DIGIT RECOGNITION USING TRISPECTRAL FEATURES

V. Chandran[1], S. Slomka[1], M. Gollogly[1] S. Elgar[2]

[1]Signal Processing Research Centre, QUT, Brisbane, Qld 4001, Australia
[2]School of EECS, Washington State University, Pullman, WA 99164, U.S.A.

ABSTRACT

Features derived from the trispectra of DFT magnitude slices are used for multi-font digit recognition. These features are insensitive to translation, rotation, or scaling of the input. They are also robust to noise. Classification accuracy tests were conducted on a common data base of 256×256 pixel bilevel images of digits in 9 fonts. Randomly rotated and translated noisy versions were used for training and testing. The results indicate that the trispectral features are better than moment invariants and affine moment invariants. They achieve a classification accuracy of 95% compared to about 81% for Hu's moment invariants and 39% for Flusser/Suk affine moment invariants on the same data in the presence of 1% impulse noise using a 1-NN classifier. A multilayer perceptron with no normalization for rotations and translations yields 34% accuracy on 16×16 pixel low-pass filtered and decimated versions of the same data.

1. INTRODUCTION

1.1. Higher-Order Spectral Features

Higher-order spectra [1] were introduced as spectral representations of cumulants or moments of stationary processes, and are useful in the identification of nonlinear and non-Gaussian random processes. Higher-order spectral representations of real-valued deterministic signals [2, 3, 4] may also be expressed as products of Fourier coefficients. For example, the bispectrum, $B(f_1, f_2)$, of a real-valued sequence, $x(n)$, may be defined by

$$B(f_1, f_2) = X(f_1)X(f_2)X^*(f_1 + f_2) \tag{1}$$

where $X(f)$ is the discrete-time Fourier transform of the sequence at frequency f. The trispectrum may similarly be defined as

$$T(f_1, f_2) = X(f_1)X(f_2)X(f_3)X^*(f_1 + f_2 + f_3) \tag{2}$$

The bispectrum is a function of two frequencies and the trispectrum is a function of three frequencies. In contrast to the power spectrum, these functions are complex-valued in general and retain some of the phase information in the Fourier transform. In particular, for asymmetric sequences the phase is nonlinear and higher-order spectra retain the nonlinear phase information. Like the power spectrum, they are unaffected by a translation of the input. Because of these properties higher-order spectra have been used in pattern recognition [4, 5, 6].

1.1.1. Features from 1D patterns

Bispectral invariant features are defined in reference [4]. Trispectral features [7] are similarly defined as

$$P(\alpha, \beta) = \arctan\left[\frac{I_i(\alpha,\beta)}{I_r(\alpha,\beta)}\right] \tag{3}$$

where

$$I(\alpha,\beta) = I_r(\alpha,\beta) + jI_i(\alpha,\beta) = \int_{f_1=0^+}^{\frac{1}{1+\alpha+\beta}} T(f_1, \alpha f_1, \beta f_1)df_1 \tag{4}$$

is the integral along a line in trifrequency space with slope determined by α and β. The region of trifrequency space used for feature calculations is shown in figure 1. The integration preserves the translation invariant property of the trispectrum. Invariance to DC level shift is gained by integrating only along lines for f_1 greater than zero. Scale invariance is also achieved because scaling of the sequence $x(n)$ will result in a scaling of the Fourier transform. The magnitude of $I(\alpha, \beta)$ will be changed, but the phase will remain the same. Similarly an amplification of $x(n)$ will change only the magnitude, not the phase, of $I(\alpha, \beta)$. Formal proofs of invariance properties can be found in reference [4].

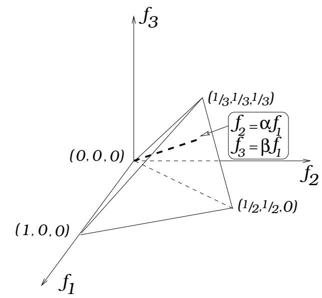

Figure 1. Trifrequency space showing the region used for feature calculations. The frequency axes, f_1, f_2, and f_3 are normalised by the Nyquist frequency.

1.1.2. Indirect Procedure Using the DFT Magnitude

Finite width patterns are not strictly bandlimited. Scale invariance relies on the finite bandwidth property. Consequently, to make the features from finite width patterns better invariant to scale an indirect procedure is adopted. The pattern is discrete Fourier transformed and the positive frequency half of the transform magnitude is used as the input sequence for feature extraction instead of the original pattern itself. The power spectrum may be used in place of the DFT magnitude, as there is a one-to-one relationship between the two. The transform of this new sequence is an analytic function whose real part is the autocorrelation of the original input sequence and whose imaginary part is the Hilbert transform of the autocorrelation of the original input sequence (except for an amplitude scale factor of 1/2 in each case). This sequence is then used in the products defining the bispectrum and the trispectrum in equations 1 and 2.

This procedure results in a loss of uniqueness for 1D patterns, and the inability to distinguish between $x(n)$ and $x(-n)$. However, the direct procedure can always be used in combination to retain uniqueness if necessary, as is done for the classification of 6 and 9 in this work. The sign of the bispectrum or the trispectrum (as defined in equations 1 and 2) depends on the left or right asymmetry of the input sequence.

1.1.3. Features from 2D images

An algorithm for 2D object recognition using features derived from the bispectrum was introduced in [6]. It decomposes the image into a set of one-dimensional functions via the Radon transform. The Radon transform projections have the following properties

(a) a shift in the 2D image results in a shift in every projection, except for the one parallel to the direction of the shift, which is unchanged,

(b) a scale change of the 2D image in the direction of projection results in multiplication of the 1D projection by a constant,

(c) a scale change of the 2D image perpendicular to the direction of projection results in a scale change of the 1D projection, and

(d) a rotation of the 2D image results in a cyclic shift in the set of projection functions.

The algorithm can also use trispectral features and extract $P(\alpha, \beta)$ from the projections. These features are invariant to translation, scaling, and amplification of the 1D projections. They provide a set $\{P(\alpha, \beta)(\theta)\}$ of 1D functions of θ, invariant to translation, scaling, and amplification of the image. Regrouping the features for each (α, β) pair results in sets of one-dimensional functions of θ where rotation of the image is equivalent to a translation of these sequences. A second stage of invariant feature extraction from these sequences will produce a new set of features $P(\alpha, \beta, \gamma, \delta)$ that also provide invariance to rotation of the image.

1.1.4. Previous work on trispectral features

The algorithm was extended to trispectral features in [7] and compared with Hu's moment invariants [8]. The focus of these studies was on testing the immunity to various types of noise. Based on cluster plots and accuracy results, trispectral features were shown to be better than moment invariants [9] in the presence of Gaussian, Uniform, and Rayleigh iid noise.

2. MULTIFONT DIGIT RECOGNITION

The present study investigates the discriminating power of the trispectral features using the problem of multifont digit recognition. A prototype data base of 9 printed fonts and 10 digits in each font was created. Each prototype was an image of 256×256 pixels, and bilevel with grelevels 0 or 255. For training and testing, 3 randomly translated and rotated versions of each prototype were created. Random impulse noise was added to generate 4 noisy versions of each digit in each font and each orientation. The effect of the impulse noise used is to set any pixel to the maximum greylevel of 255 with a prescribed probability. Effectively, a prescribed number of pixels, such as 1%, get set to this value. This yielded 144 realizations per digit. Since most artificial neural networks do not accept large sized inputs, a second data base of the same size was created by lowpass filtering and decimating the clean images to 16x16 pixels. These images

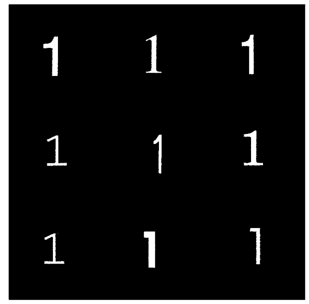

Figure 2. The digit '1' in each of the nine fonts used.

were greylevel in the range 0 to 255. The same level of impulse noise was also added to the decimated images. Accuracy results were obtained by using 3 orientations to train and the fourth orientation to test. Confusion matrices for each of the four possible splits of the data were averaged to obtain the results shown in the tables. Since 6 and 9 are rotated versions of each other in most fonts, they are considered to be in the same class. A technique to classify between 6 and 9 using higher order spectral features is described in a later section.

3. COMPARISON OF CLASSIFICATION ACCURACY

This comparison was performed using 1% impulse noise in both training and test data. 144 trispectral features were extracted from each image. Similarly, 7 moment invariants [8] and 4 affine invariant moments [10] were extracted for the same data. All features were used and a 1-NN classifier was trained in each case. The results are therefore indicative of the quality of the features in terms of discriminating power and robustness to noise. They are independent of the classifier and of any feature selection process.

	0	1	2	3	4	5	6	7	8
0	97.2						2.1		0.7
1		98.6				0.7		0.7	
2	2.8		93.7					3.5	
3			0.7	88.2		2.1	7.6		1.4
4					100				
5	1.4			2.1		95.8	0.7		
6			0.7	1.05			97.9		0.35
7		4.9	3.5					91.6	
8				2.1			1.4		96.5

Table 1. Confusion Matrix for classification using the proposed trispectral features. Blank entries are zeroes. Accuracies are in percentages.

	0	1	2	3	4	5	6	7	8
0	83.3	1.4	0.7	0.7	4.2	1.4	5.6	1.4	1.4
1		92.4	2.1	2.1		2.8		0.7	
2	0.7	0.7	88.2	7.6			1.4	1.4	
3			4.9	70.1	6.3	9.7		7.6	1.4
4	4.9		1.4	7.6	82.6		2.1		1.4
5	4.2	0.7		9.0		68.8	14.6	2.8	
6	3.1		2.4		0.7	5.9	81.3	1.4	5.2
7	1.4		2.1	7.6			5.6	81.2	2.1
8	0.7			1.4	1.4		11.1	0.7	84.7

Table 2. Confusion Matrix for classification using Hu's moment invariants.

	0	1	2	3	4	5	6	7	8
0	34.7	4.2	5.6	11.8	4.2	4.2	18.1	7.6	9.7
1	3.5	45.8	8.3	3.5	1.4	4.2	8.3	21.5	3.5
2	6.2	9.0	35.4	11.1	4.2	19.4	9.7	6.9	2.8
3	10.4	4.2	15.3	43.7	2.8	12.5	5.6	1.4	4.2
4	6.9	2.1	8.3	3.5	25.7	13.2	24.3	4.9	11.1
5	9.7	2.1	13.9	12.5	15.7	25.7	16.7	0.7	6.9
6	9.4	5.6	3.8	2.4	9.7	6.9	46.2	5.2	10.8
7	5.6	22.2	2.1	2.8	3.5		10.4	46.5	4.9
8	11.8	2.1	2.8	0.7	7.6	7.6	19.4	5.6	43.8

Table 3. Confusion Matrix for classification using affine moment invariants.

4. COMPARISON WITH ARTIFICAL NEURAL NETWORKS

4.1. Multilayer Perceptron

The Multilayer Perceptron (MLP) trained using backpropagation can be used for classfication of digits but it does not yield good results unless the input can be normalised for translations and rotations. A lower resolution data base of the same data used for testing trispectral and moment features was created to compare with those obtained from a fully connected MLP with one hidden layer. The 256x256 pixel images were low pass filtered and decimated to yield 16×16 pixel grayscale images. The resulting images are small enough to be input straight into an MLP without requiring exorbitantly large number of interconnections and consequently training times. The same testing procedure as for the previous comparisons mentioned above was used to obtain the confusion matrix shown in table 4. The results confirm the well known fact that the MLP cannot generalize very well to large unseen translations and rotations of a pattern. Most successful methods of digit and character recognition using MLPs normalize the input before presenting to the network.

4.2. Modified MLPs

Modified MLPs such as the Le Cun architecture [11] have obtained higher classification accuracies for handwritten digit classification than those obtained with the proposed trispectral features. However, they (a) require preprocessing to normalize the inputs, (b) can only work with lower resolution data at practical network sizes, (c) require training with large number of inputs, and (d) are not fully rotation invariant.

	0	1	2	3	4	5	6	7	8
0	58		11		6	8	3	11	3
1	3	31	3	8	6	6	3	36	6
2		11	19	8	11	6	6	19	19
3		6	11	8	8	17	3	25	22
4		8	3		58	8		6	17
5		6	11	6	11	17	6	31	14
6	7	8	10	6	13	15	10	7	25
7		8	8	6	3			72	3
8	3		3	6	25	14	11	6	33

Table 4. Confusion Matrix for classification using an MLP. Accuracies shown are in percentages.

4.3. Neocognitron

The Neocognitron [12, 13] achieves high classificaton accuracies for digit recognition and is very robust to translations, scale changes, deformations in shape, noise and missing data. However, it is not completely rotation invariant. It is also very computationally intensive, requiring tens of million interconnections even for 19×19 pixel inputs. The proposed trispectral feature extraction has an order of complexity, roughly $N_s N \log_2 N N_f$, where N_s is the number of slices, N is the size of input, and N_f is the number of features. For 32×32 inputs and 20 features, this is about one tenth of a million. The Neocognitron is not inherently suited for greyscale data with poor signal to noise ratio, while the proposed features have excellent immunity to greyscale data with noise.

5. CHANGING THE DIMENSION OF THE TRISPECTRAL FEATURE VECTOR

To investigate the effects of changing the dimension of the trispectral feature vector, fully connected MLPs with one hidden layer were used. A different MLP structure was constructed for each feature vector dimension. These MLPs were trained with two rotations and tested with two distinct rotations. The 9 fonts were used and the digits 6 and 9 were again combined into one class. The table below shows that the dimension of the feature vector can be reduced from 144 to 10 with only a small reduction in accuracy.

No. of Features	5	10	20	50	144
Accuracy	74.3%	89.0%	87.8%	89.7%	92.9%

Table 5. Effect of changing the dimension of the trispectral feature vector.

6. CLASSIFICATION OF 6 AND 9

Full rotation invariant features cannot distinguish between 6 and 9. Higher order spectral features can be used to decide between the two by using only one or two projections, and the direct procedure described in section 1.1. Projections along the x axis or the y axis show different types of symmetry for 6 and 9.

Figure 3. Digit '6' in one of fonts with 1% impulse noise.

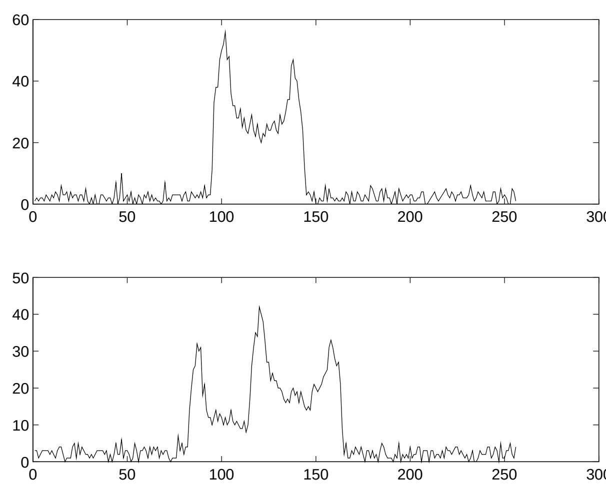

Figure 4. Projections along the x and y directions for the image in figure 3.

Figure 5. Digit '9' in the same font with 1% impulse noise.

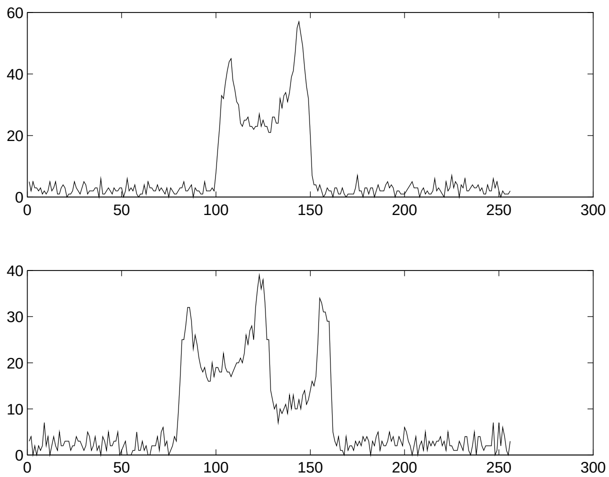

Figure 6. Projections along the x and y directions for the image in figure 5.

The phase of the bispectrum changes sign when the input changes from left-asymmetric to right-asymmetric. The use of a single bispectrum value is susceptible to noise and scale changes. The phase of the integrated bispectrum is insensitive to these effects. The phase of the bispectrum of projections along the x and y axes, integrated along the 45 degree line in bifrequency space was used for classification accuracy testing. The use of two projections takes care of fonts for which the x axis projection may be symmetric. It also permits a choice depending on the magnitude of the imaginary part. By choosing the projection which yields a larger imaginary part of the integrated bispectrum, the decision is more robust to noise.

Classification accuracy results for upright 6 and 9 in the nine different fonts for 1%, 2% and 5% impulse noise are given below.

Noise level	1%	2%	5%
Error	1.11%	1.7%	7.8%

Table 6. Error in classification of 6 and 9 using bispectral features. The results are based on a test of 180 upright images at each noise level.

7. CONCLUSION

A technique for extracting translation, rotation and scale insensitive features from two dimensional patterns is presented and used for classification of digits. The features are shown to posses better discriminating ability and noise immunity than moment invariants and affine moment invariants, on a common data base of digits in nine fonts. The technique is superior to artificial neural network approaches in accommodating high resolution inputs and large rotations, requiring less training and fewer computations.

Acknowledgements

S. Elgar and V. Chandran acknowledge funding from the US office of Naval Research. The authors are grateful to the Center for Intelligent Robotics, Universidade Nova de Lisboa, Portugal for the use of their public domain pattern recognition toolbox *TOOLDIAG*.

REFERENCES

[1] D. Brillinger, "An introduction to polyspectra," *The Annals of mathematical statistics*, vol. 36, pp. 1351–1374, 1965.

[2] C. Nikias and M. Raghuveer, "Bispectrum estimation: A digital signal processing framework," *Proceedings of the IEEE*, vol. 75, pp. 869–889, 1987.

[3] B. Sadler, "Shift and rotation invariant object recognition using the bispectrum," in *Workshop on HOS analysis*, (Vail, Colorado), pp. 106–111, 1989.

[4] V. Chandran and S. Elgar, "Pattern recognition using invariants defined from higher-order spectra–one-dimensional inputs," *IEEE Transactions on Signal Processing*, vol. 40, pp. 205–212, January 1991.

[5] M. Tsatsanis and G. Giannakis, "Object and texture classification using higher order statistics," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, pp. 733–750, July 1992.

[6] V. Chandran and S. Elgar, "Position, rotation, and scale invariant recognition of images using higher-order spectra," in *Proceedings of ICASSP '92*, (San Francisco), 1992.

[7] B. Carswell and V. Chandran, "Trispectral features for object recognition in the the presence of additive gaussian noise," in *Proceedings of DICTA '95*, (Brisbane, Australia), 1995.

[8] M. Hu, "Visual pattern recognition by moment invariants," *IRE Transactions on Information Theory*, vol. IT-8, pp. 179–187, February 1962.

[9] V. Chandran, B. Carswell, B. Boashash, and S. Elgar, "Pattern recognition using invariants defined from higher-order spectra - two dimensional inputs," *IEEE Transactions On Image Processing*, accepted.

[10] J. Flusser and T. Suk, "Affine moment invariants: a new tool for character recognition," *Pattern Recognition Letters*, pp. 433–436, April 1994.

[11] Y. L. Cun and et al, *Handwritten Digit Recognition using a Backpropagation Network*, vol. 2. San Mateo, California: Morgan Kaufmann Publishers, 1990.

[12] K. Fukushima, S. Miyake, and T. Ito, "Neocognitron: A neural network model for a mechanism of visual pattern recognition," *IEEE Transactions on Systems. Man and Cybernetics*, vol. SMC-13, no. 5, pp. 826–834, 1983.

[13] K. Fukushima and S. Miyake, "Neocognitron: A new algorithm for pattern recognition tolerant of deformations and shifts in position," *Pattern Recognition*, vol. 15, no. 6, pp. 445–469, 1984.