

## CLASSIFICATION OF REDUCTION INVARIANTS WITH IMPROVED BACKPROPAGATION

S. M. SHAMSUDDIN, M. DARUS, and M. N. SULAIMAN

Received 3 November 2000

Data reduction is a process of feature extraction that transforms the data space into a feature space of much lower dimension compared to the original data space, yet it retains most of the intrinsic information content of the data. This can be done by using a number of methods, such as principal component analysis (PCA), factor analysis, and feature clustering. Principal components are extracted from a collection of multivariate cases as a way of accounting for as much of the variation in that collection as possible by means of as few variables as possible. On the other hand, backpropagation network has been used extensively in classification problems such as XOR problems, share prices prediction, and pattern recognition. This paper proposes an improved error signal of backpropagation network for classification of the reduction invariants using principal component analysis, for extracting the bulk of the useful information present in moment invariants of handwritten digits, leaving the redundant information behind. Higher order centralised scale-invariants are used to extract features of handwritten digits before PCA, and the reduction invariants are sent to the improved backpropagation model for classification purposes.

2000 Mathematics Subject Classification: 68T10.

**1. Introduction.** The curse of many or most neural network applications is that the number of potentially important variables can be overwhelming. There are problems whenever we deal with a very large number of variables. These include the sheer size of the computational burden can slow even the fastest computers to the point of uselessness and there can be substantial correlation between variables. The method of principal components is primarily a data-analytic technique that obtains linear transformations of a group of correlated variables such that certain optimal conditions are achieved. The most important of these conditions is that the transformed variables are uncorrelated [7].

Moment invariants have been proposed as pattern sensitive features in classification and recognition applications. Hu (1962) was the first to introduce the geometric moment invariants which are invariant under change of size, translation, and orientation [2]. Moments and functions of moments can provide characteristics of an object that uniquely represent its shape and have extensively employed as the invariant global features of an image in pattern recognition and image classification since 1960s.

This paper discusses the use of principal component analysis to reduce the complexity of invariants dimension for unconstrained handwritten digits and an improved error function of backpropagation model for classification purposes. The rest of the paper's presentation is as follows: [Section 2](#) gives a review on moment invariants and

higher order centralised scale invariants, while Section 3 gives a summary of principal component analysis and its methodology. Section 4 gives an overview of backpropagation model and the proposed error function. Finally, Section 5 shows the experimental results, and Section 6 is the conclusion.

**2. Geometric moment invariants.** The geometric moments (see [2]) of order  $p + q$  of a digital image are defined as

$$m_{pq} = \sum_x \sum_y f(x, y) x^p y^q, \quad (2.1)$$

where

$$p, q = 0, 1, 2, \dots \quad (2.2)$$

The translation invariant central moments are obtained by placing the origin at the centroid of the image,

$$\mu_{pq} = \sum_x \sum_y f(x, y) (x - x_0)^p (y - y_0)^q, \quad (2.3)$$

where

$$x_0 = \frac{m_{10}}{m_{00}}, \quad y_0 = \frac{m_{01}}{m_{00}}. \quad (2.4)$$

Under the scale transformation (the change of size),

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \alpha & 0 \\ 0 & \alpha \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}, \quad \alpha \text{ is a constant}, \quad (2.5)$$

each coefficient of any algebraic form  $f(x, y)$  will be an algebraic invariant, by the definitions of invariants (see [6]),

$$a'_{pq} = \alpha^{p+q} a_{pq}. \quad (2.6)$$

Then, for moment invariants,

$$\mu'_{pq} = \alpha^{p+q+2} \mu_{pq}. \quad (2.7)$$

By eliminating  $\alpha$  between the zeroth-order relation,

$$\mu' = \alpha^2 \mu, \quad (2.8)$$

and the remaining ones, the absolute scale moment invariants is generated as

$$\eta_{pq} = \frac{\mu'_{pq}}{(\mu'_{00})^{((p+q)/2)+1}} = \frac{\mu_{pq}}{\mu_{00}^{((p+q)/2)+1}}, \quad p + q = 2, 3, \dots, \quad (2.9)$$

$$\mu'_{10} = \mu'_{01} \equiv 0, \quad (2.10)$$

in which the image is assumed to have equal scaling in the  $x$  and  $y$  directions.

Consider the following linear transformation which performs different scalings in the  $x$  and  $y$  directions (see [1]),

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \alpha & 0 \\ 0 & \beta \end{bmatrix} \begin{bmatrix} x' \\ y' \end{bmatrix}. \tag{2.11}$$

Improved scale invariants can be derived from (2.9) using higher order centralised moments and algebraic invariants for unequal scaling as (see [5])

$$\eta_{pq} = \left( \frac{\mu_{20}^{p+1/2} \mu_{02}^{q+1/2}}{\mu_{40}^{p+1/2} \mu_{04}^{q+1/2}} \right) \mu_{pq}. \tag{2.12}$$

**3. Principal component analysis.** Principal component analysis (PCA) is a multivariate technique in which a number of related variables are transformed to a set of uncorrelated variables. The starting point for PCA is the sample covariance matrix  $\mathbf{S}$ . For a  $p$ -variable problem (see [3])

$$\mathbf{S} = \begin{bmatrix} s_1^2 & s_{12} & \cdots & s_{1p} \\ s_{12} & s_2^2 & \cdots & s_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ s_{1p} & s_{2p} & \cdots & s_p^2 \end{bmatrix}, \tag{3.1}$$

where  $s_i^2$  is the variance of the  $i$ th variable,  $x_i$  and  $s_{ij}$  is the covariance between the  $i$ th and  $j$ th variables.

The principal axis transformation will transform  $p$  correlated variables  $x_1, x_2, \dots, x_p$  into  $p$  new uncorrelated variables  $z_1, z_2, \dots, z_p$ . The coordinate axes of these new variables are described by the characteristic vectors  $\mathbf{u}_i$  which make up the matrix  $\mathbf{U}$  of direction cosines used in the transformation

$$\mathbf{z} = \mathbf{U}'[\mathbf{x} - \bar{\mathbf{x}}], \tag{3.2}$$

where  $\mathbf{x}$  and  $\bar{\mathbf{x}}$  are  $p \times 1$  vectors of observations on the original variables and their means.

The transformed variables are called the principal components of  $\mathbf{x}$ . The  $i$ th principal component is

$$\mathbf{z}_i = \mathbf{u}'[\mathbf{x} - \bar{\mathbf{x}}]_i, \tag{3.3}$$

and has mean zero, variance  $l_i$ , and the  $i$ th eigenvalue. In other words, the vectors that define the principal components are the eigenvectors of the covariance matrix, and that the eigenvalues are the variances of the principal components. Figure 3.1 is an example of a screen plot for PCA and its eigenvalues. In this paper, we choose PCA of eigenvalue greater than 1.0 as a mean for the extracted components accordingly.

**3.1. Methodology of PCA on unconstrained handwritten digits.** The complexity of data is because of its highly correlation among variables. As such, PCA is meant to choose the variables that really represent it. Thus, in our experiment, we pick variables of correlation greater than 0.7 on unconstrained handwritten digits. We group

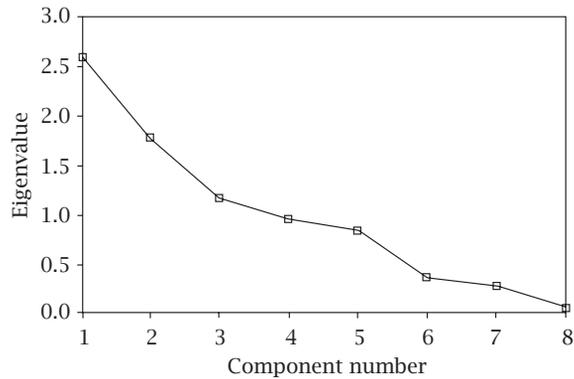


FIGURE 3.1. Principal component and its eigenvalues.

TABLE 3.1. The total variance explanation of PCA.

Component	Initial eigenvalues	% of variance	Cumulative %	Extraction sums of squared loadings	% of variance	Cumulative %
	Total			Total		
1	2.604	32.549	32.549	2.604	32.549	32.549
2	1.775	22.192	54.741	1.775	22.192	54.741
3	1.172	14.654	69.396	1.172	14.654	69.396
4	0.944	11.801	81.197			
5	0.840	10.500	91.697			
6	0.356	4.452	96.150			
7	0.264	3.298	99.447			
8	4.422e-02	0.553	100.000			

these digits based on individual handwritings. Thus we have 200 samples of handwritten digits from five persons. Each individual has 40 samples of handwritten digits with different styles of writing. Table 3.1 shows the total variance explained on unconstrained handwritten digits of group I. The column that represents component is the number of variables, whereby in this study is higher order centralised scale invariants. The percentage of variance informs us that only 3 extracted components are sufficient to represent the whole informations about the invariants of group I.

**4. Backpropagation model.** The backpropagation network is introduced by Rumelhart and McClelland [4]. This network has served as a useful methodology to train multilayered neural networks for a wide variety of applications. The backpropagation model is a supervised learning algorithm using feedforward networks which make use of target values. Backpropagation network is basically a gradient descent method and its objective is to minimize the mean squared error (see Figure 4.1)

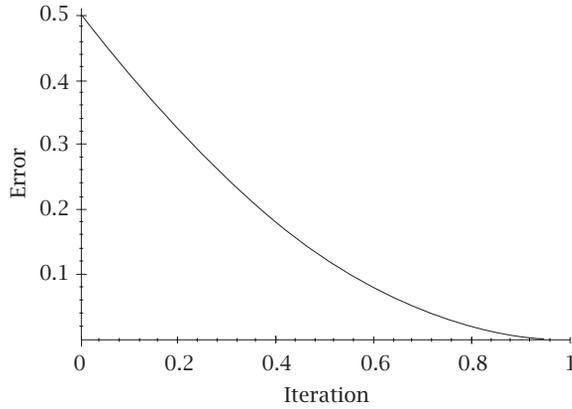


FIGURE 4.1. Mean square error of backpropagation model.

between the target values and the network outputs. Thus the mean square error function (MSE) is defined as

$$E = \frac{1}{2} \sum_k (t_{kj} - o_{kj})^2, \tag{4.1}$$

where  $t_{kj}$  is the target output and  $o_{kj}$  is the network output.

The proposed error function for standard backpropagation ( $mm$ ) is defined implicitly as

$$mm = \sum_K \rho_K \tag{4.2}$$

with

$$\rho_k = \frac{E_k^2}{2a_k(1 - a_k^2)}, \tag{4.3}$$

where  $E_k = t_k - a_k$ ,  $E_k$  is the error at the output unit  $k$ ,  $t_k$  is the target value of the output unit  $k$ , and  $a_k$  is an activation of the unit  $k$ .

The updating weight of standard backpropagation model is

$$\left( \Delta w_k \alpha - \frac{\partial E}{\partial w_k} \right). \tag{4.4}$$

Thus, in this case, (4.4) can be rewritten as

$$\frac{\partial \rho_k}{\partial w_k} = \frac{\partial \rho}{\partial \text{Net}_k} \cdot \frac{\partial \text{Net}_k}{\partial w_k}. \tag{4.5}$$

Knowing that  $\text{Net}_k = \sum_k w_{kj}o_j + \theta_k$ , thus by taking the partial derivative, we have

$$\frac{\partial \text{Net}_k}{\partial w_k} = o_j. \tag{4.6}$$

Substituting (4.6) into (4.5) gives

$$\frac{\partial \rho}{\partial w_k} = \frac{\partial \rho}{\partial \text{Net}_k} \cdot o_j. \quad (4.7)$$

Assume that  $\partial \rho / \partial \text{Net}_k = \delta_k$ . By the chain rule gives

$$\frac{\partial \rho_k}{\partial \text{Net}_k} = \frac{\partial \rho_k}{\partial a_k} \cdot \frac{\partial a_k}{\partial \text{Net}_k}. \quad (4.8)$$

Knowing that  $a_k = f(\text{Net}_k)$ , so for the proposed method, sigmoid function of  $1/1 + e^{-2x}$  is used.

By taking partial derivatives of the  $a_k$ , and simplify it by substituting in terms of  $a_k$ ,

$$\frac{\partial a_k}{\partial \text{Net}_k} = 2a_k(a_k - 1). \quad (4.9)$$

Substituting (4.10) into  $a_k$  gives

$$\frac{\partial \rho_k}{\partial \text{Net}_k} = \frac{\partial \rho_k}{\partial a_k} \cdot 2a_k(a_k - 1). \quad (4.10)$$

It is known that

$$\rho_k = \frac{E_k^2}{2a_k(1 - a_k^2)}. \quad (4.11)$$

Thus by taking partial derivatives with respect to the activation function  $a_k$  gives

$$\frac{\partial \rho_k}{\partial a_k} = \frac{-[(4a_k(t_k - a_k)(1 - a_k^2) + 2(t_k - a_k)^2(1 - 3a_k^2))]}{4a_k^2(1 - a_k^2)^k}. \quad (4.12)$$

Simplifying (4.12), it becomes

$$\frac{\partial \rho_k}{\partial a_k} = -\left(\frac{E + \rho(1 - 3a_k^2)}{a_k(1 - a_k^2)}\right). \quad (4.13)$$

Substituting (4.13) into (4.10) gives the proposed error signal of backpropagation for the output layer as

$$\frac{\partial \rho_k}{\partial \text{Net}_k} = \frac{2(E + \rho(1 - 3a_k^2))}{1 + a_k} = \delta_k, \quad (4.14)$$

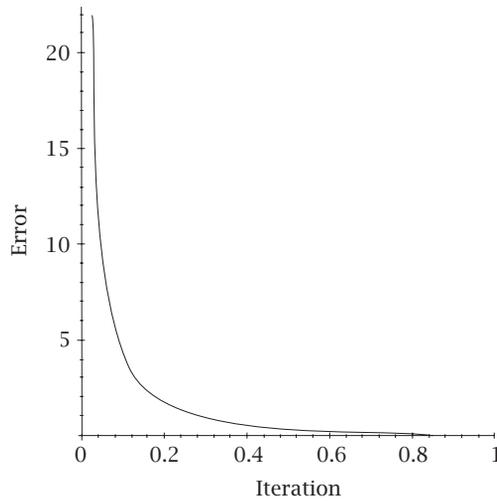


FIGURE 4.2. Proposed error function of backpropagation model.

and an error signal for modified backpropagation of the hidden layer is the same as standard backpropagation,

$$\delta_j = \sum \delta_k w_j f'(a_j). \quad (4.15)$$

A proposed backpropagation error function can be illustrated geometrically as in [Figure 4.2](#), with its errors reducing rapidly compared to MSE, thus giving less iterations for convergence.

**5. Experimental results on unconstrained handwritten digits.** We tested 200 samples of unconstrained handwritten digits from 0 through 9 with various shapes for classifications. Due to computer space and memory limitations, we categorise these samples into five groups, that is, group I-group V. Each group has 40 samples of unconstrained handwritten digits with various shapes and styles of writing. The learning rate and momentum parameter were set to 0.9 and 0.2 for proposed backpropagation with sigmoid as an activation function. Each group has different PCA, therefore, for this paper, we choose group I (see the appendix) as a sample for training using proposed backpropagation. After PCA process of group I, we choose 3 extracted components with eigenvalues greater than 1 which are invariants of the third order with almost 70% of variations.

[Figure 5.1](#) shows the convergence rate for scale invariants using proposed backpropagation for unconstrained handwritten digits before and after the reduction process of group I. Classification rates for proposed backpropagation using scale invariants are successfully recognised after using PCA. [Table 5.1](#) shows the convergence rate and iteration for proposed backpropagation for unconstrained handwritten digits of group I. [Table 5.2](#) shows the variations for each group and number of components extracted from invariants of those handwritten digits accordingly.

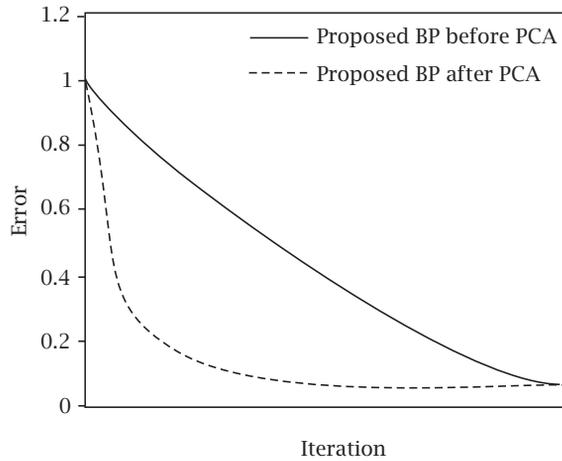


FIGURE 5.1. Convergence rate for handwritten digits before and after PCA of group I.

TABLE 5.1. Convergence rate for handwritten digits before and after PCA of group I.

Unconstrained handwritten digits	Before PCA	After PCA
Proposed BP	Time = 1495 seconds	Time = 1158 seconds
Iterations	15840	9880

TABLE 5.2. Total variations for each group.

Group	Total variations	# Component extracted
Group I	70%	3
Group II	82%	4
Group III	70%	3
Group IV	72%	3
Group V	82%	3

**6. Conclusion.** We presented the use of PCA as a mean to reduce invariants complexity of unconstrained handwritten digits, and classifications of those digits using proposed backpropagation. Higher order centralised scale invariants are used to extract digit images before PCA technique is applied for dimensionality reduction. From the experiments, we find that PCA is able to reduce the number of invariants for these digits without losing their useful informations. In other words, after PCA, we use extracted components of third order moments before we proceed for classifications using proposed backpropagation model. Dimensionality reduction using PCA indirectly saves the computation time and space by using less number of invariants variables for unconstrained handwritten digits. From [Figure 5.1](#) and [Table 5.1](#), we see

that the convergence rate is faster using the extracted components of invariants for these handwritten digits accordingly.

## Appendix

Figure A.1 shows samples of handwritten digits of group I.

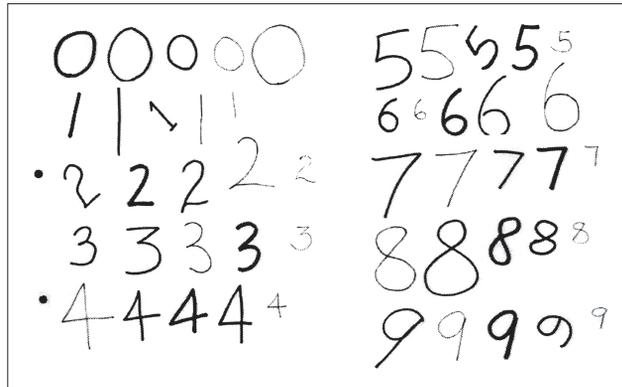


FIGURE A.1. Samples of handwritten digits.

## REFERENCES

- [1] P. Feng and K. Mike, *A new set of moment invariants for handwritten numeral recognition*, IEEE Intern. Conf. on Image Processing, 1994, pp. 154-158.
- [2] M. K. Hu, *Visual problem recognition by moment invariant*, IRE Trans. Inform. Theory IT-8 (1962), 179-187.
- [3] J. E. Jackson, *A User's Guide to Principal Components*, John Wiley & Sons, New York, 1991.
- [4] D. Rumelhart and J. McClelland (eds.), *Parallel Distributed Processing*, MIT Press, Massachusetts, 1986.
- [5] S. M. Shamsuddin, M. N. Sulaiman, and M. Darus, *Improved scale-invariant moments for deformation digits*, Int. J. Comput. Math. 74 (2000), 439-447.
- [6] B. Spain, *Analytical Quadrics*, International Series of Monographs on Pure and Applied Mathematics, vol. 14, Pergamon Press, Oxford, 1960.
- [7] M. Timothy, *Advanced Algorithms for Neural Network: A C++ SourceBook*, John Wiley & Sons, New York, 1995.

S. M. SHAMSUDDIN: FACULTY OF COMPUTER SCIENCE AND INFORMATION SYSTEM, UNIVERSITI TEKNOLOGI, MALAYSIA

E-mail address: [mariyam@fsksm.utm.my](mailto:mariyam@fsksm.utm.my)

M. DARUS: FACULTY OF SCIENCES AND TECHNOLOGY, UNIVERSITI KEBANGSAAN, MALAYSIA

E-mail address: [maslina@pkriscc.ukm.my](mailto:maslina@pkriscc.ukm.my)

M. N. SULAIMAN: FACULTY OF COMPUTER SCIENCE AND INFORMATION TECHNOLOGY, UNIVERSITI PUTRA, MALAYSIA

E-mail address: [nasir@fsktm.upm.edu.my](mailto:nasir@fsktm.upm.edu.my)