Iterative Identification Framework for Robust Hand-Written Digit Recognition under Extremely Noisy Conditions

Hosun Lee, Sungmoon Jeong, Tadashi Matsumoto, and Nak Young Chong

***Abstract*— A new classification framework for noise invariant hand-written digit recognition is proposed, which is based on Turbo decoding technique and Viterbi algorithm known from communication field. The digit image is modeled two-dimensionally correlated Markov Chain Model (MCM) and iteratively exploited on horizontal and vertical direction. In order to increase discriminant of Markov Chain models for each digit, a novel sequence learning algorithm is proposed to obtain sequence map and Markov chain models which represent the each class by minimizing an entropy information of MCMs within individual digit classes. The effectiveness of the proposed algorithm is verified through experiments that can significantly enhance the character classification accuracy under several noise conditions.**

I. Introduction

Recently, many classification methods for pattern and character recognition have been proposed and investigated in detail, including the advantages and remaining problems [1]. For character recognition, four types of classifiers are generally used: support vector machines, artificial neural networks, statistical methods, and ensemble methods. These methods are well-known with good recognition performance, however one remaining main topic is noise problem in sensor system. Fig.1 shows examples of noise situation when the character images are distorted by 50% of noise and hard to distinguish the digit class of each image. To overcome this problem, classifier has to be robust against to the presence of noise in sensed information.

Many papers and textbooks on information and communication systems are based on the reliable and efficient transmission of information over noisy channels. Because of this requirement, several powerful coding methodologies are proposed. These coding schemes have not only good coding performance but the capability of detecting or correcting errors for practical applications. The basic structure of a digital communication system is shown which represents the architecture of the communication systems [2]. The transmitter generates the signal that is used to transmit the original information across the channel. The transmitted signal is disturbed due to the noisy nature of the channel. The sensor-based systems also receive the information across the noisy environment. Therefore, we can consider digital communication techniques if the information belongs to a finite set of discrete signals. The nonlinear recursive filter based on a coding methodology, Viterbi Algorithm [3], shows a superior performance over the limitations of the linear filters. However, these communication techniques are developed for the coding processes. Therefore, it is difficult to be implemented directly to sensor-based system.

H. Lee, S. Jeong, T. Matsumoto and N. Y. Chong are with the School of Information Science, Japan Advanced Institute of Science and Technology, Ishikawa, Japan {Hosun_LEE, jeongsm, matumoto, nakyoung}@jaist.ac.jp

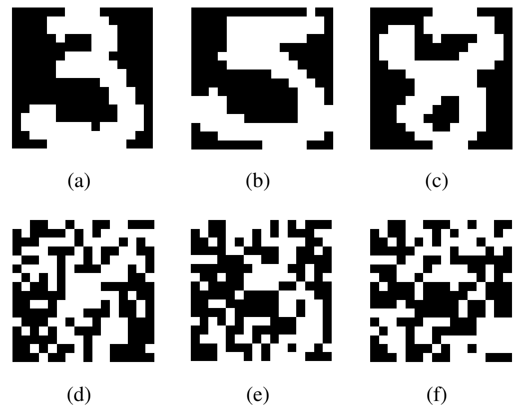

Fig. 1. Example cases of handwritten digit: original images, (a) 3, (b) 5, (c) 8, and original images are effected with 50% of noise, (d) 8, (e) 3, (f) 5.

It is the objective of this paper to extend those findings in communication theory to the noise problem of sensor-based system. Particularly, we focus on the character recognition which decides the class of the received handwritten digit image under the extreme noisy condition. Currently, the Viterbi algorithm with the hidden Markov model is generally used for character recognition [4] [5] [6]. In this paper, an iterative identification framework is proposed based on Turbo coding technique to obtain robust recognition performance. First, we model the handwritten image with the two-coupled Markov Chain Model (MCM) by considering the horizontal and vertical correlation property of 2D-image. We propose the sequence learning algorithm to obtain effective MCM for each digit class by transforming the sequence of the image pixels. The generated sequences and MCMs are used to exploiting the image data with two Viterbi decoder on horizontal direction and vertical direction. The effectiveness of the proposed algorithm is verified through experiments that can significantly enhance the character classification accuracy under several noise conditions.

In summary, the main contributions of this paper are as follows:

- We propose an iterative identification framework for hand-written digit recognition under noisy sensing condition.
- We propose the sequence learning algorithm to effec-

tively represent the digit class with the improved MCMs by generating the transform sequence.
- The proposed algorithm shows a novel approach to improve the system performance by learning the effective access sequence of the sensed data.

II. Markov Chain Model

A Markov chain is a probabilistic model that exhibits a special type of dependence: given the present, the future does not dependent on the past. In formulas, let $u_t, t \in \{0, 1, \cdots, T\}$ be a sequence of random variables taking values in the state space $s_l, l \in \{1, 2, \cdots, k, \cdots, L\}$. The sequence is a Markov chain or Markov process if

$$\begin{aligned} \Pr\{u_t = s_k | u_{t-1} = s_l, u_{t-2} = s_1, \cdots, u_0 = s_2\} \\ = \Pr\{u_t = s_k | u_{t-1} = s_l\} =: p_{lk} \end{aligned} \tag{1}$$

where the symbol "|" is the symbol for conditional probability.

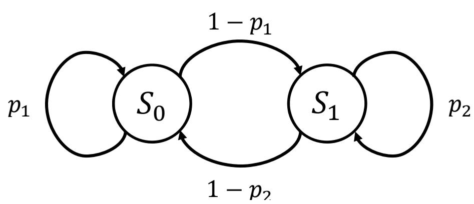

Fig. 2. First-order binary MCM

In this paper, we model a first-order binary MCM to deal with the binary source. The MCM is described by a single transition probability matrix. The transition probabilities, p_1 and p_2, correspond to relative frequencies of transitions from a state S_0 to another state S_1 (Fig. 2). These transition probabilities can be arranged in a square matrix from:

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = \begin{bmatrix} p_1 & 1-p_1 \\ 1-p_2 & p_2 \end{bmatrix}. \tag{2}$$

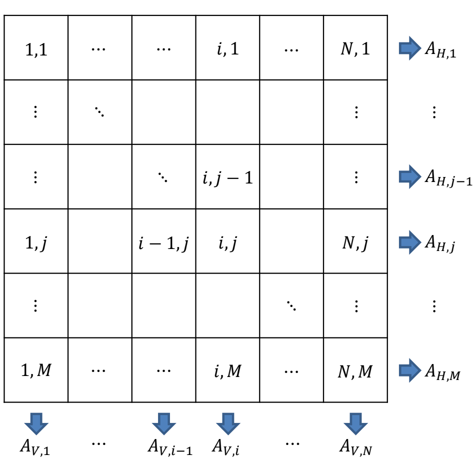

Fig. 3. 2D domain for the coupled MCM

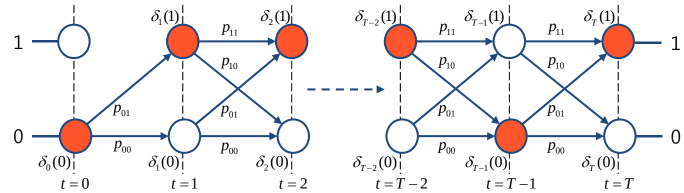

Fig. 4. The recursive computation of Viterbi algorithm works on the trellis diagram: Red nodes show the transitions between the states.

A. The coupled Markov Chain Model (MCM) for two dimensional images

For the two dimensional spatial source, the two-coupled MCM [7] can be used considering a two-dimensional domain of cells as shown in Fig. 3. The random variable $u_(i,j)$ of the current state is determined by two factors; the previous state in the horizontal direction, $u_(i-1,j)$, and that in the vertical direction, $u_(i,j-1)$. The current state also determines the next states in the horizontal and the vertical direction; $u(i+1,j)$ and $u(i,j+1)$. Therefore, in case of $M \times N$ two dimensional image (Fig. 3), $M+N$ transition probability matrices are respectively obtained from M horizontal and N vertical pixel sequences; $\mathbf{A}_{H,1}, \cdots, \mathbf{A}_{H,M}$ and $\mathbf{A}_{V,1}, \cdots, \mathbf{A}_{V,N}$.

B. Trellis diagram and Viterbi algorithm

An efficient algorithm to find the most likely path through a trellis which represents a graph of a finite set of states is introduced by Viterbi in [8]. This algorithm is currently the most popular decoding procedure for convolutional codes. The Viterbi algorithm is also applied to numerous other applications such as speech recognition and equalization for transmission over channels with memory.

The Viterbi algorithm is a dynamic programming algorithm. In computer science, dynamic programming is a method for reducing the runtime of algorithms exhibiting the properties of optimal substructures, which means that optimal solutions of sub-problems can be used to find the optimal solution of the overall problem. The sub-problems of the Viterbi algorithm are finding the most likely path from the starting node to each node in the trellis (Fig. 4). The optimal solution for the node of time step $i+1$ is however, simple if we know the solutions for all nodes of time step i. Therefore, the Viterbi algorithm traverses the trellis from left to right, finding the overall solution X, the maximum likelihood p_{max} estimate, when the terminating node is reached. The detail of the algorithm is described in Algorithm 1.

In the hand-written digit recognition, two Viterbi likelihood estimators are implemented to compute the maximum likelihood values of each digit class with respected to the received sequences of the image. By comparing these maximum likelihood values, the class of the target hand-written digit can be decided.

III. Sequence Learning Algorithm

MCM is the most important parameters in the classification process. Basically the parameters of MCM are automatically computed from training data by counting the

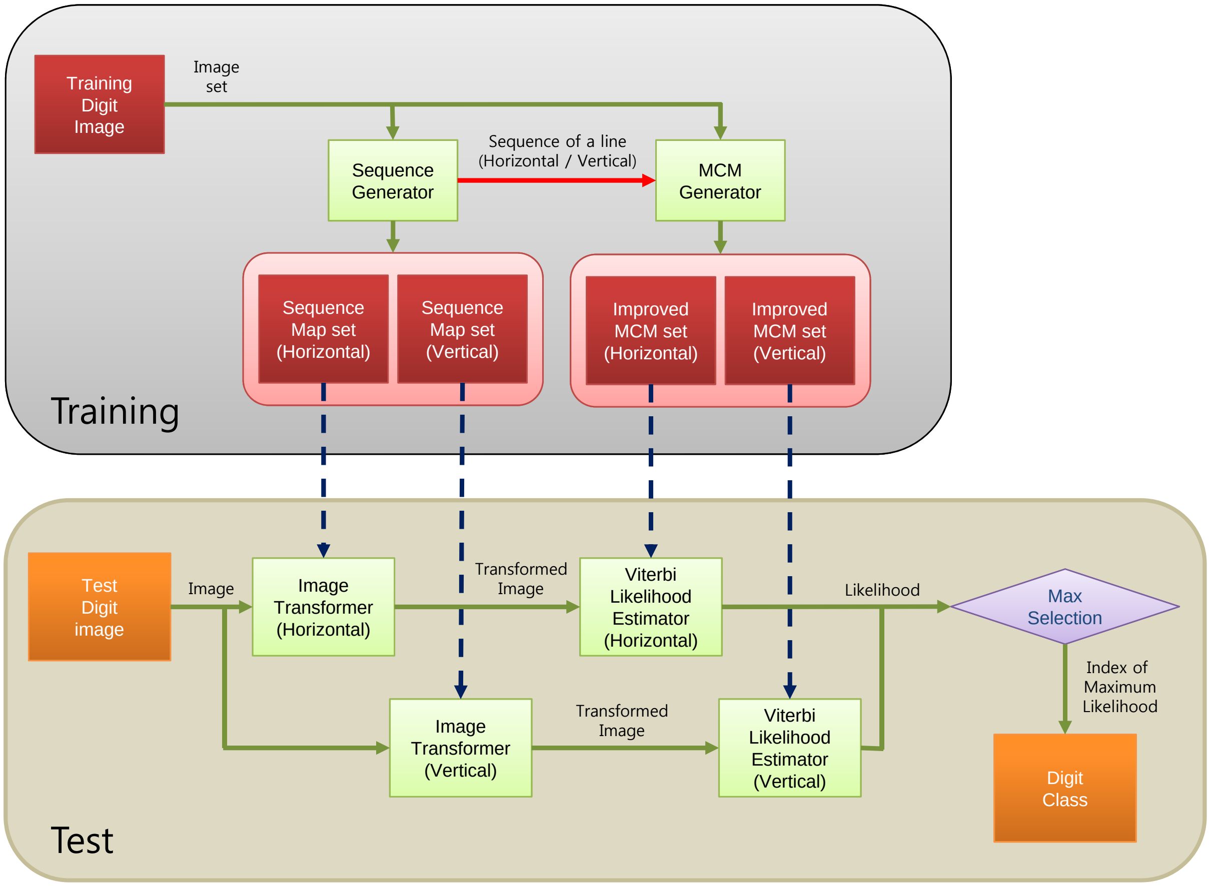

Fig. 5. The framework of handwritten digit recognition

Algorithm 1: Viterbi Algorithm

Step 1: **Initialization**

$$\begin{aligned} s, s' &\leftarrow \{0,1\} \\ \delta_0(s) &\leftarrow \Pr\{u_0 = s\} \\ \Psi_0(s) &\leftarrow 0 \\ X &\leftarrow \{x_1, \cdots, x_T\} \end{aligned}$$

Step 2: **Time Recursion**

for $t = 1$ **to** T **do**

$$\begin{aligned} \delta_t(s) &\leftarrow \max_{s'}[\delta_{t-1}(s') \cdot a_{s',s}] \\ \Psi_t(s) &\leftarrow \arg\max_{s'}[\delta_{t-1}(s') \cdot a_{s',s}] \end{aligned}$$

Step 3: **Termination**

$$\begin{aligned} p_{max} &\leftarrow \max_{s}[\delta_T(s)] \\ x_T &\leftarrow \arg\max_{s}(\delta_T(s)) \end{aligned}$$

Step 4: **Backtracking**

for $t = T-1$ **to** 1 **do**

$$x_t \leftarrow \Psi_{t+1}(x_{t+1})$$

transition from successive pixels. And then, mean transition probabilities within each class are defined as a model for that class. If different classes have similar parameters such as transition probabilities then it causes to decrease the classification accuracy. In order words, we are able to enhance the discriminant of each MCM of different classes by optimizing the sequences of pixels to calculate the transition probabilities.

In this paper, we propose a supervised learning approach to modify the parameters of MCM by changing the sequence of the pixels. That sequence can be selected by minimizing the entropy (H) of the transition probabilities in each class to increase the correlation coefficient between MCM parameters of each image in the same class. It can be rewritten as the following optimization problem formally:

$$\min H\left[A([seq,b]|D)\right], \tag{3}$$

where seq is the sequence of the previously selected pixels. $[seq,b]$ denotes the new sequence after the addition of a pixel, b. $A([seq]|D)$ denote the transition probability matrix with the dataset D within the new pixel sequence $[seq,b]$. The selection of the pixel is recursively carried by applying the greedy algorithm to find suitable pixel sequences for the training images. The detail computation is described in Algorithm 2. The sequence learning algorithm is placed in the training process of whole classification framework (Fig. 5) as Sequence Generator, which scans the training data horizontal and vertical direction to build Sequence Map sets for two directional sequences. Simultaneously, modified sequence for each training data is used to model the improved MCMs for each class. In the test session, Sequence Maps and improved MCMs calculated within training session are respectively used for the image transformers and Viterbi likelihood estimators to transform the test image and calculate the maximum likelihood.

Algorithm 2: Sequence generator

input : dataset D, data length M
output: sequence seq

begin
  Set $B \leftarrow \{1, 2, \cdots, M\}$;
  Select $seq \in \arg\min_{[b_1,b_2]} H\left[A([b_1, b_2]|D)\right]$;
  Set $B \leftarrow B \setminus \{b_1, b_2\}$;
  while $B \neq \emptyset$ **do**
    Select $b^* \in \arg\min_b H\left[A([seq, b]|D)\right]$;
    Set $seq \leftarrow [seq, b^*]$;
    Set $B \leftarrow B \setminus \{b^*\}$;
  end
end

IV. Iterative Identification Framework

Turbo codes get their name because the decoder uses feedback, like a turbo engine. The turbo code inventors [9] worked with two parallel concatenated and interleaved rate $1/2$ recursive systematic convolutional codes decoded iteratively with two MAP decoders. However, the source from the environment does not contain the product code. Therefore, the basic turbo decoding structure is simplified in the sensing problem.

In this paper, we propose an iterative identification framework based on the basic concept of Turbo decoding, and the sequence map and improved MCMs trained in off-line learning process are used to make the framework perform its function. The test process in Fig. 5 shows a structure of the proposed recognition framework that helps illuminate the iterative procedure. By the image transformers, the received test digit image is transformed in horizontal and vertical direction based on the sequence maps, and two transformed images are generated for each class. Then these two transformed images are exploited respectively by Viterbi likelihood estimators with improved MCMs. Finally, maximum likelihood values for each class are obtained, and the digit class of the target image is decided by comparing these maximum likelihood values.

V. Experimental Verification

A. Experimental setup

To evaluate the effectiveness of the proposed hand-written digit recognition algorithm,"USPS Handwritten digit dataset" is used, which contains selected 7000 handwritten digit images (size: 16×16 pixel) categorized into 10 class (10class $\times$ 700 images) [10]. The scale of the whole images, basically generated in gray level, is reduced to a binary scale by using Otus's method [11]. Fig. 6 shows the sample images in each class.

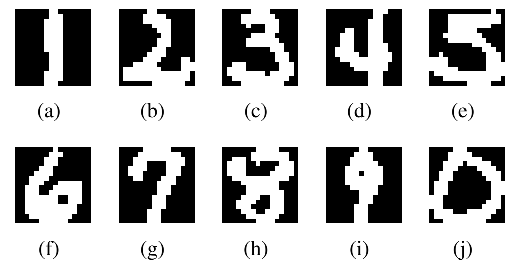

Fig. 6. An example of USPS handwritten digit dataset

B. Results and Discussion

To demonstrate the effectiveness of the sequence learning, the transition matrices of two images from the different classes are compared. Fig. 7 show the original images of digit 3 and 5 and horizontally and vertically transformed images.

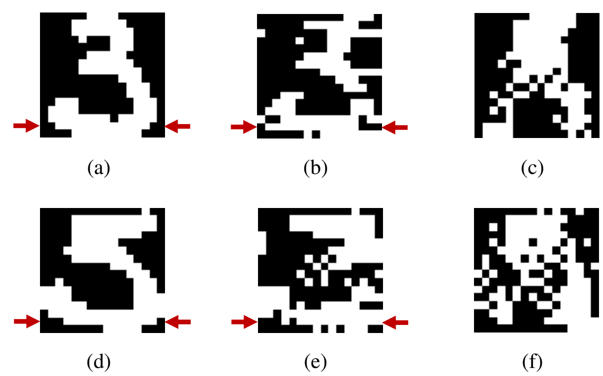

Fig. 7. Example of the transformed image: digit 3 - (a) original, (b) horizontal transform (c) vertical transform, digit 5 - (d) original, (e) horizontal transform (f) vertical transform

In figure, red arrow points the pixel line 15 which is compared. The parameters of the transition matrices computed from the original images (Fig. 7(a) and Fig. 7(d)) are same as

$$\mathbf{A}_{3H,15} = \mathbf{A}_{5H,15} = \begin{bmatrix} 0.6667 & 0.3333 \\ 0.0833 & 0.9107 \end{bmatrix} \tag{4}$$

After the transformation, the transition matrix of the class 5 is changed and the class 3 has same parameter as

$$\mathbf{A}_{3H,15} = \begin{bmatrix} 0.6667 & 0.3333 \\ 0.0833 & 0.9107 \end{bmatrix}, \tag{5}$$

$$\mathbf{A}_{5H,15} = \begin{bmatrix} 0.5000 & 0.5000 \\ 0.0909 & 0.9091 \end{bmatrix} \tag{6}$$

Therefore, the classifier can make distinguishable result in this pixel line. We also compare the correlations between all class models in Fig. 8. In each figure, the result value is 1 when the model is compared with itself. Figure shows that the sequence learning increases the discriminant of the MCM.

Next, the improvement of the recognition performance are verified by comparing two classifier. One is equipped with sequence map for image transformation and the improved MCM. The other uses general MCM. the recognition performances of two classifier are compared in Table. I. The recognition rate is dramatically improved with the sequence learning and the improved MCMs. And the proposed classified shows robust capability to noise. The recognition rate of the proposed classified at 30% noise is similar with the performance of the general MCM-based classifier at clear condition. And one detailed confusion matrix in the case with noise 50% is Table. II. We can also confirm that the proposed framework performs the recognition task in the extreme noisy condition mentioned in the beginning of this paper.

VI. Conclusion

In this paper, an iterative identification algorithm is proposed for noise invariant hand-written digit recognition. The goal of the proposed algorithm is to enable the system to decide the class of the received digit image which is disturbed by noise. The class models are modeled with two Markov Chain Model for vertical direction and horizontal direction. A new sequence learning algorithm is proposed to obtain sequence map for effective MCM generation and data acquisition. Turbo decoder equipped with two Viterbi decoders is designed to estimate the max a priori of each sequence. By comparing the likelihood values which are computed with given models, the class of the received data can be decided. The proposed character recognition framework shows robust and high identification result with the proposed sequence learning approach.

As future works, various experiments and analyses will be examined with various training data. In detail, we are considering improvements in the sequence learning to generate more effective data acquisition. This algorithm can be improved by considering various coding techniques in communication field. Then, it can be also extended to wide field of robotics.

References

[1] C.L. Liu and H. Fujisawa, "Classification and learning methods for character recognition: Advances and remaining problems", *Machine learning in document analysis and recognition*, ser. Studies in Computational Intelligence, vol. 90, Springer, pp. 131-161, 2008

[2] S.B.Z. Azami, P. Duhamel and O. Rioul, "Joint Source-Channel Coding: Panorama of Methods", *Proceedings of CNES workshop on Data Compression*, Toulouse, France, pp. 1232-1254, 1996

[3] G.D. Forney, "The Viterbi Algorithm", *Proceedings of the IEEE*, vol. 61 no. 3, pp. 268-278, 1973

[4] Y.H. Tseng, and H.J. Lee, "Recognition-based handwritten Chinese character segmentation using a probabilistic Viterbi algorithm", *Pattern Recognition Letters*, vol. 20 no. 8, pp. 791-806, 1999

[5] F. Einsele, R. Ingold, and J. Hennebert, "A language-independent, open-vocabulary system based on HMMs for recognition of ultra low resolution words", *Proceedings of the 2008 ACM symposium on Applied computing*, New York, NY, USA, pp. 429-433, 2008

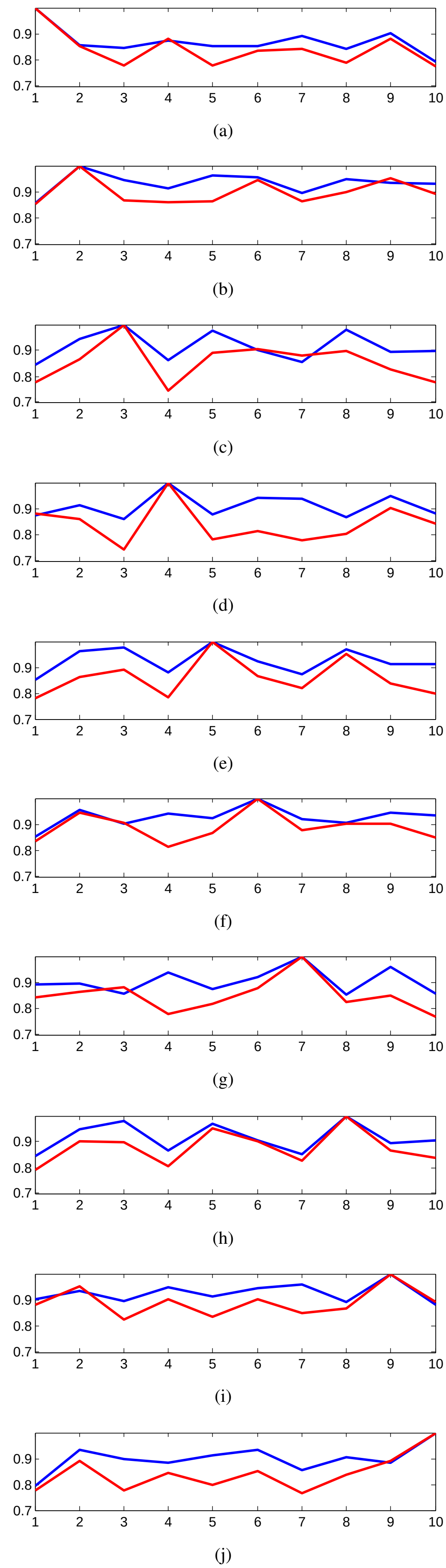

Fig. 8. Correlation between classes: Original image dataset (blue) and Transformed image dataset (red)

TABLE I

RECOGNITION RATE

noise (%)	0	10	20	30	40	50	60	70	80	90	100
Non-transformed	77.81	69.07	62.80	56.31	50.10	45.57	39.74	31.69	21.57	15.24	9.59
Transformed	91.76	87.22	82.66	77.59	72.29	65.81	57.06	46.61	34.31	21.10	15.04

TABLE II

CONFUSION MATRIX (IN PERCENT) BETWEEN THE CLASSES.(NOISE: 50%)

		identified class									
		1	2	3	4	5	6	7	8	9	10
test class	1	98.00	0.14	0.00	0.00	0.00	0.43	0.14	1.14	0.14	0.00
	2	0.86	56.29	4.14	6.71	8.14	7.14	2.86	5.71	0.71	7.43
	3	1.00	2.57	61.29	1.86	9.86	1.14	4.14	9.29	2.71	6.14
	4	3.43	3.43	3.57	65.57	0.71	1.71	2.57	3.29	10.57	5.14
	5	0.57	4.57	22.14	2.86	36.57	4.29	2.29	11.86	3.57	11.29
	6	2.57	13.29	0.86	1.86	3.14	60.57	0.29	10.14	0.29	7.00
	7	2.00	0.86	2.71	2.71	0.57	0.14	79.57	2.86	7.00	1.57
	8	3.14	1.57	5.29	3.86	7.14	3.29	0.71	62.57	6.14	6.29
	9	5.29	0.43	3.29	8.00	2.14	0.14	6.57	5.57	65.43	3.14
	10	0.00	3.29	3.86	1.71	4.00	4.14	0.57	9.86	0.29	72.29

[6] Y. Kessentini, T. Paquet, and A.B. Hamadou, “Off-line handwritten word recognition using multi-stream hidden Markov models”, *Pattern Recognition Letters*, vol. 31 no. 1, pp. 60-70, 2010

[7] A. Elfeki and M. Dekking, “A Markov Chain Model for Subsurface Characterization: Theory and Applications”, *Mathematical Geology*, vol. 33 no. 5, pp. 569-589, 2001

[8] A.J. Viterbi, “Error Bounds for Convolutional Codes and an Asymptotically Optimum Decoding Algorithm”, *IEEE Transactions on Information Theory*, vol. 13 no. 2, pp. 260-269, 1967

[9] C. Berrou, A. Glavieux and P. Thitimajshima, “Near Shannon limit error-correcting coding and decoding: Turbo codes (1)”, *Proceedings of IEEE International Conference on Communications*, Geneva, Switzerland, pp. 1064-1070, 1993

[10] J.J. Hull, “A database for handwritten text recognition research”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 16, no. 5, pp. 550-554, 1994.

[11] N. Otsu, “A Threshold Selection Method from Gray-Level Histograms”, *IEEE Transactions on Systems, Man and Cybernetics*, vol. 9, no. 1 pp. 62-66, 1979