main idea of feature selection is to select a set of features, by discarding the insignificant as well as the redundant features that are strongly correlated. Many practical pattern classification tasks (e.g., medical diagnosis) require that a given input pattern (typically represented by a feature vector) be categorized into one of the given set of classes by a classifier. The choice of features affects the accuracy of the classifier and the time required for classification. The goal is to select the minimum subset of features to enhance the classification accuracy and also to shorten the training and execution times.

In this paper, we study the feature selection on FKP's. In [12], an acquisition device is developed to capture the FKP images, and a region of interest is cropped for feature extraction. A feature extraction scheme which amalgamates both the orientation and magnitude information due to Gabor filtering is developed. The local convex direction map of the FKP image is extracted in [12] based on which a local coordinate system is established to align the images. In [13], both local and global information for the FKP verification is exploited, where the orientation information extracted by the Gabor filters is treated as the local feature and by increasing the scale of Gabor filters to infinite, the Fourier transform of the image is obtained, and hence the Fourier transform coefficients of the image can be taken as the global features. The final matching distance of two FKPs is a weighted average of local and global matching distances. In [14], the

Jyotsana Grover

Electrical Engineering Department Indian Institute of Technology, Delhi, India

SIFT (Scale invariant feature transform) is applied after the Gabor enhancement to improve the performance. The finger knuckle identification is undertaken in [15] using the orientation features from the finite Radon transform.

Techniques for feature subset selection can be classified into the following categories: wrapper, embedded, filter and hybrid. The wrapper approach requires one classification model and it selects features with the aim of improving the generalization performance of that model. An algorithm in [1] uses a constructive approach involving correlation information while selecting features and determining Neural Network architectures. A wrapper method is presented in [2] for feature selection, namely IAFN-FS (Incremental Analysis Of Variance and Functional Networks-for Feature Selection). The method uses an induction algorithm called ANOVA and Functional Networks. It follows a backward non-sequential strategy on the complete set of features (thus allowing to discard several variables in one step, and so reducing computational time); and is able to consider "multivariate" relations between features. A new feature selection in [3] for binary classification uses SVM. A wrapper feature selection is presented is presented in [4] with multilayer perceptron and the sequential backward selection (SBS) procedure. Filter models utilize an evaluation function that relies only on the characteristics of the data, and it is independent of any particular learning algorithm. A stochastic algorithm in [5] based on the Greedy Randomized Adaptive Search Procedure (GRASP) uses the well-known filter algorithms such as Relief and FCBF for the constructive phase, and also the classical wrapper feature subset selection algorithms.

A feature selection algorithm in [6] is based on dynamic mutual information wherein, mutual information of each candidate feature is recalculated on unlabeled instances, rather than the whole sampling space. Here it can exactly measure the relevance between candidate features and the class labels by following the selection procedure. A rough set approach for feature selection based on Ant Colony Optimization is proposed in [7], which adopts mutual information based feature significance as heuristic information. The theoretical analysis is presented in [19] on class-augmented principal component analysis (CA-PCA), which is composed of processes for encoding the class information, augmenting the encoded information to data, and extracting features from class-augmented data by applying PCA. In [20], feature selection method based on fuzzy entropy measures is introduced and it is tested together with similarity classifier. A model is implemented in [21] to hybridize the particle swarm optimization and support vector machines to improve the classification accuracy with an appropriate feature subset

This paper is organized as follows: Section 2 describes the feature extraction method for the FKP's. Section 3 discusses about the feature selection method using the t-norms. Section 4 presents the Reinforced PSO-BF hybrid and its utility to the feature selection. Section 5 is devoted to the experimental results. Finally the conclusions are given in Section 6.

2. FEATURE EXTRACTION

The Finger Knuckle Prints (FKP's) are utilized for extracting the features and for the subsequent authentication of persons. The Region of interest (ROI) of FKP is operated on in eight directions using the Sobel compass operator to find edges. The Sobel masks used are shown in Fig. 1 and the results of application of Sobel operator are displayed in Fig. 2.



Fig. 1 3x3 Sobel operators in the eight directions.



Fig. 2: An example of directional responses derived using Sobel operator in the eight different directions and the centered image is the original ROI of FKP.

2.1 Local Binary Pattern

The original LBP operator by Ojala [9] is employed here for the texture analysis of FKP's. This operator assigns a label to every pixel in an image by thresholding the eight neighborhood pixels with respect to the gray value of its center. The feature vector is formed by concatenating the thresholded binary bits in the anti-clockwise direction, shown in Fig. 3.

The LBP code is invariant to any monotonic transformation on the pixel values. Therefore, it is very appropriate for representing the knuckles which are illuminated from a fixed distance but with varying light intensity. An extension to LBP code can be made using a subset of LBP string. An LBP string is called uniform if it consists of at most 2 bit-wise transitions from 0 to 1 or vice-versa. For example, the



1*1 + 1*2 + 0*4 + 1*8 + 1*16 + 0*32 + 1*64 + 0*128 = 91

Fig. 3 Obtaining a Binary label for a pixel

LBP strings 1111110 and 11101111 are uniform whereas 11101011 and 10110101 are non-uniform. There are 58 possible labels of uniform patterns and the rest 198 are non-uniform, which are stored under the 59th label. The operations one of which is shown in Fig. 3 result in 59 labels. A histogram of these labels can be constructed as:

$$H_{l} = \sum \sum \{ L(i, j) = l \}, \qquad l = 1, 2....59 \qquad (1)$$

The histogram of labels is treated as a texture descriptor of the image. Following the approach in [10] the images are divided into N equal sub-windows and the texture descriptor is extracted from each sub-window in eight directions by the application of the Sobel operator. The local texture descriptors are then averaged out to get an average descriptor for each sub-window of the image. These descriptors are then concatenated to form the global texture descriptor. Thus, the texture descriptor for a given FKP will have a size of {59(No. of labels) x N (No. of sub-windows)}.

3. FEATURE SELECTION

3.1 T-Norms

As we want to make use of t-norms [11] for their ability to accentuate the differences between two feature vectors; which is useful in the feature selection, a brief description of t-norms is given here.

A t-norm is a function T: $[0, 1] \times [0, 1] \rightarrow [0, 1]$ that satisfies the following properties:

- Commutativity : T(a, b) = T(b, a)
- Monotonicity: $T(a, b) \le T(c, d)$ if $a \le c$ and $b \le d$
- Associativity: T(a, T(b, c)) = T(T(a, b), c)
- Identity: T(a, 1) = a
- Table 1 gives the three t-norms specially suited for the feature selection and classification. As seen from this table these norms contain a parameter, which can be found either by experimentation or by learning. However, we resort to learning in this work.

Table 1	Different	t-norms
---------	-----------	---------

t-norm	Formulation $T(x,y)$
Hamacher (p>0)	$\frac{xy}{p+(1-p)(x+y-xy)}$
Frank t-norm $p \in]0, +\infty[\setminus\{1\}]$	$\log_p \left[1 + \frac{(p^x - 1)(p^y - 1)}{p - 1} \right]$
Yager (p>0)	$\max(0,1-((1-x)^{p}+(1-y)^{p})^{1/p})$

3.2 Feature Selection based on t-norms

We concatenate the features of all the FKP's resulting in a single feature vector. As we have a large number of features for the representation of FKP's, we are faced with the problem of dimensionality. To mitigate this problem of dimensionality, we go in for the feature selection based on t-norms of divergence.

Divergence: When the two features are used to represent the same pattern, the absolute difference between the features is termed as the divergence. We will now see how the t-norms can be used for the feature selection by the procedure outlined here.

The data needs to be normalized first. The min-max criterion for normalization is initiated here. It is very convenient to form the data matrix consisting of all sample features of a user. Let the data matrix of l^{th} user be denoted by X^{l}

$$X^{l} = \begin{bmatrix} x_{11}^{l} \dots x_{1j}^{l} \dots x_{1n}^{l} \\ \vdots \\ \vdots \\ x_{s1}^{l} \dots x_{sj}^{l} \dots x_{sn}^{l} \end{bmatrix}$$
(2)

where each row of the matrix represents a training sample. The number of normalized features is n and the number of training samples is s. The superscript l denotes the l^{th} user.

The divergence is computed between the two samples of the same user. As we need two such divergences for applying t-norms, these are computed from two pairs of samples as

$$e_{mkj}^{l} = \left| x_{mj}^{l} - x_{kj}^{l} \right| e_{hgj}^{l} = \left| x_{hj}^{l} - x_{gj}^{l} \right|$$
 (3)

Note that the t-norm of e_{nki}^{l} and e_{hei}^{l} in the jth feature gives the spread between the divergences, which bestows the discriminating power. If the jth feature is significant, then both e_{mkj}^{l} and e_{hgj}^{l} will be small and so the t-norm of these divergences will still be smaller. It is now intended to tap this discriminating power of the features by way of the above normed divergences for feature selection. There are two ways of accomplishing this task: One, directly using the normed divergences and second using the entropy of the normed divergences. The uncertainty in the t-normed divergences is necessitated for the sheer discrimination of how important are these from the point of view of feature selection. We will now use Hanman - Anirban entropy [18] to find the uncertainty. The non normalized Hanman-Anirban entropy is defined as: $H(p) = \sum_{i} p_{i} e^{-(ap_{i}^{3} + bp_{i}^{2} + cp_{i} + d)}$ (4)

By taking $p_i = T(e_{mkj}^l, e_{hgj})$, a=0, b=0, and summing over all the users, we get the divergence function for the jth feature denoted by S_i as

$$S_{j} = \sum_{l} \sum_{m,k,h,g} T(e_{jmk}^{l}, e_{jhg}^{l}) \exp(-cT(e_{jmk}^{l}, e_{jhg}^{l}) + d)$$

$$\forall j = 1, 2,, n$$
(5)

To avoid learning and for the ease of computation, we choose c=1 and d=0 in (5). The power of the feature selection comes from the use of this divergence function. If S_j is very low, it means that the corresponding jth feature is significant. So the procedure for the feature selection lies in arranging S_j (j=1,2,...,n) in the ascending order and then taking the leading features by specifying a threshold. The features of all FKP's are concatenated/ fused. Fig.4 depicts the block diagram for the feature level fusion.



Fig. 4 Block diagram for feature level fusion

3.3 Algorithm for feature level fusion

1. Form the data matrix as in (2), $X = \begin{bmatrix} X^1, X^2, \dots, X^n \end{bmatrix}^T$

where n is the total number of features after concatenation.

Each row represents one training sample consisting of the features of all the knuckles (LI, RI, LM, RM).

2. Normalize the data X with n features and m training samples using the min-max criterion.

3. Evaluate the divergence between two jth features using: $e_{mkj}^{l} = \left| x_{mj}^{l} - x_{kj}^{l} \right|$ $e_{hgj}^{l} = \left| x_{hj}^{l} - x_{gj}^{l} \right|$ (6)

where the subscripts m, k, h, and g denote the four samples of the same lth user, e_{mkj}^{l} denotes the divergence in the jth feature between the mth and the kth sample of the lth user. Here we calculate all possible t-norm combinations of divergences. For example, if we have 4 training samples then we obtain ${}^{4}C_{2} = 6$ divergences. While taking t-norms of these divergences we get ${}^{6}C_{2}$ combinations.

- 4. Find the divergence function $S = [S_1, ..., S_j, ..., S_n]$.
- 5. Arrange *S_j*'s in the ascending order by taking the leading features.

3.4 Estimation of parameters of t-norms

T-norms contain an unknown parameter 'p' in the t-norm which needs to be estimated. An optimal value of 'p' can be learned using the Reinforced hybrid BF-PSO, which requires an objective function, which is taken as the identification accuracy. The main aim of this function is to maximize the recognition rate (RR). Given a test sample features the Euclidean distance is computed with respect to the features of all the samples in the training set. The lowest value of this distance associated with the training sample gives the identity of the test sample (i.e. its user). This is repeated for all the test samples and the counts of the number of both successes and failures are made. The objective function (J) is defined as follows:

$$J = (1 - RR)^{2}$$
Where,
$$RR = \frac{No.of \ successes}{Total \ no.of \ tests}$$
(7)

The above function in (7) is evaluated in the proposed Reinforced BF-PSO hybrid described now.

4. REINFORCED HYBRID BF-PSO

The hybrid of Bacterial Foraging (BF) and Particle Swarm Optimization (PSO) in [17] aims at tapping the ability of PSO in exchanging the social information and that of BF in finding a new solution during the elimination and dispersal step. Further, the incorporation of reinforced learning permits us to use the past information. In the BF-PSO hybrid, the velocity term in PSO algorithm becomes the direction for BF. In this hybrid the step size is one of the determining factors for the convergence of the global optima. Because of the fixed step size the hybrid algorithm has several problems [16]:

i) If the step size is very small then it requires a large number of iterations to attain the optimum solution.

ii) If the step size is very high then the bacteria reach the optimum value quickly but the solution suffers from low accuracy.

iii) Bacteria may get stuck up around the initial positions or the local optima, because the dispersion event happens after the specified number of reproduction processes.

These problems are removed in the selection of step size by the reinforcement learning (RL). RL facilitates any algorithm to learn its behavior based on the feedback from the environment. In the case of hybrid BF-PSO, it adjusts the direction of movement of bacteria. The actions performed by an organism become feedback, which in turn is translated into a negative or positive reward for that action. When the bacterium moves to the non-favorable environment it encounters meager source of food. This negative feedback will force the bacterium to change the direction (tumble). In the case of favorable situation the bacterium ventures to move in the same direction (swim) compelled by the positive feedback. Reward of this positive feedback is the right stimulant for the optimal solution.

Reinforcement is a way to exploit the reuse policy of the past information, i.e. the error. It helps accelerate the process of exploitation. Exploitation is still possible without recourse to the reinforcement but is not definitely effective. Any policy utilizing the past information is bound to enhance the exploitation since the current information provides a local view whereas the accumulated information provides a global view. However, how best the accumulated information is made use of is the concern of this investigation. To bring the reinforcement into effect, the past information in the form of error is accumulated for each bacterium. An exploitation mechanism through RL by way of the past information makes the updating strategies of the variables involved in PSO-BF hybrid adaptive. This will be elaborated now.

If P(i, j, k, l) denotes position of the *i*th bacterium in *j*th chemotactic step, k^{th} reproduction step, and *ell*th elimination and dispersal step, the updating strategy in BF-PSO is given by:

Tumble:
$$P(i, j+1, k, l) = P(i, j, k, l) + c(i)V(i)$$
 (8)

Swim: P(i, j, k, l) = P(i, j, k, l) + c(i)V(i) (9)

where c(i) is the step size and V(i) is the velocity of ith bacterium given by:

$$V(i) = w^*V(i) + C_1 * R_1 * (Plbest(i) - Pcurrent(i)) + C_2 * R_2 * (Pgbest(i) - Pcurrent(i))$$
(10)

where *Plbest* is the local best position, *Pgbest* is the global best position and *Pcurrent* is the current position, C_1, C_2 :the fixed parameters of PSO, R_1, R_2 : the random numbers in PSO.

The error is taken as the accumulated average of the absolute difference of the objective functions in the consecutive chemotaxis steps of BF. The step size in the chemotaxis steps is taken as the sigmoid function of the error. The bacteria take the step size depending on the nutrients they get hold of. This is determined by the difference of the current and the previous chemotaxis steps, which leads to the converged optimum solution. When the reinforcement concept is incorporated in the hybrid BF-PSO, the equations of swim and tumble get modified to:

Tumble:

$$P(i, j+1, k, l) = P(i, j, k, l) + \left(\frac{V(i)}{1 + \exp(-error(i))}\right)$$
(11)
Swim: $P(i, j+1, k, l) = P(i, j+1, k, l) + \left(\frac{V(i)}{1 + \exp(-error(i))}\right)$ (12)

Where the step size c(i) is replaced by the sigmoid function of the *error*,

$$error(i) = \frac{\sum_{j} |J(i, j, k, ell) - J(i, j - 1, k, ell)|}{Total no. of chemotaxis steps so far}$$
(13)

where the summation is taken over the chemotaxis steps.

The initial parameters belonging to BF are:

$$n, S, N_c, N_s, N_{re}, N_{ed}, P_{ed}, S_r, c(i)(i = 1, 2...S)$$
, and

those belonging to PSO are : V, C_1, C_2, R_1, R_2 . These parameters stand for n: Dimension of the search space; S: The number of bacteria in the population; S_r : Half the total number of bacteria; N_s : Maximum number of swim length; N_c : Chemotaxis steps; N_{re} : The number of reproduction steps; N_{ed} : Elimination and dispersal events; P_{ed} : probability of Elimination and dispersal; J(i, j, k, ell) is the fitness value of the ith bacterium, at the jth chemotactic step, kth reproduction step, and *ell*th elimination and dispersal event and *Jbest*(*j*, *k*, *ell*) is the fitness of best position in jth chemotactic and kth reproduction step, *ell*th elimination and dispersal step.

Algorithm of Reinforced Hybrid BF-PSO

1. Initialize the random direction V(i), position P(i,1,1,1), error(i) for i=1:S

For $(ell=1 \text{ to } N_{ed})$

For $(k = 1 \text{ to } N_{re})$

For $(j = 1 \text{ to } N_c)$

2. Evaluate the cost function J(i, j, k, ell) for i=1, 2,..., S

3. Save the best cost function for i^{th} bacterium in *Jlast*

Jlast(i, j) = J(i, j, k, ell)

4. The best cost for each bacterium in Step 3 is selected as the local best Jlocal

Jlocal(i,j) = Jlast(i, j)

5. Update the positions using Eq. (11)

6. Evaluate the cost function J(i, j+1, k, ell).

While $(m < N_s)$ If J(i, j+1, k, ell) < Jlast

$$Jlast(i, j) = J(i, j+1, k, ell)$$

7. Update the positions using Eq. (12).

8. Compute the cost function J(i, j+1, k, ell)

9. Evaluate the current position and local cost of each bacterium from:

Pcurrent(i) = P(i, j+1, k, ell)

Jlocal(i, j+1) = Jlast(i, j)

m=m+1; End

i=i+1 (the next bacterium)

10. Obtain the local best position (Plbest) for each bacterium and global best position (Pgbest)

11. Evaluate the new direction for each ith bacterium using Eq. (10).

12. Calculate the accumulated sum of absolute differences of cost functions between the consecutive chemotaxis steps for each bacterium as:

$$sum(i) = \sum_{j} |J(i, j, k, ell) - J(i, j-1, k, ell)|$$

13. Obtain the *error* averaged over the number of chemotaxis steps for each bacterium as

$$error(i) = \frac{sum(i)}{No.of \ chemotaxis \ steps \ so \ far}$$

j=j+1 (the next chemotactic step)

Evaluate $J_{health}^{i} = \sum_{j=1}^{N_{c}+1} J(i, j, k, ell)$ and $S_{r} = \frac{S}{2}$ bacteria

with the highest cost function values die and other S_r bacteria with the best values split.

k=k+1 (the next reproduction step)

14. Eliminate and disperse each bacterium with probability *Ped*.

ell =ell + 1 (the next elimination and dispersal step)

The parameters of Reinforced hybrid learning are specified as: No. of bacteria S=20, No. of step sizes Ns=4, No. of chemotaxis steps Nc=10, No. of reproduction and elimination steps Nre=2, No. of elimination and dispersal steps Ned=2, the corresponding probability Ped=0.2.

5. EXPERIMENTAL RESULTS

The feature selection using t-norms is tested on PolyU FKP database consisting of 165 users with 12 images each of LI, LM, RI, and RM FKP's. LBP features are extracted from each of the LI, LM, RI, and RM FKP's. The ROI is divided into 9 sub windows of equal size and so the texture descriptor evaluated using LBP is of the length 531 (=59*9). The length of feature vector becomes 2124 (=531*4) after concatenation of four FKP's. Since the dimensionality becomes large, we apply feature selection using t-norms.

The database is divided into 6 training and 6 testing images. The genuine and imposter scores are computed based on the Euclidean distance. There are 990 (165*6) genuine scores and 162360 (165*164*6) imposter scores. These scores are then compared with the threshold (varying with the step size of 0.01) and error rates are calculated. If the genuine score exceeds the threshold, then it contributes to the false rejection rate (FRR) which is the ratio of the number of rejected verification attempts to the number of verification attempts by a valid user. If the imposter score is less than the threshold, then it contributes to the False acceptance rate (FAR) which is the ratio of the number of successful fraud

attempts to the total number of fraud attempts by a user. These error rates are used to plot the receiver operating characteristics (ROC) which depicts the performance of an authentication system. The ROC plot is drawn between FAR vs. GAR (GAR= 100-FRR is the genuine acceptance rate), with varying threshold values. From this plot the threshold that yields the highest GAR corresponding to the lowest FAR needs to be selected.

The Reinforced hybrid BF-PSO is used to maximize the recognition rate with respect to the parameter 'p' in t-norms. Fig. 5 compares the performance of the features of the individual knuckles with that of the feature level fusion involving four knuckles after discarding 300 features. If we remove more than 300 features then the results will

deteriorate, so we take 1824(531*4-300) significant features for the fusion. From Fig. 5, it is evident that significant performance improvement in GAR can be achieved as compared to that of individual FKP's representations. Table 2 displays the verification results of feature level fusion with and without the feature selection. Table 3 gives the identification accuracy with the three well known t-norms in addition to proving the estimates of parameter 'p' using Reinforced hybrid BF-PSO learning technique. With the feature selection, significant performance improvement in the verification as well as the identification accuracy can be seen from Tables 2 and3. Table 4 compares the performance of feature selection using t-norms with the other methods like PCA and fuzzy entropy. The t-norms have an edge over other methods.



Fig. 5 A comparison of performance of individual knuckle (LI, RI, LM, RM) with the feature level fusion using Frank t-norm with 300 features removed.

Method	GAR (%) (for FAR= 10^{-2} %)
Without feature selection	93
Frank t-norm (p=0.12)	97
Yager t-norm (p=0.08)	96.5
Hamacher t-norm (p=1.3)	95

Table 2 Verification results of feature level fusion	with	and		
without feature selection.				

Table 4 Comparison of the performance with other feature selection methods

Feature level fusion	GAR (%) (for FAR=10 ⁻² %)
Frank t-norm	97
PCA [19]	94
Fuzzy entropy[20]	92
PSO-SVM [21]	93.65

Table 3 Identification results with and without feature selection.

Feature level fusion	Identification accuracy (%)
Without feature selection	98
Frank t-norm	100
Yager t-norm	100
Hamacher t-norm	99.71

6. CONCLUSIONS

As the features contain the raw information about a pattern, its utilization in the context of multimodal biometric fusion is of utmost importance. But the features are too large to capture the effective information. Hence this study focuses on reducing the insignificant features with the help of divergence function derived from Hanman-Anirban entropy function. The concept of the divergence is necessitated to represent the difference between the features of the training samples of the same user. We need two divergences to be able to apply tnorms for getting t-normed divergences required in the formulation of the divergence function. The significance of the features is evaluated using the divergence function such that the features with the lower values of the divergence function need to be retained.

The implementation of the proposed approach on FKP gives a reduction of 15% features with 4% improvement in the performance. Beyond 15% reduction in the features, the performance is found to deteriorate. Further work is on to link the divergence function with notion of roughness for paving a way to the rough set theory.

7. REFERENCES

- Md. Monirul Kabir, Md. Monirul Islam, Kazuyuki Murase, 2010. A new wrapper feature selection approach using neural network. Neurocomputing, vol. 73, no. 16-18, pp. 3273-3283.
- [2] Noelia Sánchez-Maroño, Amparo Alonso-Betanzos, 2011. Combining functional networks and sensitivity analysis as wrapper method for feature selection. Expert Systems with Applications, vol. 38, no. 10, pp. 12930-12938.
- [3] Sebastián Maldonado, Richard Weber, 2009. A wrapper method for feature selection using Support Vector Machines. Information Sciences, vol. 179, no. 13, pp. 2208-2217.
- [4] Romero, E., Sopena, J.M., 2008. Performing Feature Selection With Multilayer Perceptrons. IEEE Trans. Neural Networks, vol. 19, no. 3, pp. 431-441.
- [5] Pablo Bermejo, Jose A. Gámez, Jose M. Puerta, 2011. A GRASP algorithm for fast hybrid (filter-wrapper) feature subset selection in high-dimensional datasets. Pattern Recognition Letters. vol. 32, no. 5, pp. 701-71, Apr. 2011.
- [6] Huawen Liu, Jigui Sun, Lei Liu, Huijie Zhang, 2009. Feature selection with dynamic mutual information. Pattern Recognition, vol. 42, no. 7, pp. 1330-1339.
- [7] Yumin Chen, Duoqian Miao, Ruizhi Wang, 2010. A rough set approach to feature selection based on ant colony optimization. Pattern Recognition Letters, vol. 31, no. 3, pp. 226-233.
- [8] Susana M.Vieira, João M.C. Sousa, Uzay Kaymak, 2011. Fuzzy criteria for feature selection. Fuzzy Sets and Systems, vol. 189, no. 1, pp. 1-18.
- [9] Timo Ojala , Matti Pietikäinen , Topi Mäenpää , 2002. Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns. IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 24, no.7, pp. 971-987.

- [10] Michael Goh Kah Ong, Tee Connie, Andrew Teoh Beng Jin, 2008. Touch-Less Palm Print Biometric System. 3rd International Conference on Computer Vision Theory and Applications, Madeira – Portugal, pp. 423-430.
- [11] V. Novák, W. Pedrycz, 1988. Fuzzy sets and t-norms in the light of fuzzy logic. International Journal of Man-Machine Studies, vol. 22, no. 2, pp. 113-127.
- [12] Zhang L., Zhang David, Zhu H., 2010. Online fingerknuckle-print verification for personal authentication. Pattern Recognition, vol. 43, no. 7, pp. 2560-2571.
- [13] Lin Zhang, Lei Zhang, David Zhang, Hailong Zhu, 2011. Ensemble of local and global information for fingerknuckle-print recognition. Pattern Recognition, vol. 44, no. 9, pp. 1990-1998.
- [14] Morales, A., Travieso, C.M., Ferrer, M.A., Alonso, J.B., 2011. Improved finger-knuckle-print authentication based on orientation enhancement. Electronics Letters, vol. 47, no. 6, pp. 380-384.
- [15] Ajay Kumar and Yingbo Zhou, 2009. Personal identification using finger knuckle orientation features. Electronic Letters, vol. 45, no. 20, pp. 1023–1025.
- [16] Datta T, Misra I. S., 2008. Improved adaptive bacteria foraging Algorithm in optimization of antenna array for faster convergence. Electromagnetic Research C, vol. 1, pp.143-157, 2008.
- [17] Wael Mansour Korani, 2008. Bacterial foraging oriented by particle swarm optimization strategy for PID tuning. Proceedings of the 2008 GECCO conference companion on Genetic and evolutionary computation, Atlanta, USA ,pp. 1823- 1826.
- [18] M. Hanmandlu and Anirban Das, 2011. Content-based Image Retrieval by Information Theoretic Measure. Defence Science Journal, vol. 61, no. 5, pp. 415-430.
- [19] Myoung Soo Park and Jin Young Choi, 2009. Theoretical analysis on feature extraction capability of class-augmented PCA. Pattern Recognition, vol. 42, no. 11, pp. 2353-2362.
- [20] Pasi Luukka , 2011. Feature selection using fuzzy entropy measures with similarity classifier. Expert Systems with Applications, vol. 38, no. 4, pp. 4600-4607.
- [21] Cheng-Lung Huang and Jian-Fan Dun, 2008. A distributed PSO–SVM hybrid system with feature selection and parameter optimization. Applied Soft Computing, vol. 8, no. 4, pp. 1381-1391, 2008.