

SHIFT AND SCALE INVARIANT DETECTION

Eugene J. Zalubas, Jeffrey C. O'Neill, William J. Williams, and Alfred O. Hero III

Department of Electrical Engineering and Computer Science
University of Michigan, Ann Arbor, MI 48109-2122
email: loz,jeffo,wjw,hero@eecs.umich.edu

ABSTRACT

Different signal realizations generated from a given source may not appear the same. Time shifts, frequency shifts, and scales are among the signal variations commonly encountered. Time-frequency distributions (TFDs) covariant to time and frequency shifts and scale changes reflect these variations in a predictable manner. Based on such TFDs, representations invariant to these signal distortions are possible. Presented here are two approaches for discriminating between signal classes where within class translation and scale variation occur. The first method uses an auto-correlation followed by a scale transform to achieve the invariances. The second method treats the TFD as a two-dimensional probability density function and applies a transformation that removes the mean and variance to provide the shift and scale invariance. Each method employs discrimination mechanisms to yield powerful results.

1. INTRODUCTION

Recognition of specific signatures in signals has long been of interest. Powerful techniques exist for their detection and classification, but these techniques are often defeated by variations in the signature. These variations include scaling and shifting in time and frequency. All time-frequency distributions (TFDs) in Cohen's class are covariant to time shifts and frequency shifts [1]. There exists a subset of Cohen's class that is also covariant to scale changes. Similar to [3], we will call this class the shift-scale covariant class. Reduced interference distributions (RIDs), introduced by Williams and Jeong [7], are a subset of the shift covariant class. RIDs have a straightforward design procedure and also attenuate troublesome cross terms. Methods can be designed that transform shift scale covariant

TFDs into representations that are invariant to shifts and scales.

Define:

$$(\mathbf{T}_\tau x)(t) = x(t - \tau)$$

$$(\mathbf{F}_\theta x)(t) = x(t) e^{j\theta t}$$

$$(\mathbf{D}_c x)(t) = \frac{1}{\sqrt{c}} x(ct)$$

Given a specific signal of interest, $x(t)$, we want to detect the class that includes signals of the form:

$$(\mathbf{T}_\tau \mathbf{F}_\theta \mathbf{D}_c x)(t)$$

for all possible values of τ , θ , and c . Since the above operators do not commute, we would also include the signals obtained by permuting the order of the operators. Using TFDs that are *covariant* to shifts and scales (e.g. RIDs), we will provide methods for creating a representation that is *invariant* to shifts and scales. Scaling of a signal causes the time and frequency axes to scale inversely. Time and frequency shifts in the signal change its location in the time-frequency plane, but do not distort it.

We propose two approaches for robust signature recognition. Both use RIDs for generating representations invariant to scales, time shifts, and frequency shifts. In addition, both detection schemes remove information regarding the location of the signal in the time-frequency plane and also the scale of the signal. Thus, the detections schemes use the actual "pattern" of the time-frequency distribution for discriminating the two classes of sounds. This scheme is most appropriate when the signal classes of interest have a consistent time-frequency structure, but can occur with varying shifts in time and frequency and also varying scales.

The first approach produces the Scale and Translation Invariant Representation (STIR) via autocorrelation and the scale transform. Classification is accomplished by comparison of projection measures of a test

This research has been supported in part by the Office of Naval Research, ONR contract nos. N00014-90-J-1654 and N00014-97-1-0072.

STIR onto subspaces orthogonal to the class STIRs. The second approach uses moments to provide invariant representations. Distance measures between test representations and class representations are used for classification. The next two sections describe the detection schemes in detail. Examples using recorded sperm whale sounds are presented to highlight the abilities of the techniques.

2. STIR - NOISE SUBSPACE CLASSIFIER

STIR classification has three steps. Autocorrelation of the 2D representations is used to remove translational effects. 2D scale transformations of the autocorrelation result is used to remove scaling effects. Classification is accomplished by projecting a test representation onto a subspace.

2.1. Computation of the 2D autocorrelation

It is well known that autocorrelation removes translational effects in images. The 2D autocorrelation may be carried out as follows:

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

where $a(n_1, n_2)$ is the image. The 0,0 lag point provides an origin from which the autocorrelation function scales. Of course for TFDs, the removal of translation effects may be accomplished by taking the magnitude in the ambiguity domain.

2.2. Direct scale transform

The 2D autocorrelation function of a RID provides invariance to time and frequency shift and a stable origin. A discrete scale transform implementation can additionally provide scale invariance. The scale transform, introduced by Cohen [2], is a specific case of the Mellin transform with the key property that for signals of equal energy the magnitude of their scale transforms are invariant to scaling effects. The one dimensional (1D) scale transform, $D(c)$, of the time domain signal $f(t)$ may be defined as

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

using the substitution $t = e^x$. Using a direct expansion of the scale transform, a discrete approximation is obtained. Assume the signal is sampled every T units and remains constant between samples. Splitting the

integral into logarithmic intervals yields

$$D(c) = \frac{1}{\sqrt{2\pi}} \left\{ \int_{-\infty}^{\ln T} f(e^x) e^{(\frac{1}{2}-jc)x} dx + \int_{\ln T}^{\ln 2T} f(e^x) e^{(\frac{1}{2}-jc)x} dx + \dots \right\}$$

$$D(c) \approx \frac{1}{\sqrt{2\pi}} \left\{ f(e^{-\infty}) \int_{-\infty}^{\ln T} e^{(\frac{1}{2}-jc)x} dx + f(e^{\ln T}) \int_{\ln T}^{\ln 2T} e^{(\frac{1}{2}-jc)x} dx + \dots \right\}$$

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

Since the scale transform is based on exponential sampling relative to the origin, the entire autocorrelation plane cannot be dealt with at once. By symmetry, the first and fourth quadrants combined provide complete information about the entire autocorrelation plane. Thus, 2D scale transform these quadrants separately. A 2D scale transform is implemented by sequentially scale transforming the rows then columns of a matrix. The magnitude of normalized scale transforms for first and fourth quadrants of the autocorrelation plane define the STIR image of the original 2D input.

2.3. Classification of patterns

Our technique for pattern classification uses STIR images decomposed into an orthonormal set of descriptors, using a concept borrowed from Pisarenko's harmonic decomposition [5, 8]. The Karhonen-Loève transform is a means of accomplishing this. The singular value decomposition (SVD) provides equivalent results. The STIRs of each exemplar in a class are shaped into a row vector by concatenating rows of the two STIR matrices. These row vectors are stacked to form a matrix representing the class. The SVD is then applied to extract essential features of the set of vectors. Provided that a sufficient number of scale coefficients are calculated, singular values of zero will result. Right singular vectors corresponding to zero singular value define a subspace orthogonal to the class of STIR vector representations.

In classifying a test signal, generate its STIR vector. Compute for each class the sum of inner product magnitudes of the STIR vector with the orthogonal subspace vectors. If the sum is zero, then the test signal

must be a member of the corresponding class. In practice, one does not obtain a zero sum with the proper subspace class, but the sum resulting from the proper class has the smallest magnitude relative to sums from calculation with other class subspaces.

3. MOMENT METHOD

For this approach, we treat the TFD as a two dimensional probability density function (pdf) where time and frequency represent a two-dimensional random variable [4]. The expected value of a function of these variables, $f(t, \omega)$, will be defined as:

$$E[f(t, \omega)] = \iint f(t, \omega) C(t, \omega) dt d\omega \quad (3)$$

As a first step, we calculate the expected value of the time variable, $\mu_t = E[t]$, the expected variable of the frequency variable, $\mu_\omega = E[\omega]$, and the spread of the time variable $\sigma_t^2 = E[(t - \mu_t)^2]$. If we normalize the TFD in the following manner:

$$\tilde{C}(t, \omega) = C((t - \mu_t)/\sigma_t, (\omega - \mu_\omega)/\sigma_\omega) \quad (4)$$

then we obtain a representation that is invariant to time shifts, frequency shifts, and scale changes.

Given the representation in equation 4, we need a method for reducing the dimensionality so we can apply it to the detection problem. We propose to do this by calculating normalized moments of the probability density function in equation 4.

$$m_{p,q} = E[t^p \omega^q] \quad (5)$$

Theoretically, all of the moments completely describe the pdf. Here we hope to retain as much information as possible by using a subset of the moments.

We will use the above to characterize classes of whale sounds. Suppose we have a set of N whale sounds that represent the A class. For each of the whale sounds, we will apply the following steps:

1. compute a shift-scale covariant TFD,
2. remove the location and scale information by applying the transformation in equation 4, and
3. calculate a subset of the moments defined in equation 5.

For each whale sound in the class, we will have a moment vector, m_i , for $i = 1 \dots N$. This moment vector will be invariant to shifts and scales of the signal. We will treat these N moment vectors as instances of a jointly Gaussian random vector, $\underline{M}$. Since the moments are computed as an average, it is reasonable to

assume that the random vector, $\underline{M}$, will be approximately jointly Gaussian. From the set of N moment vectors we will estimate the mean vector, μ_A , and the covariance matrix, K_A , using standard sample estimators.

Given two different classes of whale sounds, A and B, that we wish to discriminate, we will estimate μ_A , K_A , μ_B , and K_B . Given an unknown signal, we will apply the steps above to compute the moment vector, m . Under the above assumptions, the following test

$$(m - \mu_A)^t K_A^{-1} (m - \mu_A) - (m - \mu_B)^t K_B^{-1} (m - \mu_B) \stackrel{A}{\leq} \gamma$$

chooses the most "likely" answer [6].

4. DISCUSSION

The detection schemes presented above have been applied to the recognition of sperm whale signatures¹. Recordings of sperm whale "clicks" from two different animals are used as data. The goal is to correctly associate each recorded signature to the animal which produced it, whale A or whale B.

We only had six recordings from each of two whales. To increase the number of records and to make correct classification more challenging, mathematical variation was added. Time shifts, frequency shifts, scalings, and noise were applied to the signals. As a result, the final signal data set consisted of 81 variations of each of six recordings for each whale. These signals were divided into test and training sets, each containing 81 variations of three recordings. Figure 1 shows example signals from the two whales after noise and variation have been added to the recordings. Obviously, with the added noise, discrimination between the signals is difficult in either the time or frequency domain.

In order to achieve the frequency shift invariance using TFDs, the analytic signals of the real valued data sequences were used. Calculating the binomial TFD [7] and applying the two detection mechanisms yields the receiver operating characteristics (ROCs) shown in Figure 2. It appears that the STIR method could provide better performance as a detector over the moment method. However, the moment method has a lower computational complexity.

5. REFERENCES

- [1] L. Cohen, *Time-Frequency Analysis*. Englewood Cliffs, NJ: Prentice-Hall, 1995.

¹Provided by W. Watkins at Woods Hole Oceanographic Institution

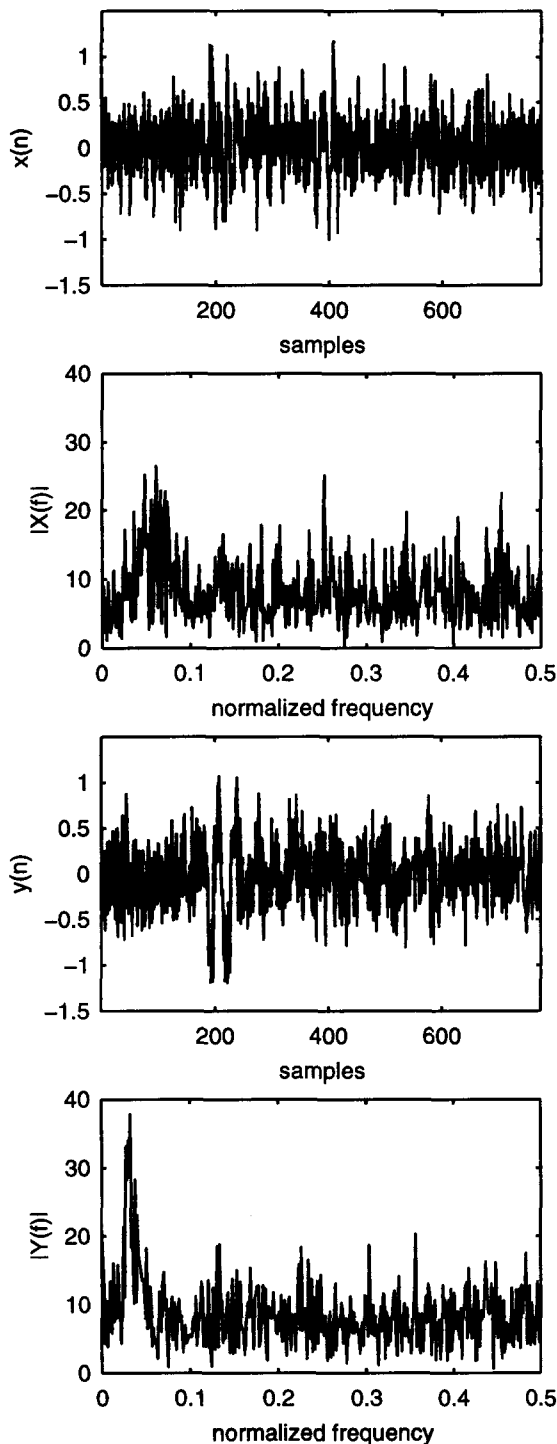


Figure 1: Examples of the whale signals with added noise and shift and scale distortions. Signal for whale A, $x(t)$, and its Fourier transform magnitude appear in the top and upper middle graphs. Signal for whale B, $y(t)$, and its Fourier transform magnitude appear in the lower middle and bottom graphs.

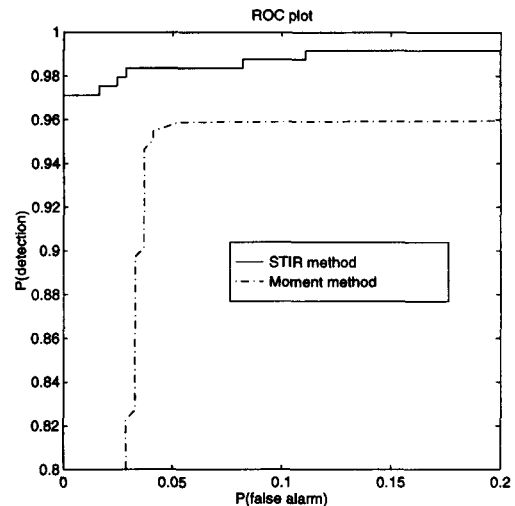


Figure 2: Receiver Operating Characteristics for the whale sounds data using the two different methods.

- [2] L. Cohen, "The Scale Representation," *IEEE Trans. on Signal Processing*, Vol.41, pp. 3275-3292, 1993.
- [3] F. Hlawatsch and R. L. Urbanke, "Bilinear time-frequency representations of signals: the shift-scale invariant class," *IEEE Trans. on Signal Processing*, vol. 42, no. 2, 1994.
- [4] J. C. O'Neill, A. O. Hero III and W. J. Williams, "Word spotting via spatial point processes," *Proc. of the IEEE Int. Conf. on Image Processing*, vol II, pp 217-220, 1996.
- [5] V. F. Pisarenko, "The retrieval of harmonics from a covariance function," *Geophys. J. Royal Astron. Soc.*, Vol. 33, pp. 347-366, 1973.
- [6] H. L. Van Trees, *Detection, Estimation, and Modulation Theory*. New York: John Wiley and Sons, 1968.
- [7] W. J. Williams and J. Jeong, "Reduced interference time-frequency distributions," *Time-Frequency Signal Analysis: Methods and Applications*, chapter 3, pp. 74-98, Longman and Cheshire, 1992.
- [8] W. J. Williams, E. J. Zalubas and A. O. Hero III, "Separating desired image and signal invariant components from extraneous variations," *Advanced Signal Processing Algorithms, Architectures, and Implementations VI*, SPIE vol. 2846, pp. 262-272, 1996.