Separating desired image and signal invariant components from extraneous variations

William J. Williams
Eugene J. Zalubas
Alfred Hero III

Electrical Engineering and Computer Science Department
University of Michigan
Ann Arbor Michigan 48109

ABSTRACT

Images and signals can be characterized by representations invariant to time shifts, spatial shifts, frequency shifts, and scale changes as the situation dictates. Advances in time-frequency analysis and scale transform techniques have made this possible. The next step is to distinguish between invariant forms representing different classes of image or signal. Unfortunately, additional factors such as noise contamination and "style" differences complicate this. A ready example is found in text, where letters and words may vary in size and position within the image segment being examined. Examples of complicating variations include font used, corruption during fax transmission, and printer characteristics. The solution advanced in this paper is to cast the desired invariants into separate subspaces for each extraneous factor or group of factors. The first goal is to have minimal overlap between these subspaces and the second goal is to be able to identify each subspace accurately. Concepts borrowed from high-resolution spectral analysis, but adapted uniquely to this problem have been found to be useful in this context. Once the pertinent subspace is identified, the recognition of a particular invariant form within this subspace is relatively simple using well-known singular value decomposition (SVD) techniques.

Keywords: Scale transform, Mellin transform, Wordspotting, Singular value decomposition, Character recognition, Time-frequency.

1 INTRODUCTION

The recognition of specific signatures in images and signals has long been of interest. Powerful techniques exist for their detection and classification, but these techniques are often defeated by changes or variations in the signature. These variations often include translation and scale changes. Methods exist for transforming the signal/image so that the result is invariant to these disturbances. Translation and scaling are well understood in a mathematical sense, so it is fairly straightforward to design methods which yield a transformed form of the data wherein these effects are removed. There are other variations which are not well described mathematically or are not mathematically tractable in terms of reasonable transformations. This paper describes a combination of techniques which allow scale and translation invariant transformations to be used as one step of the signature

0-8194-2234-7/96/$6.00

recognition process. This is followed by an approach which separates the entities to be classified into a number of subsets characterized by additional variations. A method is provided to identify the subset to which the specific entity at hand belongs so that classifiers specific to that subset can be used. A two dimensional image is the basic starting point for the technique. This may be the actual image of an object or the two dimensional form of a signal representation such as a time-frequency distribution.

Classification of alphabetical characters of different fonts and sizes serves to illustrate the methods developed. However, the approach is quite general and may be applied to a variety of problems and signals of any dimension including time-frequency and time-space-frequency representations.

A representation termed the Scale and Translation Invariant Representation (STIR) is introduced here. It has desirable properties for pattern recognition under certain conditions. The object to be recognized must have consistent shape and appear on a constant intensity background. Using autocorrelation and the scale transform STIRs of patterns are identical for examples that have been translated on the background or scaled (compressed or dilated) along one or more axes.

Concepts borrowed from high-resolution spectral analysis, but adapted uniquely to the problem of classifying these STIRs have been found to be useful in this context. In high resolution frequency estimation, the noise subspace eigenvectors of the autocorrelation matrix are used. Pisarenko harmonic decomposition[7] employs the orthogonality of the noise subspace to the signal vectors to estimate sinusoid frequencies. This idea is used in the classification of signals following STIR processing.

A standard approach to classification is to use the training data to generate templates for each class. A similarity measure, such as correlation coefficient, between the processed test data and each template is calculated and the test data is declared to be in the class corresponding to the largest similarity measure. In this noise subspace approach, an orthogonal subspace is created for each class of training data. A measure of the projection of the test data onto each of these subspace is calculated. Test data matching a given class should be orthogonal to the noise subspace for that class and yield a small projection measure.

The STIR and noise subspace classification method are applied to the example of character recognition. For a bitmap character input, the data are represented invariantly to translation and size, then categorized by font, and finally classified by letter. This combination of methods is applicable to many pattern recognition problems of any dimension.

2 MATHEMATICAL TOOLS FOR REPRESENTATION

Three distinct mathematical tools are used in the development of the method. All may be applied to suitably transformed data of any dimensionality. For simplicity of explanation the case for a 2D input signal is presented. The one dimensional case is an obvious simplification and no additional issues are raised in the use of higher dimensional signals, so explanation based on 2D signals adequately describes the characteristics of the method. In addition, bitmap character data is used to demonstrate characteristics of the representations. This ties in with the example application of character recognition presented in a later section.

The mathematical tools employed are:

- 2D autocorrelation of the 2D representations to remove translational effects.
- 2D scale transformations of the autocorrelation result to remove scaling effects.
- Partition of the results into subsets which reflect extraneous variations of the data.

Classification of the image involves two steps. These are:

- Determine the subset to which the unknown image belongs.
- Use the classifier designed for that specific subset to classify the image.

2.1 Time-frequency

The Wigner distribution has aroused much interest in the signal processing community. However, its use in image processing has been limited. Jacobson and Wechsler[3–5] have pioneered in the use of such techniques in image processing. Cristobal et al[2] have investigated the use of Wigner distributions in image pattern recognition. Jacobson and Wechsler apparently have had a keen interest in human perception and the means by which images are perceived.

A more recent paper by Reed and Wechsler[8] discusses the use of Wigner-based techniques to realize the Gestalt laws that resulted from perceptual grouping in the 1920s. It was suggested at this time that individual elements appear to group according to a set of principles including proximity, similarity, good continuation, symmetry and common fate. Reed and Wechsler go on to show that applying a relaxation procedure to the primary frequency plane of the 2D Wigner distribution is useful. Selection of the primary frequency plane reduces the representation from a N x N x N x N representation to a N x N frequency representation, of the same dimension as the original image. This is achieved by selecting pixels according to their maximum energies and retaining a number of top ranked frequencies. Then, regions of homogeneity are grouped together. They also show that this process produces a similar end result for image textures that have the various Gestalt properties in common. This work is interesting and deserves further attention. One may conclude that the surface has been barely scratched in the application of space-spatial frequency techniques to images in general. The application to word spotting seems obvious. There are a number of new time-frequency techniques such as RIDs[9] that offer a considerable improvement over the Wigner distribution. However, for the present paper a less ambitious approach using only autocorrelation is used.

2.2 Computation of the 2D autocorrelation

The scale transform has been introduced by Cohen.[1] It is a specific case of the Mellin transform with many useful and interesting properties. Signals normalized to equal energy which differ only in scale have coefficients of scale identical within a phase factor. This scale invariance property of the scale transform permits direct comparison of signals of different scale. The magnitude of the scale transform is invariant to scaling of signals, provided the transform is applied from the origin of scaling. Determining the origin from which scaling occurs is often difficult, especially when noise is present. The autocorrelation function of the signal provides a stable origin. Since the autocorrelation simultaneously sums over all points of a function, shifting of a signal on a plane does not affect the values for each lag. It is well known that autocorrelation will remove translational effects in images and specifically in optical character recognition (OCR) methods.[6]

The 2D autocorrelation may be carried out as follows:

$$A(k_1, k_2) = \sum_{n_1} \sum_{n_2} a(n_1, n_2) a(n_1 - k_1, n_2 - k_2) \tag{1}$$

where $a(n_1, n_2)$ is the image. The image need not be centered within the bitmap representation, which has finite support in n_1, n_2. Consider, for purposes of exposition, that the bitmap is infinitely padded with zeros outside of the specific bitmap support region chosen.

The 0,0 lag point provides an origin from which the autocorrelation function scales. Another feature of the 2D autocorrelation function is the symmetry $A(k_1, k_2) = A(-k_1, -k_2)$. Hence, the first and fourth quadrants together contain complete information about the entire autocorrelation lag plane. This attribute will be used in applying the scale transform. Typical autocorrelation functions for lowercase letters are shown in Figure 1

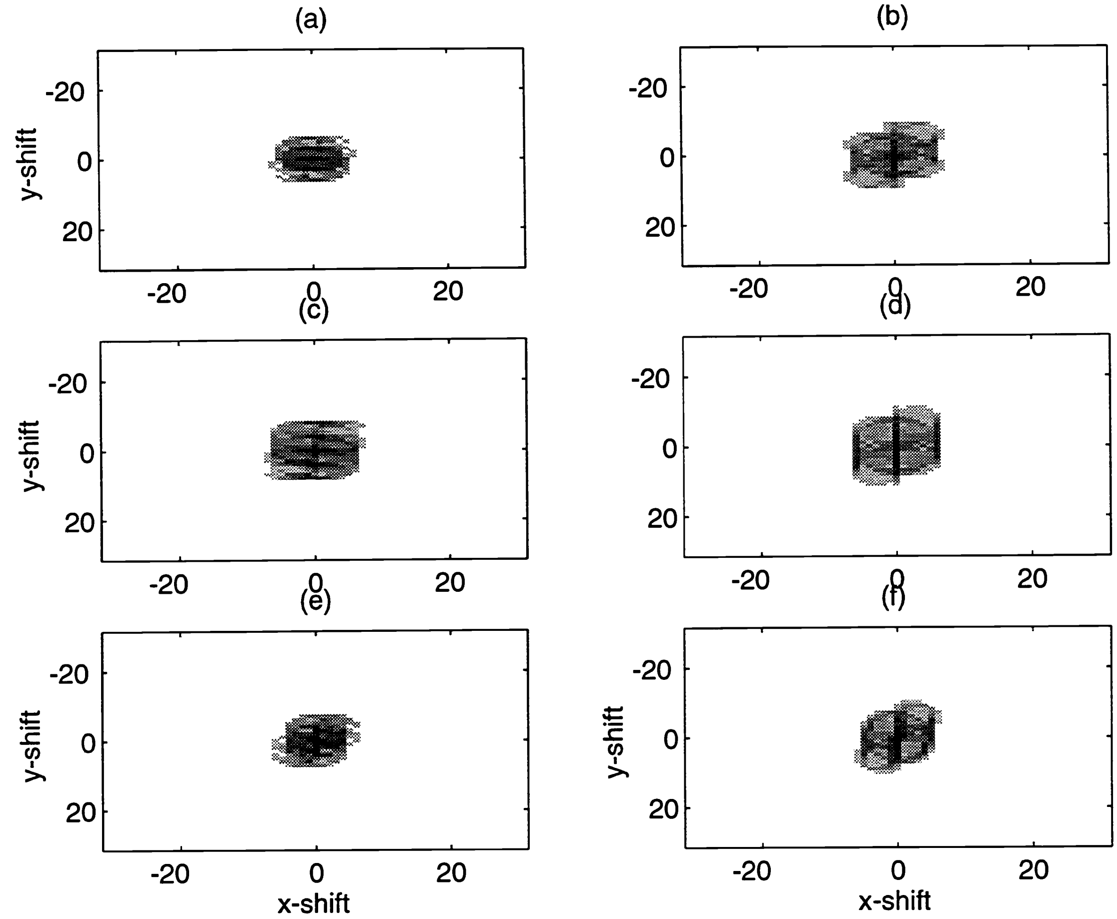

Figure 1: 2D autocorrelation result for a 63x63 pixel bitmap of two lowercase letters represented in normalized log form. (a) 'a' in Courier(12pt) , (b) 'b' in Courier(12pt), (c) 'a' in Helvetica(12pt), (d) 'b' in Helvetica(12pt), (e) 'a' in Times(12pt) , (f) 'b' in Times(12pt).

For pattern recognition purposes, one must be aware of the loss of information which results from obtaining the autocorrelation of the signal. One wishes to remove only translation effects. Unfortunately, due to the symmetry of the autocorrelation function, an ambiguity in the orientation of the original image is introduced. The autocorrelation of an image is indistinguishable from the autocorrelation of a 180 degree rotated version of the image. This is due to the masking of phase information when the autocorrelation is applied to a signal.

2.3 Direct scale transform

The 2D autocorrelation function provides invariance to translation and a stable origin. A properly applied discrete scale transform implementation can additionally provide the desired scale invariance.

The scale transform is defined in the continuous domain as[1]

$$D(c) = \frac{1}{\sqrt{2\pi}} \int_0^{\infty} f(t) \frac{e^{-jc\ln t}}{\sqrt{t}} dt \tag{2}$$

Using a direct expansion of the scale transform, which has some advantages over the previous discrete scale transform reported by our group,[10] a new discrete approximation is obtained which avoids the problem of interpolating and exponentially resampling the data. Let $t = e^x$, $dt = e^x dx$. So $\sqrt{t} = e^{t/2}$. As a function of x, the scale transform becomes

$$D(c) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(e^x) e^{(1/2-jc)x} dx \tag{3}$$

Signals are commonly sampled at uniform intervals. Since the value of the function is not known for all instants, calculations must be performed based on values at the sampled points of the signal. In this discrete approximation the signal is assumed to remain constant between samples.

Assume that the signal is sampled every T units. Since, by linearity, an integral may be broken into a summation of integrals over smaller regions, we may break the integral into logarithmic intervals. This choice of interval width in conjunction with the approximation of constant value over each interval permits calculation using only the sampled points of $f(t)$. Splitting the integral into logarithmic intervals yields

$$\begin{aligned} D(c) \quad = \quad & \frac{1}{\sqrt{2\pi}} \left\{ \int_{-\infty}^{\ln T} f(e^x) e^{(1/2-jc)x} dx + \int_{\ln T}^{\ln 2T} f(e^x) e^{(1/2-jc)x} dx + \right. \\ & \left. \int_{\ln 2T}^{\ln 3T} f(e^x) e^{(1/2-jc)x} dx + \int_{\ln 3T}^{\ln 4T} f(e^x) e^{(1/2-jc)x} dx + \ldots \right\} \end{aligned} \tag{4}$$

$$\begin{aligned} D(c) \quad \approx \quad & \frac{1}{\sqrt{2\pi}} \left\{ f(e^{-\infty}) \int_{-\infty}^{\ln T} e^{(\frac{1}{2}-jc)x} dx + f(e^{\ln T}) \int_{\ln T}^{\ln 2T} e^{(\frac{1}{2}-jc)x} dx + \right. \\ & \left. f(e^{\ln 2T}) \int_{\ln 2T}^{\ln 3T} e^{(\frac{1}{2}-jc)x} dx + f(e^{\ln 3T}) \int_{\ln 3T}^{\ln 4T} e^{(\frac{1}{2}-jc)x} dx + \ldots \right\} \end{aligned} \tag{5}$$

$$\begin{aligned} D(c) \quad \approx \quad & \frac{1}{\sqrt{2\pi}} \left\{ f(0)[\frac{(T)^{1/2-jc}}{\frac{1}{2}-jc}] + f(T)[\frac{(2T)^{\frac{1}{2}-jc} - (T)^{\frac{1}{2}-jc}}{\frac{1}{2}-jc}] + \right. \\ & \left. f(2T)[\frac{(3T)^{\frac{1}{2}-jc} - (2T)^{\frac{1}{2}-jc}}{\frac{1}{2}-jc}] + f(3T)[\frac{(4T)^{\frac{1}{2}-jc} - (3T)^{\frac{1}{2}-jc}}{\frac{1}{2}-jc} \ldots \right\} \end{aligned} \tag{6}$$

Simplifying the above equation yields the following algorithm for direct calculation of the discrete scale transform

$$D(c) \approx \left(\frac{1}{(1/2-jc)\sqrt{2\pi}} \right) \sum_{k=1}^{\infty} [f(kT-T) - f(kT)] \, (kT)^{1/2-jc} \tag{7}$$

Since the scale transform is based on exponential sampling relative to the origin, the entire autocorrelation plane cannot be dealt with at one time. Since both lag values in the first quadrant index from zero in the first quadrant, the scale transform may be directly applied. The lag axes in the fourth quadrant, however, aren't both positive, so reindexing is necessary. For each quadrant the axes must be included, since the scale transform indexing is based relative to the origin. Hence, define two discrete quadrant functions as follows:

$$Q_1(k_1, k_2) = A(k_1, k_2) \quad \text{for } k_1,k_2 \geq 0 \tag{8}$$

$$Q_2(k_1, k_2) = A(k_1, -k_2) \quad \text{for } k_1,k_2 \geq 0 \tag{9}$$

Since it is not possible to calculate the scale coefficient $D(c)$ for every scale, c, a set of scales is chosen. The scales and interval parameter, T are selected such that a unique representation is generated for each class of inputs.

A 2D scale transform approximation is implemented by applying the 1D scale transform algorithm in (7) first to the rows then to the columns of a matrix of values. Applying such a 2D scale transform to Q_1 and Q_2 and taking the magnitude of the result yields two 2D matrices of scale coefficients. The size of these matrices is determined by the number of row and column scale values selected.

Since the autocorrelation function input was not energy normalized, normalization of the scale magnitudes is required for a scale invariant representation. Since the scale transform is a linear transform, normalization may be done by a variety of methods to generate an appropriate result.

The normalized scale transformed quadrant functions represent a STIR of the original 2D input. Since only selected scale transform magnitudes are computed, the transform is not invertible. In addition to providing a scale invariant representation, other signal information is lost. The usefulness of the STIR is dependent on its implementation and application. For the very common case of a 2D function sampled into a matrix of discrete values, we have developed a usable classification scheme which can be used with STIRs as the inputs.

3 CLASSIFICATION OF PATTERNS

In addition to the invariances, STIRs have the desirable property that for a fixed set of row and column scales the sizes of all STIR matrices are identical, regardless of the size of the input matrices. Hence, inputs from different sources may be treated identically once processed into STIR images.

The initial approach taken to classify patterns was to decompose the STIR images to provide an orthonormal set of descriptors. The Karhonen-Loève transform is commonly mentioned as a means of accomplishing this. In OCR methods this is a well known approach.[6] The singular value decomposition (SVD) provides equivalent results. The STIRs of each character were reshaped into a single vector by concatenating the rows of the two STIR matrices. The specific mapping of elements from the matrices to row vector is of no importance as long as the values are unchanged. This vector contains all the information of the STIR, but in a more convenient form for processing. The row vectors were 'stacked' forming a new matrix representing all characters of interest for a range of sizes and various fonts. The SVD was then applied to extract essential features of the set of vectors. Right singular vectors corresponding to the largest singular values were chosen as features. Correlation coefficients between test STIR vectors and the selected features were used for classification. Unfortunately, in the case of character recognition, classification results were not impressive.

An idea may be borrowed from matrix theory and methods developed for high resolution spectral analysis. Suppose one has C characters to be recognized and V variations of those characters. Then there are $M = CV$

exemplars of the members of the subset. When the SVD analysis is performed on each character individually, V singular vectors provide a complete basis for the set of exemplars. The number of elements in the exemplar vectors is set to be larger than the number of exemplar vectors by selecting the row and column scales in the scale transform. The SVD will yield a larger number of orthonormal vectors than there are exemplars. The $V + 1th$ singular vector will be orthogonal to the singular vectors which form a basis for the set of V exemplars as will be the $V + 2, V + 3, V + 4, ..$ singular vectors. These are so-called 'noise' vectors. The inner product of any one of these noise vectors with any of the basis vectors will yield a zero result. Since each exemplar is formed from the set of basis vectors, it too will yield a zero inner product with the noise vectors. This provides a method for character classification. Take the inner product of the test STIR vector with a noise vector of the candidate character subset. If the result is zero, then the STIR vector must be a member of the corresponding character subset. In practice, due to noise or variations in the bitmaps one may not obtain a zero inner product with the correct character noise vectors. However, the inner product of the unknown character vector with the noise vectors correct character subset will produce the smallest magnitude result when compared to inner products between the unknown character and noise vectors of the incorrect character subset.

4 APPLICATION: MULTIFONT CHARACTER CLASSIFICATION

This example shows how STIR and the SVD noise subspace index are combined to perform as a size independent multifont character classifier. A complete character recognition system incorporates much more than the pattern classifier presented here. This application is presented to show the viability of the method for pattern classification. The approach taken for identification of characters in various fonts consists of two steps: font determination and character identification. Both steps use the STIRs and noise subspace methodology. Specifically omitted is the significant task of segmentation of an image into individual character bitmaps. Each bitmap is considered to be an isolated recognition task. Contextual information such as positioning within a word or line, adjacent characters, and character frequency is not used.

The character set consists of the lowercase letters. Courier, Helvetica, and Times were the fonts examined. Text in sizes 14, 18, 24, and 28 point in each font was used for training. Bitmaps from faxed versions of clean printed copy were used as the input signals. The text consisted of one instance of each character in each font and size combination. Bitmaps of the letter 'e' in the various font and training size combinations appear in Figure 2.

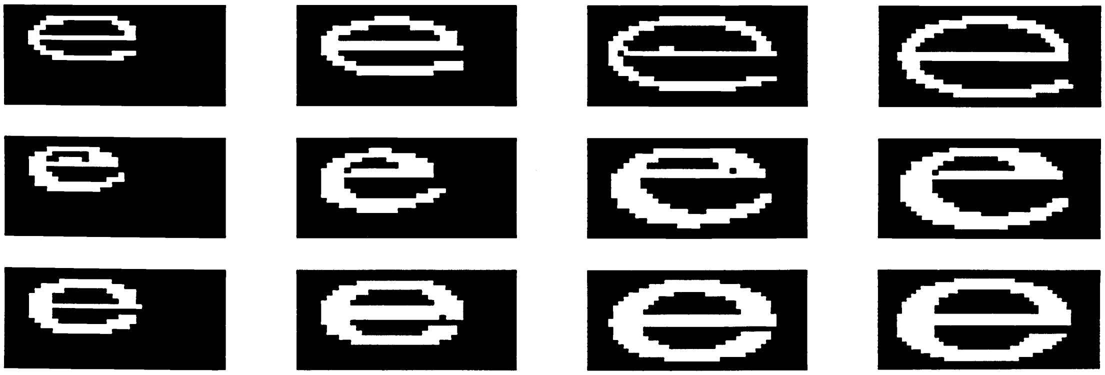

Figure 2: The letter 'e' in fonts Courier, Helvetica, Times (top to bottom) in point sizes 14, 18, 24, 28 (left to right).

The classification methodology was tested on 20 point faxed characters in each of the fonts. Hence, the recognition tool is being tested on a size of text different from any size used in training. In this character recognizer, Font is determined first. For each font, exemplars in the four training sizes are available for each of the 26 characters, a total of 104 training characters.

Every STIR row vector is generated by the steps of autocorrelation, scale transform, and reshaping to a vector. To illustrate, consider a 24 point Courier letter 'a' bitmap. Figure 3 shows its autocorrelation. The first and fourth quadrants are scale transformed using an interval distance $T = 1$ with row and column scale values of 0.1,0.4,0.7,1.0,1.3,1.6,1.9,2.2,2.5,2.8. Figure 4 shows the matrices of magnitudes of these scale transform coefficients, the STIR values. Note that the scale values are very similar for each quadrant. That is characteristic of the scale coefficient magnitudes for most functions. The magnitudes drop off roughly exponentially. These coefficient magnitudes reformed as an STIR row vector gives the appearance shown in Figure 5.

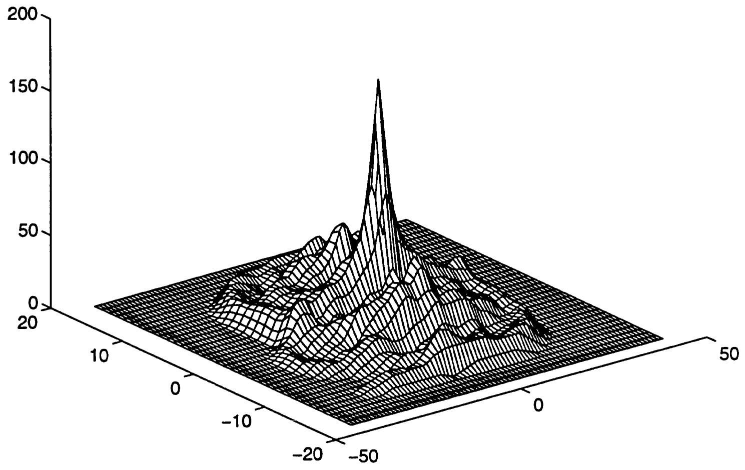

Figure 3: Autocorrelation function of 24 point Courier 'a'.

Font was classified using STIR training vectors from three matrices, one for each font. Considering each font as a single class implies an SVD on 104 STIR vectors. Since, for each font, we have four sizes each of the 26 letters. The length of each row is determined by the number of row and column scales chosen for calculation. The STIR row vectors each have 200 elements because, choice of row and column scales in the scale transform dictates a 10 by 10 matrix output for each autocorrelation quadrant. Thus, the SVD for each font will yield noise vectors corresponding to 96 singular values with zero magnitude. Calculating the sum of inner product magnitudes between these orthogonal vectors and a test STIR vector yields a selection value for each font. If the result is zero, then the unknown character must be represented in that font. In practice, one does not obtain a zero inner product with the correct font noise vectors. However, the correct font should correspond to the matrix generating the smallest selection value.

The 20 point test characters in each of the three fonts were processed into a STIR vectors and classified by smallest selection value. This worked perfectly. In all 78 test characters, the font was correctly classified.

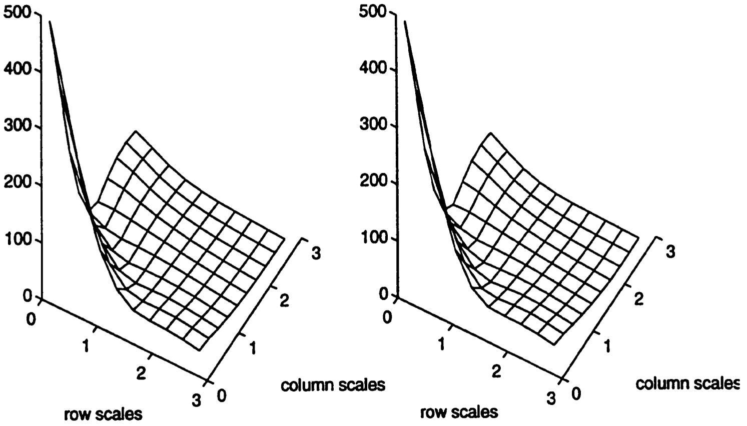

Figure 4: Magnitude of scale transform coefficients for Quadrants I and IV of the autocorrelation function of 24 point Courier 'a'.

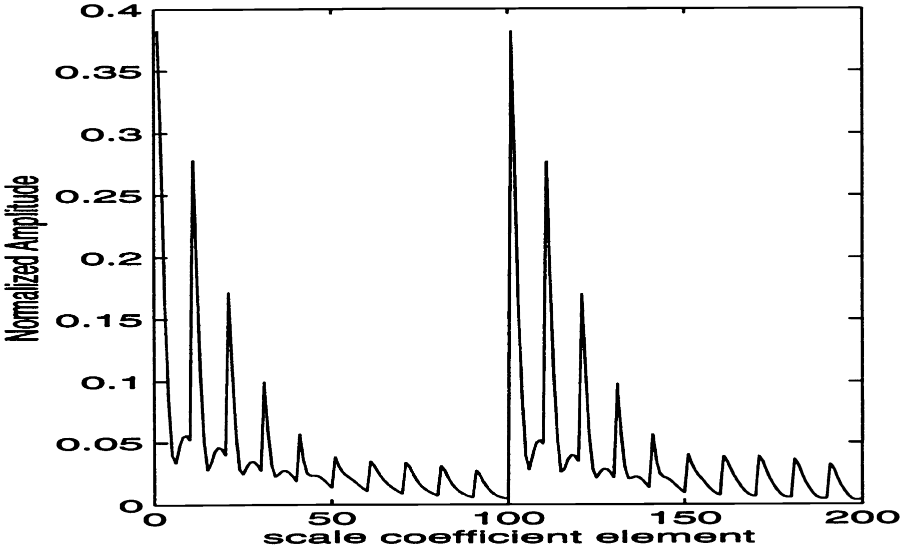

Figure 5: Scale and Translation Invariant Representation vector for a 24 point Courier 'a'.

4.1 Character determination

Once font is known, classification of the character follows the same method as classification of font. Twenty-six matrices are generated, one for each letter in the selected font. All scale transform parameters are the same as used in font classification. All STIR vectors of each training size are stacked to form a 4 by 200 matrix for each letter. SVDs were performed on each matrix. The right singular vectors corresponding to zero singular values were retained for selection value calculation. Each test input was processed into a STIR vector and its selection value was calculated for all 26 sets of noise vectors. The input was classified as the character corresponding to the set of noise vectors generating the smallest selection value.

Of the 78 test characters 3 were incorrectly classified. All Courier characters were correctly classified. In Times, the 'n' was classified as a 'u' and the 'p' was classified as a 'd'. In Helvetica, the 'd' was classified as a 'p'. Confusion between 'd'/'p' and 'u'/'n' is expected since the STIR approach does not discriminate between 180 degree rotated versions of letters. In a complete character recognition engine additional structural examination would be employed to discriminate between these characters. Figure 6 shows the bitmaps of the 20 point confused letters.

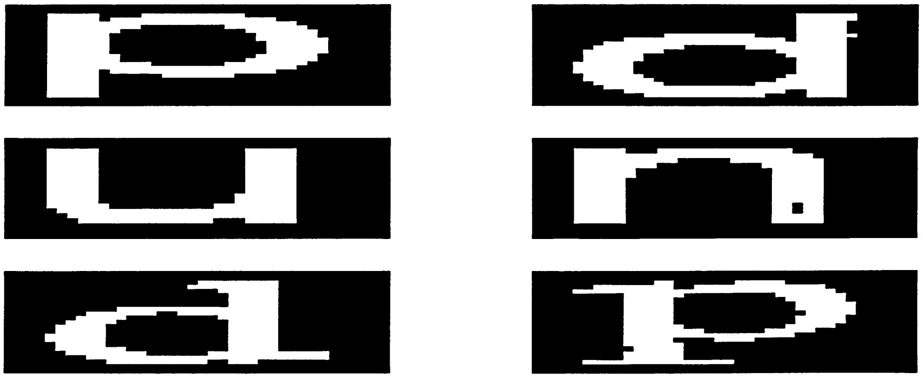

Figure 6: Test characters (20pt.) involved in misclassification. Helvetica **'p'** and **'d'** (top) Helvetica **'u'** and **'n'** (middle) Times **'d'** and **'p'**(bottom).

This example shows the potential for application of the STIR and noise subspace discrimination to character recognition. A selection value threshold could be added to reject symbols which are not among the valid set of characters. Simulations involving variations such as larger character sets, smaller font sizes, and added shot noise show that this method degrades gracefully: Errors in font classification are often still followed by correct character classification; more misclassification between similar characters occurs; classification errors increase proportionally with amount of speckle.

5 DISCUSSION

The approach we have described is general. It can be used to represent and classify patterns of any dimension. An extension of the method to space-spatial frequency representations is planned. Once in the STIR vector representation, the processing time required is identical for all sizes of inputs because the STIR vector length is determined by the number of row and column scale coefficients selected, not by the size of the bitmap input. For maximum processing speed, this number may be set to the smallest value which provides the required classification ability.

The character recognition example could easily have been a parts classification example, a word spotting example, etc. In fact, the same basic methodology can be extended to time-frequency representations (TFRs) of signals by treating the TFRs as images. Invariance would be realized for time shift, frequency shift and scale changes. The method is far from optimized, but it performs quite well, giving hope for considerable improvement in refined versions.

6 ACKNOWLEDGMENTS

This research has been supported in part by the Office of Naval Research, ONR contract no. N00014-90-J-1654, and the National Science Foundation, NSF grant no. BCS-9110571 and NSA contract no. MDA904-95-C-2157.

7 REFERENCES

[1] L. Cohen, "The Scale Representation," *IEEE Trans. Signal Processing*, Vol.41, pp. 3275-3292, 1993.

[2] G. Cristobal, J. Bescos, J. Santamaria and J. Montes, "Wigner distribution representation of digital images," *Pattern Recognition Letters,* Vol. 69, pp. 215-221, 1987.

[3] L. Jacobson and H. Wechsler, "A paradigm for invariant object recognition of brightness, optical flow and binocular disparity images," *Pattern Recognition Letters,* Vol. 1, pp. 61-68, 1982

[4] L. Jacobson and H. Wechsler, "A theory for invariant object recognition in the frontoparallel plane," *IEEE Trans. PAMI* , Vol. 6, No. 3, pp. 325-331, 1984.

[5] L. Jacobson and H. Wechsler, " Derivation of optical flow using a spatiotemporal-frequency approach," *Comput. Vision, Graphics and Image Processing,* Vol. 38, pp. 29-65, 1987.

[6] S. Mori, C. Y. Suen and K. Yamamoto, "Historical review of OCR research and development," *Proc. IEEE,* Vol. 80, pp. 1029-1092, 1982.

[7] V. F. Pisarenko, "The retrieval of harmonics from a covariance function," *Geophys. J. Royal Astron. Soc.,* Vol. 33, pp. 347-366, 1973.

[8] T. Reed and H. Wechsler, "Spatial/spatial-frequency representations for image segmentation and grouping," *Image and Vision Computing,* Vol. 9, No. 3, pp. 175-193, 1991.

[9] W. J. Williams and J. Jeong, "Reduced interference time-frequency distributions," *Time-Frequency Signal Analysis Methods and Applications,* ed. B. Boashash, pp. 74-98, Longman and Cheshire, 1992.

[10] E. J. Zalubas and W. J. Williams, "Discrete Scale Transform for Signal Analysis," Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing, Vol. 3, pp. 1557-1561, 1995.