Multi-Table Reinforcement Learning for Visual Object Recognition

Monica Piñol, Angel D. Sappa and Ricardo Toledo

Abstract This paper presents a bag of feature based method for visual object recognition. Our contribution is focussed on the selection of the best feature descriptor. It is implemented by using a novel multi-table reinforcement learning method that selects among five of classical descriptors (i.e., Spin, SIFT, SURF, C-SIFT and PHOW) the one that best describes each image. Experimental results and comparisons are provided showing the improvements achieved with the proposed approach.

Keywords Object recognition · Artificial intelligence · Reinforcement learning

1 Introduction

Bag of features (BoF) has become one of the most widely used approaches for visual object recognition (e.g., [1–4]). It consists of four steps. Firstly, it finds the interest points (detectors) and describes them (descriptors) in order to characterize

M. Piñol (✉) · A. D. Sappa · R. Toledo
Computer Vision Center and Computer Science Department, Universitat Autònoma de Barcelona, 08193 Bellaterra, Barcelona, Spain
e-mail: mpinyol@cvc.uab.es
URL: www.cvc.uab.es

A. D. Sappa
e-mail: asappa@cvc.uab.es
URL: www.cvc.uab.es

R. Toledo
e-mail: ricard@cvc.uab.es
URL: www.cvc.uab.es

Mohan S. and S. S. Kumar (eds.), *Proceedings of the Fourth International Conference on Signal and Image Processing 2012 (ICSIP 2012)*, Lecture Notes in Electrical Engineering 221, DOI: 10.1007/978-81-322-0997-3_42,

the given object at a higher abstraction level. Secondly, the extracted feature points, from all the images in the training set, are structured in a kind of dictionary of words. This dictionary of words is obtained through a learning process and will be used during in the next step. Thirdly, each of the images from the training set is represented by means of a histogram that count the number of times a given word appears. Finally, the obtained histograms are used to train a support vector machine (SVM), which classifies the given images. The BoF is a flexible architecture and the fourth steps mentioned above can be implemented through the use of different algorithms. Hence, the final result will depend on the selection of the right algorithm for each step.

From the four steps mentioned above, a particular attention should be given to the first one, since it represents the most sensible and results are highly dependent on the right descriptor selection. Since the use of different descriptors implies different results, one could think that the best option for this first step is to concatenate as much descriptors as possible. Hence, a given image will be represented by all possible different descriptors. Unfortunately, such a kind of *brute-force* strategy in most of the cases is not feasible due to the fact that it introduces noise. Recently, in [5], the authors present a method to select the best descriptor for each image using *reinforcement learning* (RL). RL is a simple method that allows to learn the best action under a *trail-and-error* framework based on a set of user defined *states*. Although interesting results have been obtained in [5], and most of the time the approach converges to the best descriptor, the problem now is how to define a reliable state to be used during the RL learning process. The current work tackles this problem by proposing a strategy that helps to select the state that maximize the result.

RL has been largely used in the robotics community during the last two decades. Recently, it has attracted the attention in the computer vision field to address problems such as image segmentation or object recognition, just to mention a few. For instance, in the segmentation domain the RL method is used to select the appropriate threshold (e.g., [6, 7]). In [8], the authors propose a RL based face recognition technique that is able to learn the best feature from each image. Similarly, in the object recognition field, [9] presents a RL technique using first order logic. Finally, RL has also been used for learning interest points [10, 11] or for selecting methods for classification [12].

In this paper we present a BoF based approach for the object recognition. As mentioned above the current work is focussed on the first step by using RL. More precisely, the current work contributes with a novel scheme for selecting the best state in the RL method. This scheme results in a multi-table formulation. Regarding the rest of steps of BoF, in the current implementation we use a kd-tree in the second step and a support vector machine in the fourth step. The reminder of the paper is organized as follow. Section 2 presents a brief summary of the RL. Then, Sect. 3 details the proposed method. Experimental results are provided in Sect. 4 and conclusions and future work are given in Sect. 5.

2 Reinforcement Learning

As mentioned above the current work proposes the use of a multi-table RL strategy for finding the best descriptor that characterizes a given image. In this section a brief description of RL is presented just to introduce the notations and the definition of the used elements (see [13] for more details).

The RL is a learning method used in those cases where the agent does not have a prior knowledge about which is the correct action to take. The RL is a Markov decision process intended to learn how an *agent* ought to take an *action* in a given *environment* so that a *reward* is maximized. These concepts are defined with the following tuple $\langle S, A, \delta, \tau \rangle$, where: S is a set of environment states; A is a set of actions; δ is a transition function $\delta{:}S \times A \rightarrow S$ and τ is a reward/punishment function, $\tau{:}S \times A \rightarrow \Re$.

By using the definitions presented above the RL methods works as follow: for a given state s_z, the agent selects the action a_h that maximize the expected reward r based on the τ function. In other words, by applying the action a_h in the state s_z a new state s_{z+1} and a reward r are obtained. Figure 1 illustrates the interaction between the agent and the environment.

The RL can be solved using dynamic programming, Monte Carlo method and temporal difference learning. The temporal difference learning is used in the current work because it does not require a model and it is fully incremental [13]. More concretely, the used framework is based on the *Q-learning* algorithm [14]. In our work, the current state s_z is only affected by the previous visits but not for the future since the Markov decision problem is of first order [8]. The δ and τ functions are nondeterministic; hence each element of the Q-table, for an iteration n, is computed as follow:

$$Q_n(s_z, a_h) \longleftarrow (1 - \alpha_n) Q_{n-1}(s_z, a_h) + \alpha_n [r + \gamma \max_{a'} Q_{n-1}(s_z, a')], \tag{1}$$

$$\alpha_n = \frac{1}{1 + \mathbf{visits}_n(s_z, a_h)}, \tag{2}$$

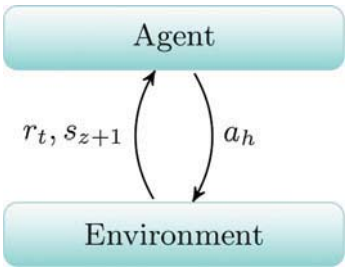

Fig. 1 Illustration of interaction between agent and the environment

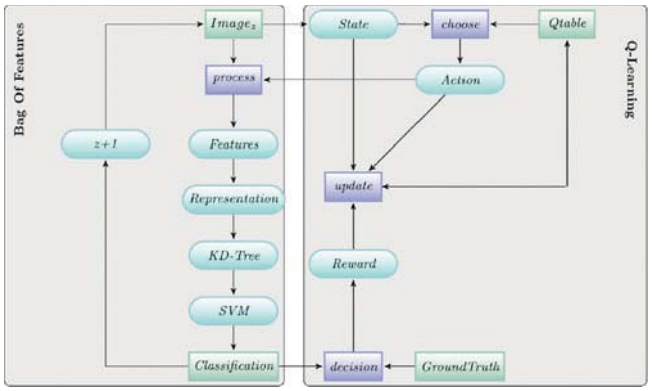

Fig. 2 Illustration of learning the best descriptor for each image using Q-learning

where γ is a discount factor for future reinforcements and is defined as $0 \leq \gamma < 1$. The Eq. 2 is the value of α_n for a nondeterministic world and **visits** is the number of iterations visiting the Q-table at the tuple (s_z, a_h) [15].

3 Proposed Method

The proposed approach, as mentioned in Sect. 1, uses the classical BoF ([1, 2]) for object recognition; our work is particularly focussed on the first step of BoF. In other words, we propose a multi-table reinforcement learning based strategy to select the best descriptor for each image from a set that contains the most widely used in the literature. This section presents the definition of the main elements of RL as well as the proposed strategy to combine the Q-table. Figure 2 illustrates the BoF (see *left-side*) with the proposed RL for the first step (see *right-side*). The three remainder steps of BoF are implemented following the state of the art, hence they are not detailed in this section (see [4] for more details).

3.1 Tuple Definition

This section aims at describing the different elements used in the RL formulation of the current work. The tuple $\langle S, A, \delta, \tau \rangle$ is defined as follows:

State definition: A state is defined as a set of characteristics from the given image. In order to tackle the challenge of defining a single representative state [5] that could be used with different databases, in the current work we propose the use four different state definitions. In all the cases the states are defined by a set of clusters obtained from the extracted vectors of characteristics—by using k-means. The vector for a given image is defined by extracting the information in a

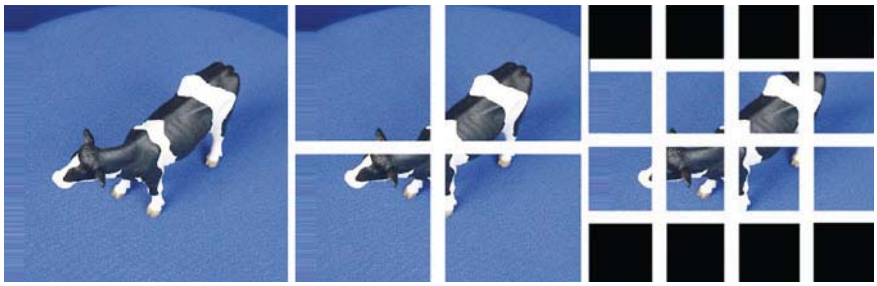

Fig. 3 (*left*) Image from ETH database [16]. (*middle*) Image splitted up into four squares. (*right*) Image splitted up into sixteen squares, but only eight of them are used

structured way. For instance, in the case of Fig. 3 the vector of characteristics is defined with 13 elements. The first element of that vector is obtained extracting information from Fig. 3 (*left*); then, Fig. 3 (*middle*) contributes with the next four elements of the vector; finally, the last eight elements are obtained from Fig. 3 (*right*). Once the vectors from all the given training images have been obtained a k-means clustering is used to compute the states that will be used for the Q-learning.

(1) $L^*a^*b^*$ ***state definition***: This state uses the $L^*a^*b^*$ color space. This color space is obtained by converting the RGB to XYZ and then, XYZ to $L^*a^*b^*$ (see [17, 18] for more details). The L^* represents the luminance of the image, a^* represents the difference between the red and green colors, and the b^* is the difference between the yellow and blue colors. As mentioned above the given image is split up into 13 squares (see Fig. 3, for each one of them the median value of $L^*a^*b^*$ is computed. Note that since $L^*a^*b^*$ has three components this state definition results in a vector of 39 elements.

(2) ***Gradient state definition***: This state definition uses the gradient in x and y directions. The gradient provides edges, but in this case, the state is defined by extracting the median of values in x and y. The partition shown in Fig. 3 is also used here. Hence, this state is defined with a vector of 26 elements.

(3) ***Entropy state definition***: The entropy measures the uncertainty of the information. In this case, the information is computed using the same partition shown in Fig. 3, but from the corresponding gray-scale image instead of the RGB color one. For each element of the partition the entropy is computed as follow:

$$E = -\sum_{i=1}^{N} p_i log_2(p_i), \tag{3}$$

where, $p(x)$ is the histogram of image data. In this case the state is defined with a vector of 13 elements.

(4) ***Histogram of interest point state definition***: This particular state is defined using all the descriptors of the set of actions. It works as follow; for each

image from the training set, it extracts all the interested points and describe them accordingly. After that, similarly to the process of BoF, it constructs a dictionary and find a representation of the image interested points. Finally, a vector with 50 elements (10 elements per descriptor) is extracted to represent the state.

Actions: In the current work, the actions are a set of descriptors. In this case, the RL learns the best descriptor for each image. Note that there is a large number of descriptors in the literature [19] the five most representative descriptors are selected for this work: SIFT [20], PHOW [4], C-SIFT [21], SURF [3] and Spin [22].

δ function: Usually, in the RL, the δ functions is defined as $\delta : s_z \times a_h \rightarrow s_{z+1}$. But, in this work, the δ function does not give a new state. In the current work, after applying an action a_h to the state s_z it generates a new representation of the image (features). The features obtained in this stage are used in the BoF for classifying the object. After that, the process continues through a new image. Summarizing, given a state s_z and applying an action a_h we obtain a new image from the training set and this new image does not have any similitude with the previous image.

τ function:

Theτ functions is defined by $\tau : s_z \times a_h \rightarrow \Re$, when the classification step gives the same label than the given object, the τ function gives a reward, and when the label does not match with the ground truth, the τ provides a punishment.

3.2 Combination of Q-tables

Like in [5], the joining process of BoF and RL is used to train the Q-table. Figure 2 shows this training process, which works as follow. For a given image, the agent extracts the state and applies a descriptor selected from the Q-table using the exploration/exploitation trade off. In the current work the ε-greedy algorithm is used as a strategy for the exploration/exploitation. After applying the descriptor, the agent follow the BoF scheme using the kdtree algorithm and the support vector machine. Once it finishes, the agent obtains a label of the classified image. The agent compares the obtained label with the ground-truth and obtains a reward/punishment. This information is used to update the Q-table according to Eq. 1, using the reward obtained before, the state and the applied action. Finally, after completing a whole iteration the agent extracts a new image from the training set and starts the training process again. In the current work this process is applied four times, one time per state definition. As a result we obtain four Q-tables. Now the question is to define which one should be used for a given image.

In the current work we propose a simple strategy for combining the four Q-tables computed as mentioned above. Actually, the information is not combined; the strategy consists in selecting the action from that Q-table where the

Fig. 4 Some of the objects contained in the nine classes of ETH database

reward is maximized. As will be presented in next section this simple strategy allows to improve results with respect to state-of-the-art, which only work with a single state definition.

4 Experimental Results

The proposed approach has been evaluated using two different databases. Additionally, it has been compared with a recent approach [5] as well as with others two BoF implementations where the first step consists of: (*i*) just a single descriptor; and (ii) a RL based approach with different states definition.

The first experiment is using the ETH database. Figure 4 shows the nine classes we have selected (i.e., apple, car, cow, cowcup, cup, dog, horse, pear and tomato) for testing and comparing the proposed approach. Each of these classes contain 45 images, which were randomly selected. These 45 images are split up into three sets: 15 images for training the BoF; 15 images for training the Q-table and 15 images for testing. The process of training starts creating for each definition of the states the corresponding Q-table. The process is repeated 60.000 times and the values of $\gamma = 0.9$ and $\varepsilon = 0.2$ are used.

In order to have a first comparison of the results obtained with the proposed approach, the performance of BoF for each descriptor is computed. Table 1 shows that the best performance is obtained when BoF uses the PHOW descriptor (74.81 % of recognition ratio). In Table 2 the performance of using the BoF with the RL method is presented. The first four rows depict the performance independently obtained for every state definition. In this case (BoF with RL) the best performance is reached with the $L^*a^*b^*$ state definition (82.4 % of recognition

Table 1 Performance of BoF using a single descriptor

Descriptor	Performance (%)
Spin	60.00
SIFT	61.48
SURF	62.96
C-SIFT	68.15
PHOW	74.81

Table 2 Performance of BoF with RL

State Definition	Performance (%)
$L^*a^*b^*$	82.2
Gradient	77.8
Entropy	78.5
Histogram of words	77.1
App. presented in [5]	81.4
Proposed approach (multi-table RL)	**83.4**

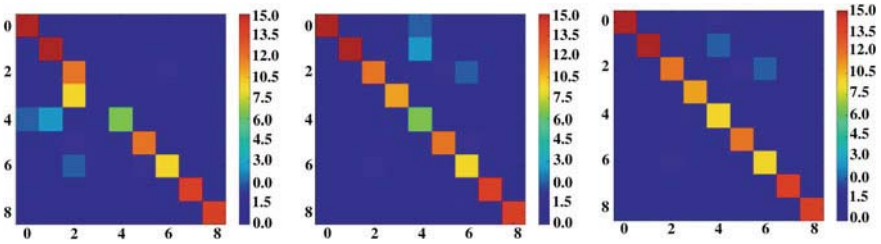

Fig. 5 Different confusion matrices for ETH database. (*left*) Using only the PHOW descriptor (74.81 %). (*middle*) Using $L^*a^*b^*$ state definition (82.2 %). (*right*) Using the proposed method (83.7 %)

ratio). Additionally, Table 2 shows the result obtained using the proposal presented in [5] (see fifth row). Finally, the performance obtained with the proposed approach (multi-table RL) is depicted in the last row. It can be appreciated that the best performance is obtained by using the strategy proposed in the current work (83.4 % of recognition ratio). Some of these results are presented in Fig. 5 by means of the corresponding confusions matrices. Figure 5 (*left*) shows the confusion matrix resulting when BoF with a single descriptor is used, in this case the PHOW descriptor. The confusion matrix presented in Fig. 5 (*middle*) corresponds to the BoF using RL and the $L^*a^*b^*$ state definition. Finally, Fig. 5 (*right*) depicts the confusion matrix resulting from the proposed approach.

A similar comparison to the one presented above has been performed with another database to validate the proposed approach. In this case, the COIL database [23], which contains 100 classes has been selected (Fig. 6 shows some of the objects contained in the COIL database). Each of these classes contain 45 images, which are split up into three sets: 15 images for training the BoF, 15 images for training RL and 15 for testing. In this case the process is repeated 600.000 times and the values of γ and ε are the same as those used in the first experiment ($\gamma = 0.9$ and $\varepsilon = 0.2$).

Fig. 6 Some of the objects contained in the COIL database (the whole database contains 100 classes, each class contains 45 images)

Table 3 Performance of BoF using a single descriptor

Descriptor	Performance (%)
Spin	83
SIFT	92.2
SURF	82.27
C-SIFT	94.47
PHOW	98.3

Table 4 Performance of BoF with RL

State Definition	Performance (%)
$L^*a^*b^*$	98.53
Gradient	98.8
Entropy	98.6
App. presented in[5]	98.3
Proposed approach (multi-table RL)	**99.0**

The performance of BoF for each descriptor is computed and presented Table 3; it can be appreciated that the best performance is again achieved using the PHOW descriptor (98.3 % of recognition ratio). In Table 4 the performance obtained when the BoF is used with the RL method is presented. The first four rows present the performance obtained for each of the state definitions introduced in Sect. 3.1. In this case the best performance corresponds to the BoF with RL and with the gradient state definition (98.8 % of recognition ratio). In order to compare the results with [5], the Table 4 presents in the fifth row the recognition ratio. Finally, in the last row of the Table 4, the performance of the proposed strategy is shown; note that in this case it reaches 99.0 % of recognition ratio. The confusion matrices corresponding to three of the examples presented in Tables 3 and 4 are presented in Fig. 7. In the left side the confusion matrix using BoF with a single descriptor (PHOW descriptor) is presented; the middle illustration corresponds to BoF with RL when the best state definition is used (gradient state definition). Finally, the right side illustration depicts the results obtained with the proposed approach. Note that in this case since the is a larger number of objects in the database and the recognition ratios are about 99 %, the confusion matrices are almost a diagonal line.

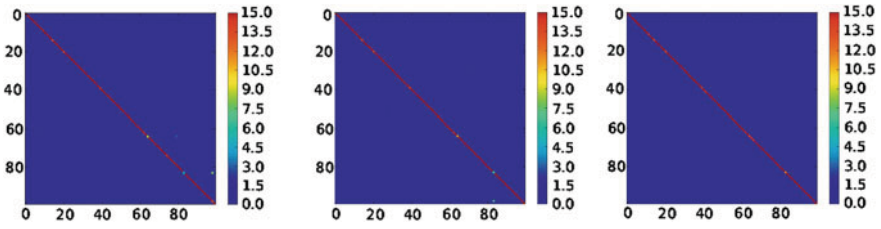

Fig. 7 Different confusion matrices for COIL database. (*left*) Using only the PHOW descriptor (98.31 %). (*middle*) Using gradient state definition (98.8 %). (*right*) Using the proposed method (99.0 %)

5 Conclusions and Future Work

This paper presents a BoF based approach for visual object recognition. We propose to improve classical BoF by means of the use of a RL strategy for selecting the best descriptor for each image. Our contribution lies on a novel method that allows the use of a multi-table strategy in the RL. This multi-table strategy allows to pick up the best state definition for each image. Experimental results are obtained using the BoF with the RL with two databases: ETH and COIL. In the first database, PHOW is the best descriptor and results in a 74.81 % of recognition ratio, the recognition ratios reaches 83.4 % using the proposed method. In the second database, the best single descriptor is also PHOW and in this case a recognition ratio of 98.3 % is reached, however, with the proposed method it could be also improved up to the 99.0 % of recognition ratio. Future work will be focused on the combination of descriptors and new state definitions in order to further improve the performance.

Acknowledgments This work was partially supported by the Spanish Government under Research Program Consolider Ingenio 2010: MIPRCV (CSD2007-00018) and Project TIN2011-25606. Monica Piñol was supported by Universitat Autònoma de Barcelona grant PIF 471-01-8/09.

References

1. Csurka G, Dance CR, Fan L, Willamowski J, Bray C (2004) Visual categorization with bags of keypoints. In: Workshop on statistical learning in computer vision, Proceedings of the European conference on computer vision (2004), pp 1–22
2. Fei-Fei L, Perona P (2005) A bayesian hierarchical model for learning natural scene categories. In: Proceedings of IEEE conference on computer vision and pattern recognition, pp 524–531
3. Bay H, Tuytelaars T, Gool LV (2006) Surf: speeded up robust features. In: Proceedings of the European conference on computer vision, pp 404–417
4. Bosch A, Zisserman A, Muñoz X (2007) Image classification using random forests and ferns. In: Proceedings of international conference on computer vision

5. Piñol M, Sappa AD, López A, Toledo R (2012) Feature selection based on reinforcement learning for object recognition. In: adaptive learning agent workshop, pp 4–8
6. Shokri M, Tizhoosh HR (2008) A reinforcement agent for threshold fusion. Appl Soft Comput 8:174–181
7. Sahba F, Tizhoosh HR, Salama M (2007) Application of opposition-based reinforcement learning in image segmentation. In: IEEE Symposium on Computational Intelligence in Image and Signal Processing, Honolulu, HI, pp 246–251
8. Harandi MT, Ahmadabadi, MN, Araabi, BN (2004) Face recognition using reinforcement learning. Proc IEEE Int conf image process 4:2709–2712
9. Häming K, Peters G (2010) Learning scan paths for object recognition with relational reinforcement learning. In: Proceedings of the 7th IASTED international conference on signal processing, pattern recognition and applications, vol 678. Innsbruck, Austria, p 253
10. Jodogne S (2005) Reinforcement learning of perceptual classes using q learning updates. In: Proceedings of the 23rd IASTED international multi-conference on artificial intelligence and applications, pp 445–450
11. Jodogne S, Piater JH (2004) Interactive selection of visual features through reinforcement learning. In: Proceedings of 24th SGAI international conference on innovative techniques and applications of artificial intelligence, pp 285–298
12. Bianchi R, Ramisa A, de Mántaras R (2010) Automatic selection of object recognition methods using reinforcement learning. Adv Mach Learn 1:421–439
13. Sutton R, Barto A (1998) Reinforcement learning: an introduction Cambridge Univ Press, MA (1998)
14. Watkins CJCH (1989) Learning from delayed rewards. Ph.D. thesis, King's College, Cambridge
15. Mitchell TM (1997) Machine learning. McGraw-Hill Science/Engineering/Math, New York
16. Leibe B, Schiele B (2003) Analyzing appearance and contour based methods for object categorization. Proc IEEE Conf Comput Vis Pattern Recogn 2:409
17. Ruzon MA, Tomasi C (2001) Edge, junction, and corner detection using color distributions. IEEE Trans Pattern Anal Mach Intell 23:1281–1295
18. Martin DR, Fowlkes CC, Malik J (2004) Learning to detect natural image boundaries using local brightness, color, and texture cues. IEEE Trans Pattern Anal Mach Intell 26:530–549
19. Mikolajczyk K, Schmid C (2005) A performance evaluation of local descriptors. IEEE Trans on Pattern Anal Mach Intell 27(10):1615–1630
20. Lowe D (2004) Distinctive image features from scale invariant keypoints. Int J Comput Vision 2:91–110
21. van de Sande KEA, Gevers T, Snoek CGM (2010) Evaluating color descriptors for object and scene recognition. IEEE Trans Pattern Anal Mach Intell 32:1582–1596
22. Lazebnik S, Schmid C, Ponce J (2005) A sparse texture representation using local affine regions. IEEE Trans Pattern Anal Mach Intell 27(8):1265–1278
23. Nene SA, Nayar SK, Murase H (1996) Columbia object image library (COIL-100). Technical report (Feb 1996)

DFT-Based Feature Extraction and Intensity Mapped Contrast Enhancement for Enhanced Iris Recognition

S. M. Rakesh, G. S. P. Sandeep, K. Manikantan and S. Ramachandran

Abstract Iris Recognition (IR) under varying contrast conditions with low gradience is challenging, and exacting contrast invariant features is an effective approach to solve this problem. In this paper, we propose two novel techniques viz., *Intensity Mapped Contrast Enhancement (IMCE)* and *Double symmetric rectangular hyperbolic based DFT (DsrhDFT) extraction*. IMCE is a preprocessing technique used to increase the gradience between brighter and darker pixels of an image, thereby obtaining the salient iris features. DsrhDFT is used to extract prominent shift-invariant features, and a Binary Particle Swarm Optimization (BPSO) based feature selection algorithm is used to search the feature space for optimal feature subset. Individual stages of the IR system are examined and an attempt is made to improve each stage. Experimental results obtained by applying the proposed algorithm on Phoenix, MMU and IITD iris databases, show the promising performance of the IMCE+DsrhDFT for iris recognition. A significant increase in the recognition rate and a substantial reduction in the number of features is observed.

Keywords Iris recognition · Feature extraction · Feature selection · Discrete fourier transform · Binary particle swarm optimization

S. M. Rakesh · G. S. P. Sandeep · K. Manikantan (✉)
M S Ramaiah Institute of Technology, Bangalore, Karnataka, India
e-mail: kmanikantan2005@yahoo.co.in

S. M. Rakesh
e-mail: rakeshsm91@gmail.com

G. S. P. Sandeep
e-mail: sandeepgsp17@gmail.com

S. Ramachandran
S J B Institute of Technology, Bangalore, Karnataka, India
e-mail: ramachandr@gmail.com

Mohan S. and S. S. Kumar (eds.), *Proceedings of the Fourth International Conference on Signal and Image Processing 2012 (ICSIP 2012)*, Lecture Notes in Electrical Engineering 221, DOI: 10.1007/978-81-322-0997-3_43,

1 Introduction

Biometrics [1] is giving a significant contribution in automated person identification as it aims in identifying each individual using various physiological characteristics such as fingerprints, face, iris, retina etc. Human recognition using Iris has been the most significant discovery as each individual has different iris patterns compared to others and the main advantage is that these patterns are invariant to ageing factors. In pattern recognition the key issue is the relation between interclass and intraclass variability. Identification can be reliably classified only if the variability among the different instances of a given class is less than the variability between different classes which is also satisfied by the properties of the Human Iris. Reference [2] explains Iris Recognition based on 2D-Gabor filters including many other normalizing, enhancement and invariant properties which are covered in Ref. [3]. Iris Recognition using Circular Symmetric Filters to enhance the capture of texture information of iris and a Nearest Feature Line (NFL) based iris matching approach was proposed in Ref. [4]. Extracting the tissue edge features from the iris and their matching using Identity verification, Texture analysis and Image segmentation are introduced in Ref. [5]. Reference [6] proposes Iris Recognition based on boundary localization using DWT as the feature extractor with SVM classifier. A novel iris coding method based on differences of discrete cosine transform coefficients of overlapped angular patches from normalized iris images is introduced in Ref. [7]. A comparative study of Iris recognition using Log-Gabor, Haar wavelet, DCT and DFT based features and related research works are discussed in Ref. [8].

Figure 1 represents the general flow of an Iris recognition system. The test iris image is acquired and preprocessed. The key features are obtained using an extractor. The matching is done comparing these features with the stored features. In this paper, *Log transformation* is used as an image preprocessing step to enhance the intensity level of the image. Two dimensional DFT is used as a feature extractor. *Binary Particle Swarm Optimization (BPSO)* [9] is adopted to select the optimized features from the extracted feature set, thereby further reducing the

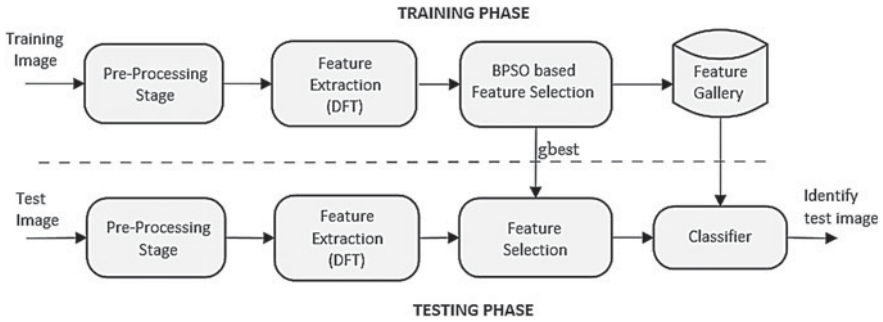

Fig. 1 General Iris recognition system

computational time. The obtained features are pixels which represent the properties of the whole image such as illumination, edges etc. proportional to their respective amplitudes. *Euclidean Classifier* performs Iris matching.

In this paper, we propose two novel approaches to enhance the Iris Recognition based on Two-Dimensional DFT as a feature extractor, which are shown below:

i. *Intensity Mapped Contrast Enhancement (IMCE)*:
Iris images in the databases used suffer from light insufficiency, lack of focus etc. during their capture. The proposed IMCE technique brings out a good contrast by increasing the gradience between brighter and darker pixels, resulting in enhanced iris features.
ii. *Double symmetric rectangular hyperbolic based DFT (DsrhDFT) extraction*:
The efficiency of an extractor depends on the extraction of required key features through precise geometrical shapes based on the spectrum obtained. The resultant shifted distribution of features after DFT is not only concentrated at the center, but also extended along perpendicular bisectors to edges of the spectrum, which is efficiently extracted by the proposed Double symmetric rectangular hyperbola.

The rest of the paper is organized as follows: Sect. 2 deals with image preprocessing based on the illumination of the iris image. Section 3 introduces the proposed contrast enhancement technique to obtain key features of the iris. The application of 2D-DFT as a feature extractor and the proposed *Dsrh* shape are discussed in Sect. 4. Section 5 introduces BPSO which selects optimized features saving the computational time along with Euclidean classifier for eye matching. Section 6 illustrates the proposed IR system and the experimental results obtained with three databases namely Phoenix, IITD, MMU. Section 7 summarizes the results obtained comparing with the proposed techniques.

2 Fundamental Image Preprocessing

Image preprocessing is a basic step done to improve the performance of subsequent procedures so that relevant information loss is reduced [10]. The Iris image obtained after converting to grayscale is not suitable for explicit feature extraction since the tissue edges are not distinct which may lead to erroneous recognition.

2.1 Logarithmic Transform

The original Iris image is converted to logarithmic domain thereby enhancing the image with respect to illumination. The above technique is performed by processing the image according to Eq. 1.

$$I(enhanced) = c \times \log(1 + I) \tag{1}$$

The resultant illumination varies proportionally with c. The value of the constant c is chosen based on intensity level of iris image in respective database.

3 Proposed Intensity Mapped Contrast Enhancement Technique (IMCE)

Contrast enhancement is based on the emphasis of the difference between the brightness levels in an image to improve its perceptual quality [11]. Various contrast enhancement techniques [12] in spatial domain have been proposed using histogram modifications [13] and several Unsharp Masking techniques [14].

The original iris image (Fig. 2a) and the image obtained after logarithmic transform have smooth variations thereby not reflecting the detailed pattern of the iris as shown in Fig. 2b. To overcome this problem, a high contrast iris image containing detailed tissue edge features is obtained by IMCE as shown in Fig. 3. The above process is implemented by our proposed technique indicated by Eqs. 2–4, where $f(x, y)$ is the Iris image to be preprocessed and $F(x, y)$ is the resultant intensity mapped contrast enhanced image and n depends on the bit map of the respective Iris image. The images used here are 8 bit mapped, hence the maximum intensity level that can be obtained is 255.

Block diagram in Fig. 3 shows the flow of the proposed IMCE technique.

i. The original iris image (Fig. 3i) of Phoenix database is complemented through Eq. 4 to obtain Fig. 3ii which results in brighter pixels.
ii. The original image is now subtracted from the complemented image to obtain an intermediate image shown in Fig. 3iii. During this process, brighter pixels (higher value) of the complemented image are subtracted from the darker pixels (lower value) of the original image preserving the brightness and the darker pixels are subtracted from brighter pixels that results in negative amplitudes, which are considered as 0 value or absolute black. This brings out appreciable difference between the darker and brighter parts of the image.

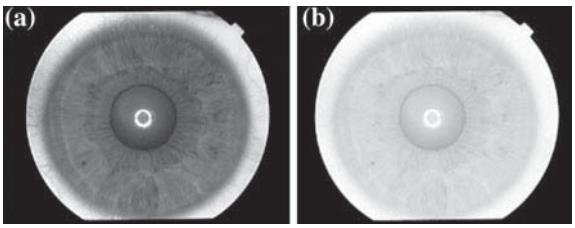

Fig. 2 Sample phoenix image **a** RGB to gray **b** applying log transform

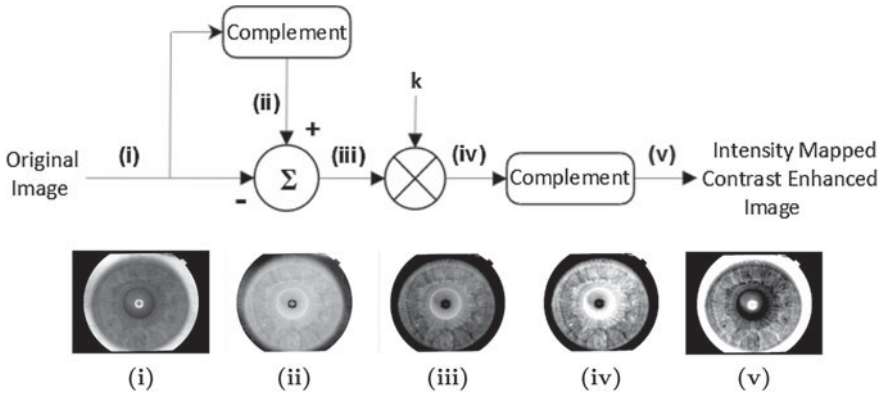

Fig. 3 Various blocks in IMCE (*upper*), and the images obtained at the output of each block (*lower*)

iii. The obtained image is now scaled by a factor k (greater than 1). The scaling factor further increases the mid-range intensity levels whereas it has no effect on the negative amplitudes as it is still treated as 0. This process further increases the gradient resulting in a high contrast image (Fig. 3iv).
iv. The resultant image is still in the complemented form as the original image is subtracted from the complemented image. Hence the image is again complemented resulting in IMCE image as shown in Fig. 3v.

$$\overline{f(x,y)} = complement(f(x,y)) \tag{2}$$

$$F(x,y) = \overline{(\overline{f(x,y)} - f(x,y)) \times k} \tag{3}$$

$$complement = [2^n - 1] - f(x,y) \tag{4}$$

The value of k is always chosen greater than 1 as a value below 1 will decrease the contrast of the iris image. An optimum k value is selected in the range of 1–1.5 for the databases used, to obtain the required contrast.

4 Proposed DsrhDFT Based Feature Extraction

DFT is a complex transform which converts the image from spatial domain to frequency domain whose equation is described in Eq. 5. The frequency components obtained are complex in nature having both real and imaginary parts. Since

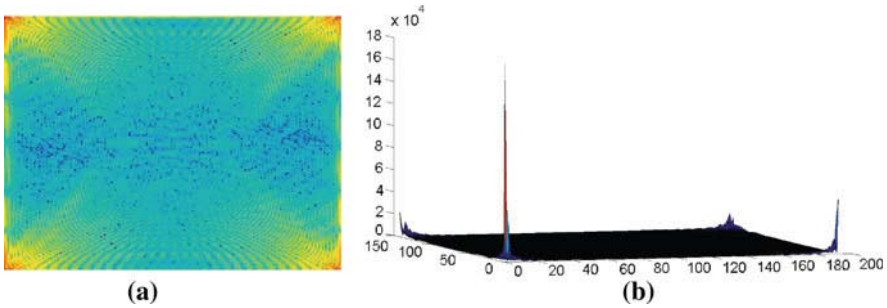

Fig. 4 DFT of Fig. 3v. **a** Frequency spectrum **b** surf diagram

the frequency components are in complex, for further processing only the magnitude of each component is calculated and processed as discussed in Ref. [15].

$$F(u,v) = \frac{1}{MN}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1} f(x,y)e^{-j2\pi\left(\frac{ux}{M}+\frac{vy}{N}\right)} \tag{5}$$

Two-Dimensional DFT is applied to the preprocessed image to obtain the frequency spectrum as shown in the Fig. 4a. The features of an image are the low frequency components having higher amplitudes which are adequate for representing the key patterns of the respective image. The features are distributed at the 4 corners of the obtained spectrum. The high amplitudes of the important features are clearly shown in Fig. 4b which have peaks at the corners. Since the extraction of these features is complicated, the spectrum is shifted to the center in a symmetric manner, exploiting the shift invariance property of DFT. The application of Logarithmic Transform as a preprocessing step results in a feature distribution which are not only concentrated at the center rather few extend along the perpendicular bisectors to the edges of the spectrum as shown in Fig. 5a and in the surf diagram illustrated in Fig. 5b. Considering the above distribution of the features a novel idea of *Double symmetric rectangular hyperbola* is used for extracting the features as shown in Fig. 6.

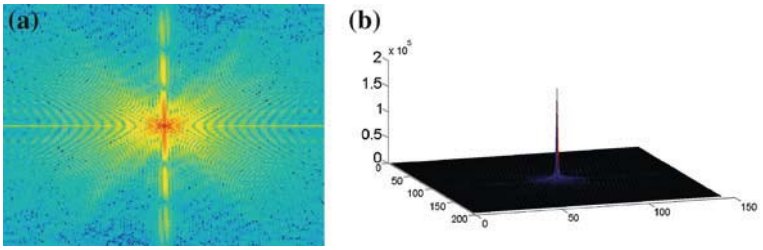

Fig. 5 Centered spectrum using fftshift [21] **a** frequency spectrum **b** surf diagram

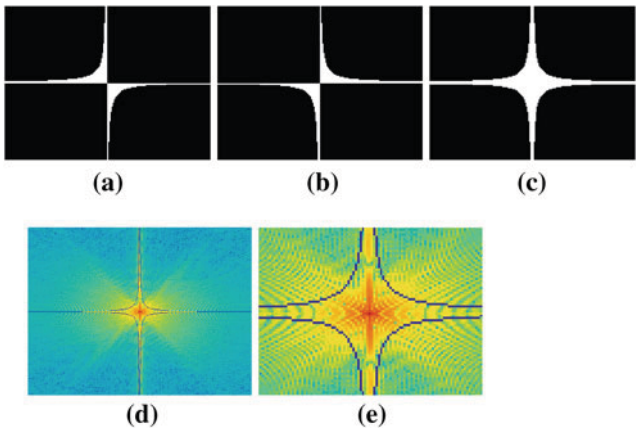

Fig. 6 Generation of Double symmetric rectangular hyperbola mask. **a** $xy = r^2$ **b** $xy = -r^2$ **c** $xy = \pm r^2$ **d** DSRH on centered DFT spectrum **e** magnified centered spectrum

This optimized Double symmetric rectangular hyperbolic based extractor mask is generated by the combination of two rectangular hyperbolas as shown in Fig. 6a, b whose equation is given by $xy = \pm r^2$ (Fig. 6c). From the property of Hyperbolas, a circle can be drawn with its center coinciding with the center of the spectrum touching the 4 parts of the Dsrh whose radius is $\sqrt{2}$ times the value of r. The value r is chosen such that the features at the center are completely enclosed in this circle which also ensures extraction of features along the asymptotic region, which is found to be in the range of 1–3. The size of the iris image also affects the value r, a higher value of r is chosen for larger iris image. The extraction of the features using the above Dsrh mask is illustrated in Fig. 6d, e shows the extraction of the center features ensuring enhanced extraction.

5 Feature Selection Using Binary Particle Swarm Optimization

Particle Swarm Optimization (PSO) was introduced by Eberhart and Kennedy in 1995 based on the idea of collaborative behavior and swarming in biological populations inspired by the social behavior of bird flocking or fish schooling [16]. PSO uses a set of swarms associated with two variables namely the position and the velocity. PSO keeps updating the velocity of the particle relative to its previous velocity until an optimized convergence is achieved. The above methodology is obtained using Eq. 10 in which v_i^t and x_i^t represent the velocity and position of the particle respectively with c_1 (*cognitive factor*) and c_2 (*social factor*) decide the rate

of convergence and ω is the inertial weight. The values of c_1, c_2 and ω are 2, 2, and 0.6 respectively with a swarm size of 30. The velocity of particle depends on $pbest_i$ which is the best position previously visited by the particle and *gbest* which is the overall best position visited by any particle. The fitness function evaluates the quality of evolved particles in terms of their ability to maximize the class. The pbest and gbest are assigned based on the highest Fitness Function value which is described in Eq. 11. Let $N_1, N_2, N_3 \ldots N_L$ denote the images within the classes, Let $M_1, M_2, M_3 \ldots M_L$ denote the class means of the respective classes which are calculated by Eq. 6 and M_0 is the grand mean in the feature space calculated by Eq. 7. Accordingly the fitness function is described in Eq. 11.

$$M_i = \frac{1}{N_i} \sum_{j=1}^{L} W_j^{(i)} \tag{6}$$

$$M_0 = \frac{1}{N} \sum_{i=1}^{L} N_i M_i \tag{7}$$

$$f(x) = \frac{1}{1 + e^{v_i{}^{t+1}}} \tag{8}$$

$$x_{id} = \begin{cases} 1 & \text{if } rand_3 < f(x), \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

where $W_j^{(i)}$, $j = 1, 2, 3 \ldots N_i$ represents the sample images of the corresponding class w_i.

$$v_i{}^{t+1} = \omega \times v_i{}^t + c_1 \times rand_1 \times (pbest_i - x_i{}^t) + c_2 \times rand_2 \times (gbest - x_i{}^t) \tag{10}$$

Binary Particle Swarm Optimization (BPSO) [9] is a modified binary version of the PSO in which the continuous velocities of the swarms are represented by binary 1's and 0's. There are various feature selective algorithms like BPSO, GA (Genetic algorithm) etc. In GA, chromosomes share information with each other continuously resulting in the movement of whole group together towards the optimum area. In BPSO the global best particle shares information among the remaining particles resulting in a one way information sharing. BPSO is thereby preferred in recognition [17] as it reduces the time to converge to the optimum area reducing the computational time. The particle velocity $v_i{}^{t+1}$ is restricted to (0,1) using the sigmoid transformation given by Eq. 8. The position of each particle is updated depending on the condition Eq. 9. Thus the extracted features are selectively mapped to 1's which are the essential features of the iris image discarding

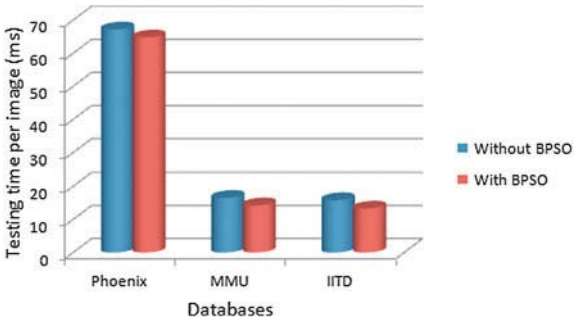

Fig. 7 Illustration of reduced testing time with BPSO

the 0's, resulting in a reduced feature set avoiding complexity and less computational time. Figure 7 shows that the addition of BPSO reduces the testing time of the recognition process in all the iris databases used.

$$F = \sqrt{\sum_{i=1}^{L} (M_i - M_0)^t (M_i - M_0)} \tag{11}$$

$$E = \sqrt{\sum_{i=1}^{N} (p_i - q_i)^2} \tag{12}$$

where $(rand_1, rand_2, rand_3) \in (0,1)$. The fitness function used for feature selection computes a feature vector which minimizes intraclass variance, while maximizing interclass variance ensuring reliable classification. Euclidean classifier calculates the distance between two corresponding points. This technique is used to measure the similarity between the features of the test image with the feature gallery obtained during the training process. The N-dimensional distance between them is calculated using Eq. 12, where p_i is one of the feature vector, q_i is the

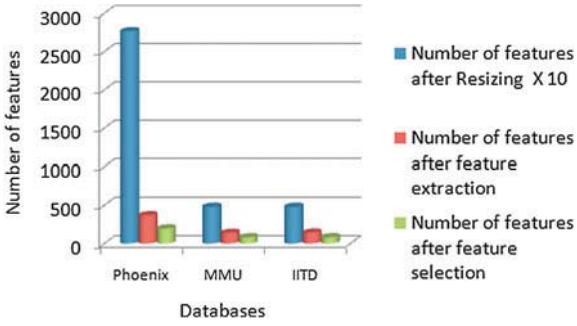

Fig. 8 Dimensionality reduction at various stages for three databases

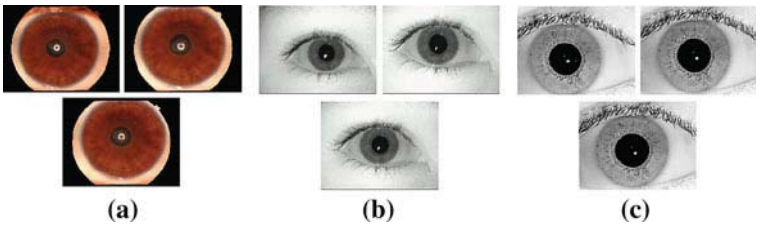

Fig. 9 Sample images of **a** Phoenix database **b** MMU database **c** IITD database

feature vector of the test image and N is the number of features extracted. The feature vector which gives the least distance is the matched image (Figs. 8, 9).

6 Discussion of Proposed IR System and Experimental Results

The block diagram of the proposed IR system is shown in Fig. 10. The experiments following the proposed method are carried out on three different databases namely Phoenix, Multi Media University (MMU) and Indian Institute of Technology Delhi (IITD). These databases have been selected to test the proposed IR system under various conditions, Phoenix database contains high resolution images of only the iris whereas the other two databases contain images of eye with noise from eye lashes. Figure 10 is applicable for Phoenix and MMU databases, Log transformation is discarded for IITD database as it is sufficiently illuminated, resulting in an other experiment.

The constant k of proposed Intensity Mapped Contrast Enhancement technique in Sect. 3 is database dependent. The variation of recognition rate for different values of constant k is shown in Fig. 11d. The optimum values of k are chosen to be 1.1, 1.2 and 1.3 for IITD, MMU and Phoenix databases respectively.

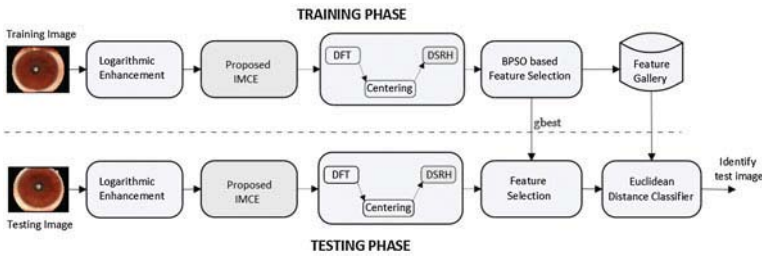

Fig. 10 Block diagram of the proposed Iris recognition system

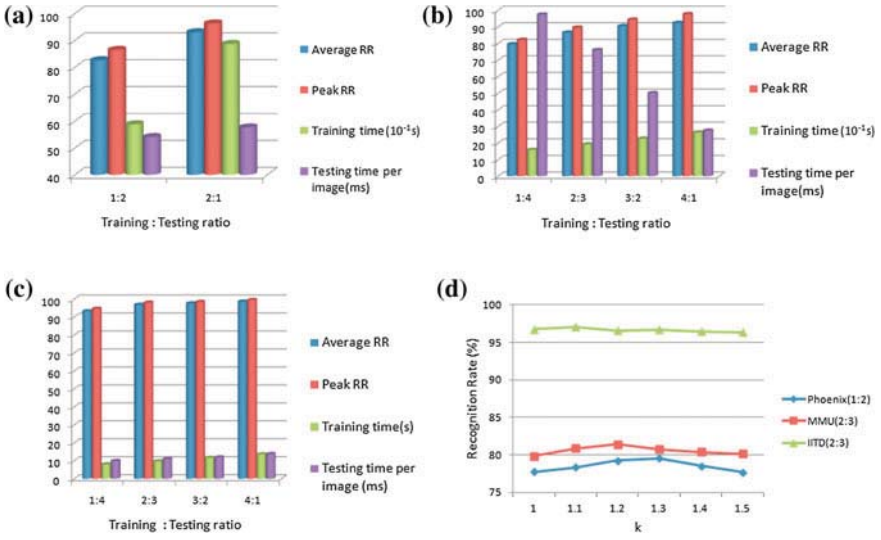

Fig. 11 Iris recognition results using the proposed IMCE+DsrhDFT method for **a** Phoenix database **b** MMU database **c** IITD database **d** Recognition rate versus k

6.1 Experiment 1 : Phoenix and MMU Database

The experiments were performed on Phoenix database [18] and MMU database [19] whose sample images are shown in Fig. 9a, b respectively. Phoenix database consists of 64 subjects with 3 images of each eye. The size of each image is 768×576 pixels which is scaled down to 192×144 pixels. The optimized features are obtained from these original pixels through various stages as illustrated in Fig. 8. The first 3 experiments tracks the importance of the image preprocessing steps and are conducted for all training to testing set ratios. The proposed IMCE technique with DFT has significantly improved the recognition rate by 13 % and further application of Log transformation lays foundation for better contrast enhancement resulting in an increase of 5 %, thereby enhancing Iris Recognition as shown in Table 1a. Further experiments covers the contribution of the geometry on Extractor, among which the proposed *Dsrh* provides better recognition rate using less number of features relative to circular and rectangular shapes as seen in Table 1d. The experimental results of proposed method for all Training:Testing ratios is shown in Fig. 11a.

MMU database contains 45 subjects with 5 images for each eye. The size of each image is 320×240 which is scaled down to 80×60 pixels. A 2 % increase is observed with application of IMCE technique and Dsrh proves to be a better extractor shape compared to the rectangular and circular based feature extractor which is indicated in Table 1b. The experimental results of proposed method for all Training to Testing set ratios is as shown in Fig. 11b.

Table 1 Recognition results highlighting the advantage of the proposed method

(**a**)

Method Adopted	Recognition rate (%) vs TTR[a] 1:2	2:1
Only DsrhDFT	64.95	76.41
IMCE + DsrhDFT	77.53	87.08
LOG + IMCE + RDFT	79.06	91.25
LOG + IMCE + CDFT	80.36	91.56
LOG + IMCE + DsrhDFT	**82.64**	**93.04**

[a] Training Testing ratios

(**b**)

Method Adopted	Recognition rate (%) vs TTR 1:4	2:3	3:2	4:1
Only DsrhDFT	67.48	78.56	83.55	85.48
IMCE + DsrhDFT	67.61	80.34	85	87.4
LOG + IMCE + RDFT	72.25	82.81	86.9	89.18
LOG + IMCE + CDFT	73.77	84.04	87.48	90.48
LOG + IMCE + DsrhDFT	**79.59**	**86.56**	**90.51**	**92.44**

(**c**)

Method Adopted	Recognition rate (%) vs TTR 1:4	2:3	3:2	4:1
Only DsrhDFT	92.06	96.00	97.28	98.31
IMCE + RDFT	92.56	96.29	97.36	98.21
IMCE + CDFT	93.10	96.42	97.67	98.66
LOG + IMCE + DsrhDFT	92.99	95.57	96.99	97.41
IMCE + DsrhDFT	**93.28**	**96.86**	**97.67**	**98.60**

(**d**)

Extraction Method	Recognition Rate (%)	Number of features Selected
Square	86.96	87
Circle	87.48	84
Dsrh	**90.51**	**82**

(a) Phoenix database (b) MMU database (c) IITD database (d) Optimized results using Dsrh compared to conventional square and circular DFT extractions

6.2 Experiment 2 : IITD Database

The experiments were performed on IIT Delhi database [20]. The sample images are shown in Fig. 9c. The database contains 224 subjects with 5 images for each

eye. The size of each image is 320×240 pixels which is scaled down to 80×60 pixels. The database is sufficiently illuminated for efficient working of IMCE technique discarding the need of Log transformation. The removal of Log transformation results in the accumulation of all the features at the center. The application of Dsrh results in the extraction of redundant features, as a result, the change in the Recognition Rate with respect to CDFT is negligible as shown in Table 1c. The experimental results of proposed method for all Training to Testing set ratios is shown in Fig. 11c.

7 Conclusions

A novel approach for a flexible Iris Recognition (IR) system is proposed which uses the combination of Intensity Mapped Contrast Enhancement (IMCE) for enhancing the finer details of the iris edges, Double symmetric rectangular hyperbolic DFT (DsrhDFT) for feature extraction, and a BPSO-based feature selection. $IMCE + DsrhDFT$ have played a key role and have been the main contributors for the high recognition rates (RR) being obtained. Due to the application of DFT and BPSO, a substantial reduction in the number of features has been observed. A successful attempt has been made to equally handle all image variations (low contrast, illumination, shift-variance). The proposed method exhibits extremely good performance under low contrast (Phoenix database). The experimental results indicate that the proposed method has performed well under severe illumination conditions with top RR having reached 97.7 % for MMU database. It is also successful in tackling the most challenging task of shift variance in IR with average RR of 96.8 % for IITD database with a Training to Testing ratio of 2:3. Using the technique of IMCE, top RR of 96.31 % was obtained for Phoenix database with a Training to Testing ratio of 2:1. On a PC with Intel(R) Core 2 Duo, 2.4 GHz CPU and 3 GB RAM, $IMCE + DsrhDFT$ costs an average testing time of 57.63 ms per image (Phoenix database) using MATLAB(R) [21]. This may still be a limitation of IMCE for real time applications. Hence, a future research issue could be to develop fast computation methods for IMCE.

This paper uses a simple Euclidean classifier. By using other classifiers such as SVM, Random Forest etc, and using suitable noise removal techniques, the performance of the IR system is expected to improve substantially.

References

1. Jain A, Bolle R, Pankanti S (2000) Biometrics: the future of identification. IEEE comput soc 33(2):46–49
2. Daugman J (2004) How Iris recognition works. IEEE Trans Circuits Syst Video Technol 14(1):21–30

3. Daugman J (2007) New methods in Iris recognition. IEEE Trans Syst Man Cybern 37(5):1167–1175
4. Li M, Yunhong W, Tan T (2002) Iris recognition using circular symmetric filters. In: Proceedings of 16th international conference on pattern recognition, vol 2, pp 414–417
5. Kevin B, Karen H, Flynn, P (2008) Image understanding for Iris biometrics: A survey. Comput Vis Image Underst 110:281–307
6. Sung H, Lim J, Park J-H, Lee Y (2004) Iris recognition using collarette boundary localization. In: 17th international conference on pattern recognition, vol 4, pp 857–860
7. Monro D, Rakshit S, Zhang D (2007) DCT-based Iris recognition. IEEE Trans Pattern Anal Mach Intell 29(4):586–595
8. Kumar A, Passi A (2010) Comparison and combination of Iris matchers for reliable personal authentication. Pattern Recogn 43(3):1016–1026
9. Kennedy J, Eberhart R (1997) A discrete binary version of the particle swarm algorithm. IEEE Int Conf Syst Man Cybern 5:4104–4108
10. Duda R, Hart P, Stork D (2000) Pattern classification, 2nd edn. Wiley-Interscience, New York
11. Gonzalez R, Woods R (2008) Digital image processing, 3rd edn. Prentice Hall, Upper Saddle River
12. Wang D, Vagnucci AH (1983) Digital image enhancement. Comput Vis Graph Image Process 24(3):363–381
13. Sun C, Ruan SJ, Shie MC, Pai TW (2005) Dynamic contrast enhancement based on histogram specification. IEEE Trans Consum Electron 51(4):1300–1305
14. Badamchizadeh MA, Aghagolzadeh A (2004) Comparative study of unsharp masking methods for image enhancement. In: 3rd international conference on image and graphics, pp 27–30
15. Imtiaz H (2011) A spectral domain local feature extraction algorithm for face recognition. Int J Secur 5(2):62–73
16. Kennedy J, Eberhart RC (1995) Particle swarm optimization. In: IEEE international conference on neural network, vol 4, issue 4, pp 1942–1948
17. Ramadan RM, Abdel-Kader RF (2009) Face recognition using particle swarm optimization-based selected features. Int J Signal Process Image Process Pattern Recognit 2(2):57–59
18. Phoenix Database, http://www.inf.upol.cz/iris/
19. MMU Database, http://pesona.mmu.edu.my/~ccteo/
20. IIT Delhi database, http://web.iitd.ac.in/biometrics/DatabaseIris.htm
21. Matlab, http://www.mathworks.in/

Programmable Gaussian Noise Generator to Test and Characterize Onboard Subcarrier Systems in Satellite Communication

K. K. Raghunandana, P. N. Ravichandran, Sunil Kulkarni, H. S. Vasudeva Murthy and M. Vanitha

Abstract The paper presents design, simulation and implementation of programmable Gaussian noise generator in hardware. The communication system performance is evaluated for additive Gaussian noise, this necessitates the design of highly accurate, programmable noise generator which is dealt in this paper. The Box-Muller method is used for Gaussian noise generator and the programmable variance is implemented in hardware through multiplying DAC. The Box-Muller method requires two uniform random generators inputs and mathematical operations for these inputs which are implemented in hardware using CORDIC algorithms. The Uniform noise generators and Gaussian noise generator are simulated in MATLAB Simulink®, finite word length effect is analyzed with bit true simulation and the same were implemented in hardware. The whole design is accommodated in a Xilinx xc4vsx35-10ff668 FPGA, the multiplying DAC circuit is realized as a separate circuit, with control inputs for noise variance control. The design is tested in real time for different subcarrier frequency and data modulation and receiver performance is evaluated in terms of BER characterized by E_b/N_0.

K. K. Raghunandana (✉) · P. N. Ravichandran · S. Kulkarni · H. S. Vasudeva Murthy · M. Vanitha
Digital Systems Group, ISRO Satellite Centre, Bangalore 560017, India
e-mail: kkraghu@isac.gov.in

P. N. Ravichandran
e-mail: pnravi@isac.gov.in

S. Kulkarni
e-mail: kulkarni@isac.gov.in

H. S. Vasudeva Murthy
e-mail: hmurthy@isac.gov.in

M. Vanitha
e-mail: vani@isac.gov.in

Mohan S. and S. S. Kumar (eds.), *Proceedings of the Fourth International Conference on Signal and Image Processing 2012 (ICSIP 2012)*, Lecture Notes in Electrical Engineering 221, DOI: 10.1007/978-81-322-0997-3_44,

Keywords Gaussian · Noise · CORDIC · Box-muller · DAC · Variance

1 Introduction

The quality of the Gaussian noise is very crucial in evaluating the performance of the communication systems. The non availability or expensive off the shelf Gaussian noise generators to test and characterize the on board receivers necessitated the development of simple, low cost, high performance and accurate noise generator.

The Gaussian noise can be generated in both analog and digital domain. The analog methods can produce truly random numbers but are sensitive to operational environment, hence digital methods for generating Gaussian random variables are preferred over analog methods. The random numbers generated by digital methods are pseudo random, but the period can be made sufficiently large such that the sequences never repeat themselves even in the largest practical situations.

The method chosen for the Gaussian noise generator is based on transformation of the uniform random variables, which can be implemented effectively in hardware with area, speed efficient and has programmable noise variance.

The programmable Gaussian noise generator will be used to test and characterize the performance of the on board subcarrier demodulator system. The subcarrier signal is BPSK/QPSK modulated and the demodulator performance is evaluated in terms of Bit Error Rate (BER) characterized by E_b/N_0. Onboard systems are qualified for BER of 10^{-6} using implemented noise generator.

The generation of Gaussian noise in digital domain and converting to analog signal with DAC gives limited range of programmability for noise variance. To overcome this limitation a new method is adopted which is dealt in detail.

The Gaussian noise generator is implemented using Box-Muller method. This method requires two independent uniform random generators, Mersenne twister and Tausworthe generators are used. The Box-Muller method requires trigonometric, logarithmic and arithmetic functions which are implemented using Coordinate Rotational Digital Computer (CORDIC) [1]. The mathematical operations are implemented in hardware in a fixed point format.

The Multiplying DAC provides the control signals to vary the amplitude of the reference signal through the control inputs these control inputs are used to provide programmable noise variance.

The paper is organized as follows: Sect. 2 deals with Gaussian noise generator Architecture, Sect. 3 deals with the selection of uniform random generators and hardware efficient algorithms Sect. 4 deals with the simulation and implementation. Section 5 deals with results and discussion. Section 6 deals conclusion at the end.