A Proof of Hirschman Uncertainty Invariance to the Order of Rényi Entropy for Picket Fence Signals, and Its Relevance in a Simplistic Recognition Experiment

Kirandeep Ghuman and Victor DeBrunner
Department of Electrical and Computer Engineering
Florida State University
Tallahassee, FL, 32310
{kg08h,victor.debrunner}@fsu.edu

***Abstract*—In [1] we developed a new uncertainty measure which incorporates Rényi entropy instead of Shannon entropy. This new uncertainty measure was conjectured to be invariant to the Rényi order $\alpha > 0$ for the case of the optimizer signals of Hirschman Uncertainty (Picket Fence functions whose lengths are a perfect square). In this paper, we prove this invariance, and test whether this invariance is predictive in the problem of a simple texture classification for digital images. In the preliminary results, we find that it certainly influences the recognizer performance. Specifically, we find that the recognition performance does not depend significantly on the Rényi parameter α. We hope that these results will be extended to other problems where Rényi entropy is used.**

***Index Terms*—Uncertainty, Entropy, Textural features, Classification**

I. Introduction

Classically, the Heisenberg uncertainty relation, $\triangle\, x \triangle\, p \geqq \frac{h}{4\pi}$ [2], is based on the standard deviations $\triangle\, x$ and $\triangle\, p$ that define the position x and moment p of a electron. This relationship forms the backbone of modern physics, but in and of itself is not useful when applied to probability distributions, and more specifically to information theory and signal processing. Here, approaches like Shannon, Rényi and Tsallis entropy measures are used. Shannon introduced the notion that the amount of uncertainty found in a given probability distribution $P = \{p_1, p_2, p_3, \ldots, p_n\}$ is the amount of entropy found in the distribution P. However, while the information measures can be viewed as measures of uncertainty because uncertainty is associated with missing information, their direct use is not satisfactory because then we will ignore the spectral information. The Hirschman Uncertainty [3] is defined by the average of the Shannon entropies of a discrete-time signal and its Fourier transform. The picket fence function defines the optimal bases of the Hirschman uncertainty in [4]. To explore the fundamental nature of Hirschman Uncertainty, we extended Hirschman Uncertainty to include Rényi entropy [1] and discrete fractional fourier transform [5].

II. Hirschman Uncertainty Principle

The discrete Hirschman Uncertainty measure U_{p} [6] conveys the compactness of a discrete-time signal in the sample-frequency (phase) plane. It is well-known that for the continuous time, the only signals for which equality in the Heisenberg Uncertainty holds are obtained from the Gaussian $2\pi^{\frac{1}{2}}\, e^{\frac{-t^2}{4}} \quad (t\epsilon\mathbb{R})$, by applying translations, dilations, or modulations or multiplication by a complex number of magnitude 1. It is less well known that Heisenberg Uncertainty is ill-posed for digital signals. The uncertainty measure obtained by a naïve discretization of the continuous case Heisenberg measure is, among many problems, not translation invariant (a more complete list of problems can be found in [4]). The more general Hirschman Uncertainty measure uses deterministic entropies with respect to densities in time and frequency.

Formally, let $\mathcal{D} = \{0, 1, 2, \ldots, N-1\}$ and let $\mathcal{H}_N$ denote the Hilbert space of squences $x : \mathcal{D} \rightarrow \mathbb{C}$ with squared norm

$$\|x\|_2^2 = \sum_{n=0}^{N-1} |x\,[n]|^2$$

We also have the (discrete) Fourier transform

$$X\,[k] = Fx\,[n] = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} x\,[n]\, W_N^{nk} \tag{1}$$

where $W_N = e^{-j\frac{2\pi}{N}}$ is the standard twiddle factor and F is the defining Fourier matrix. This defines and isometry on $\mathcal{H}_N$ with inverse given by

$$x\,[n] = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} X\,[k]\, W_N^{-nk} \tag{2}$$

By the digital phase plane we mean the set of all points $(n, k) \in \mathcal{D} \times \mathcal{D}$.

Definition 1. Let $x \in \mathcal{H}_N$ with $\|x\|_2 = 1$. The Shannon Entropy is

$$H(x) = -\sum_{n=0}^{N-1} \mid x\,[n] \mid^2 ln(\mid x\,[n] \mid ^2) \tag{3}$$

A general Uncertainty measure is created by considering the weighted average

$$U_p(x) = pH(x) + (1-p)H(Fx),\ p \in [0,1]$$

The parameter p allows us to trade-off concentration in time and frequency. In the extreme cases, the Uncertainty is either ignoring the frequency localization (when $p = 1$) or the time localization (when $p = 0$). Strictly speaking, we have

Definition 2. Let $x \in \mathcal{H}_N$ with $\|x\|_2 = 1$. The Hirschman Uncertainty is

$$U_{\frac{1}{2}}(x) = \frac{1}{2}H(x) + \frac{1}{2}H(Fx) \tag{4}$$

and so the localization in both time and frequency is considered of equal interest. Before describing the minimizers of (4), we define periodization:

For $N = KL$, the periodization of $v \in \mathbb{C}^k$ is defined as $x[sK+n] = (1/\sqrt{L})v[n]$ for $0 \le s \le L-1$ and $0 \le n \le K-1$. We refer to the sequence $v \in \mathbb{C}^k$ given by $v[0] = 1, v[1] = 0, \cdots v[K-1] = 0$, as the Kronecker delta or impulse (unit sample) sequence, without specifying the signal length K. We proved the following theorem in [7]:

Theorem 3. *The only sequences $x \in \mathbb{C}^k$, with $\|x\|_2 = 1$, for which $U_{\frac{1}{2}}(x)$ is minimal are obtained from the Kronecker delta sequence by applying any composition of periodization, translation, modulation, the DFT, and multiplication by a complex number of unit magnitude.*

III. HIRSCHMAN-RÉNYI UNCERTAINTY

In [1] we extended the Hirschman Uncertainty to include Rényi entropy. To develop this Hirschman-Rényi Uncertainty, we first recall the Rényi entropy.

A. Rényi entropy

Rényi, or alpha-order, entropy (denoted as α-order), is

Definition 4. Let $u \in \mathcal{H}_N$ with $\|u\|_2 = 1$. The Rényi Entropy is

$$H^\alpha(u) = \frac{1}{1-\alpha}\log\sum_{n=0}^{N-1}\left(|u[n]|^2\right)^\alpha,\ \alpha \in [0,\infty) \tag{5}$$

In the special case where $\alpha = 1$, the Rényi entropy is identically the Shannon entropy.

$$\lim_{\alpha\to1} H^\alpha(u) = H^1(u) = H(u) = -\sum_{n=0}^{N} \mid u[n] \mid^2 \ln(\mid u[n] \mid)^2$$

$$\lim_{\alpha\to0} H^\alpha(u) = H^0(u) = \log N = H^{max}$$

$$\lim_{\alpha\to\infty} H^\alpha(u) = H^\infty(u) = -\log[max_i(\mid u(n) \mid^2)]$$

For any $\alpha \geqslant 0$, the Rényi entropy $H^\alpha(u)$ is non-negative and decreasing function of α (i.e. for $\alpha_1 < \alpha_2$, $H^{\alpha_2}(u) \leqslant H^{\alpha_1}(u)$ for all u) decaying from $H^0(u)$ to H^∞.

B. New Uncertainty Measure

Combining the Rényi entropy with the Hirschman Uncertainty, the following definition seems reasonable:

Definition 5. Let $u \in \mathcal{H}_N$ with $\|u\|_2 = 1$. The Hirschman-Rényi Uncertainty is

$$U^\alpha_{\frac{1}{2}}(u) = \frac{1}{2}H^\alpha(u) + \frac{1}{2}H^\alpha(Fu) \tag{6}$$

The picket fence signal is a minimizer of the Hirschman Uncertainty [8]. This signal and its Fourier transform as computed in Eq. (1) have the same appearance, and so their Shannon entropy given in Eq. (3) and their Rényi entropy given in Eq. (5) are also the same.

Theorem 6. *The Hirschman-Rényi Uncertainty of Eq. (6) is invariant to the Rényi order $\alpha > 0$ for a picket fence signal whose length is a perfect square.*

$$U^\alpha_{\frac{1}{2}}(u) = U^1_{\frac{1}{2}}(u) = U^\infty_{\frac{1}{2}}(u) = \log K = constant$$

Proof: For the picket fence, the signal and its Fourier transform are the same i.e. $u = Fu$. Therefore, $H^\alpha(u) = H^\alpha(Fu)$. We must ensure that only the non-zero values (probabilities, so to speak) are included. Considering the picket fence signal u has sample length $N = K^2$ and has K non-zero samples, ∎

$$\lim_{\alpha\to1} H^\alpha(u) = H^1(u) = H(u) = -\log\frac{1}{K} = \log K$$

$$\begin{aligned}\lim_{\alpha\to\infty} H^\alpha(u) &= H^\infty(u) = -\log[max_i(\mid u(n) \mid^2)] \\ &= -\log\frac{1}{K} = \log K\end{aligned}$$

Therefore,

$$U^1_{\frac{1}{2}}(u) = \frac{1}{2}\log K + \frac{1}{2}\log K = \log K$$

$$U^\infty_{\frac{1}{2}}(u) = \frac{1}{2}\log K + \frac{1}{2}\log K = \log K$$

Fig. 1 displays the invariance of the Hirschman-Rényi Uncertainty to Rényi order $\alpha > 0$ for a picket fence signal with length $N = 25$. The Hirschman Uncertainty is shown as the star at $\alpha = 1$ (where Rényi and Shannon entropies are identical). At $\alpha = 0$, since Rényi entropy $H^\alpha(u)$ is given by $\log N$ (where N represents the length of the signal), the Hirschman-Rényi uncertainty $U^\alpha_{\frac{1}{2}}$ obtained from eq. (6) is also reduced to $\log N$ (following convention $0^0 = 1$ while caluclating). Therefore, we can summarize two main points in case of picket fence signal:

1) At $\alpha = 0$, Hirschman-Rényi uncertainty $U^\alpha_{\frac{1}{2}} = \log N$.
2) For $\alpha > 0$, $U^\alpha_{\frac{1}{2}}$ is constant for all N.

Consider the case of the $N = 64$ point Identity (Kronecker delta) function shown in Fig.2 (repeated from [7] for easy reference). The Identity function is the solid curve, slanting down from left to right, with a Concentration of about 4.1 nits for extreme time localization and 0 nits for extreme

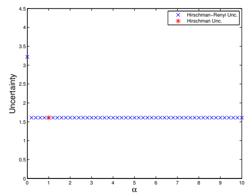


Figure 1. Hirschman Uncertainty and Hirschman-Rényi Uncertainty for the picket fence signal N=25

frequency localization. The concentration parameter p provides the analyst a mechanism to examine a signal's time or frequency characteristics as appropriate. Again, the Hirschman uncertainty is specifically for $p = \frac{1}{2}$. The optimal signal (in this case, a 64-point picket fence signal) is shown as the constant 2.07 nit dashed line.

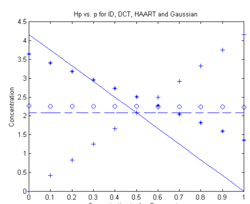


Figure 2. Concentration vs. p for various signals, $N = 64$ (identity(-), gaussian (o), Haar (*), cosine (+), picket fence(–))[7]

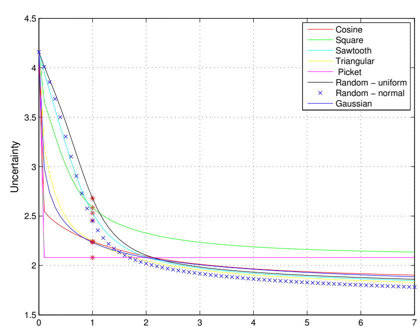


Figure 3. Hirschman Uncertainty & Hirschman-Rényi Uncertainty for discrete signals N = 64

Fig. 3 shows the Hirschman Uncertainty & Hirschman-Rényi Uncertainty for various discrete signals with N = 64. It is noted that at $\alpha = 0$, for all discrete signals Hirschman-Rényi Uncertainty is equal to $\log N$, i.e. it only depends on the length of the signal. The Hirschman-Rényi Uncertainty remains constant for the picket fence signal and for all other signals it decays from $\alpha = 0$ to $\alpha = 7$. For $\alpha > 1.5$, there is very little change in the uncertainty of signals whereas from $\alpha = 0 - 1.5$, the change in uncertainty are noticeable with the change in α values. Also, the picket fence has the minimum uncertainty for $\alpha < 1.7$.

Theorem 7. *Let $u \in \mathcal{H}_N$ with $\|u\|_2 = 1$. For $\alpha \epsilon (0, \infty)$, Hirschman-Rényi Uncertainty for discrete signals (other than picket fence signal) has the following limiting equation*

$$\log N \geq U^{\alpha}_{\frac{1}{2}}(u)$$

and

$$U^{\alpha}_{\frac{1}{2}}(u) \quad > \quad \frac{1}{2}[(-\log(max_i\left(\mid u(n) \mid^2\right)) + (-\log(max_i\left(\mid \hat{u}(n) \mid^2\right))]$$

where $\hat{u} = Fu$, fourier transform of the signal u.

Proof: Considering any discrete signal $u \in \mathcal{H}_N$ with $\|u\|_2 = 1$, $H^0(u) = \log N = H^{max}$ and $H^{\infty}(u) = -\log[max_i(\mid u(n) \mid^2)]$. Rényi entropy $H^{\alpha}(u)$ is a decreasing function of α, decaying from $H^0(u)$ to H^{∞}. Subsituting the values of $H^0(u)$ and H^{∞} in eq. 6, we get the upper and lower bounds of $U^{\alpha}_{\frac{1}{2}}(u)$ for the discrete signals. ∎

From fig. 2 and fig. 3 it is evident that the picket fence signals are invariant to both concentration trade-off parameter p and the Rényi order α.

IV. Classification Based on Haralick Textural Features

Image texture can be defined as a function of the spatial variation in pixel intensities [9]. Texture classification involves two major steps: 1) feature extraction and 2) classification. In this paper, we calculate 12 textural features (energy, entropy, contrast, correlation, variance, sum average, sum variance, sum entropy, difference variance, difference entropy, information measures of correlation) proposed by Haralick [10] for image classification. This is a classic texture recognition paper. We have conducted the preliminary experiment using five textures (D9, D12, D19, D68, D112) shown in figure 4 from the Brodatz texture database. Haralick suggested the use of the Gray Level Co-occurence Matrix (GLCM) for extracting textural features. GLCM is a MxM square matrix where M is the number of gray levels in the image. Element (i,j) of the GLCM matrix is the number of occurences of the pair of gray levels i and j which are d distance apart. Four GLCM matrices are calculated in 4 different directions of adjacency (0^0, 45^0, 90^0, 135^0). The Forward Subset Selection (FSS) is used to reduce the dimentionality of the feature set and a subset of the 12 textural features is used for the final classification process. The subset of the features obtained using FSS include: correlation, energy, entropy, sum entropy and information measures of correlation.

Algorithm

Step1: Create 100 32x32 subimages from the original 512x512 images from the Brodatz database. All the images are histogram equalized for contrast adjustment. Then, calculate the GLCMs in 4 different directions of adjacency (0^0, 45^0, 90^0, 135^0).

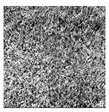
Step2: Feature extraction: Obtain a set of 4 values for each of the 12 textural features calculated from 4 GCLMs in different directions. The mean of these 12 features, averaged over 4 values comprise of the feature set.

Step3: Repeat the steps 1 and 2 for each texture class.

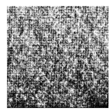
Step4: Feature Selection: The feature dimentionality is reduced using forward subset selection (FSS) and a subset of features is obtained. This subset of features is used for the classification.

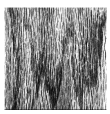
Step5: Nearest Neigbor Classifier: A 500x6 feature dataset is used to train the classifier while a 500x6 test dataset is then input to the trained classifier for the classification purpose.

Step6: The whole classification process (steps 1-5) is repeated 5 times on different sets of data from each texture class and then averaged to get the classification rate for that class.


(a) D9 Grass


(b) D12 Bark


(c) D19 Woolen cloth


(d) D68 Wood grain

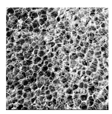

(e) D112 Plastic bubbles

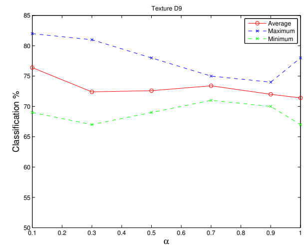
Figure 4. Five textures used in the texture classification experiment

The original images in the Brodatz database are 512x512 pixels in size. 100 32x32 images were obtained from these 512x512 original images in each of the 5 texture classes. In each texture class, 100 images are used in training, while an additional 100 are used as test images. The final training set consist of 500 subimages and the test set consists of 500 subimages. The nearest neighbor classifier is used for the classfication using the Euclidean distance measure. A 500x6 feature dataset is used for training the classifier and then another 500x6 test dataset is input to the trained classifier for final classification. The classification results are shown in fig. 5. First the classification experiment is conducted with Shannon entropy (Rényi entropy with $\alpha = 1$) as one of the features in the feature set. Then, Rényi entropy with different values of α replace the Shannon entropy as one of the 6 features in the feature subset. The classification is carried out on 5 different data sets in each texture class and the average of 5 classifications is shown in fig. 5(f).

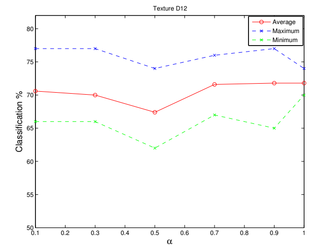
The classification varies by 1% - 5% in each texture class with variation in α values from 0.1 to 1 whereas the classification rate drops for the larger values of α in most of the texture classes. The classification rates with Rényi entropy as one of the features with $\alpha \leq 1$ are comparable or better than Shannon entropy as one of the features in the feature set. Table I shows the average classification rate carried on 5 different dataset in each texture class. If we eliminate entropy from the feature set, we find the average classfication rate drops from 77.08% to 76.84%. The drop in the classification rate is only 0.24%.

V. Conclusions

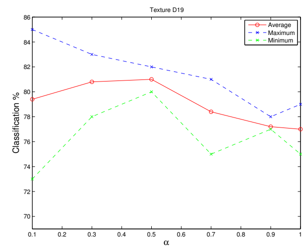
The Hirschman-Rényi Uncertainty is defined as a generalization of the previously defined Hirschman Uncertainty. The Uncertainty of the minimizers, the picket fence signal, is not only invariant to concentration trade-off parameter p, but also to the Rényi order α. We perform a classical texture recognition experiement, and find that the recognition performance follows directly as our developed Hirschman-Rényi Uncertainty theory suggests. Specifically, we find that the Rényi entropy parameter α in the commonly used range barely, if at all, impacts classification performance; and using large values of α only decreases classification performance.



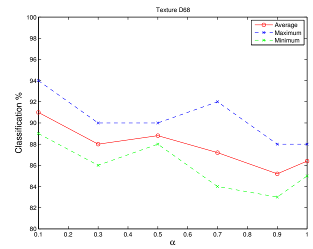
(a)



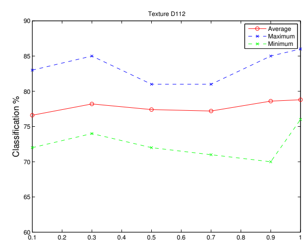
(b)



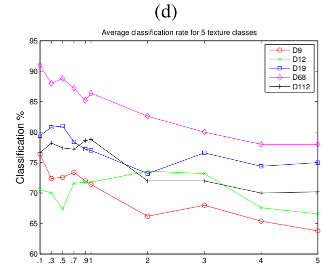
(c)



(d)



(e)



(f)

Figure 5. Classification rate using Nearest Neighbor Classifier with euclidean distance measures for different texture classes (a) Texture D9 (b) Texture D12 (c) Texture D19 (d) Texture D68 (e) Texture D112 (f) Average classification rate for 5 texture classes

Table I
Results of classification on 5 classes of texture

Texture	Classification% with entropy	Classification% without entropy
D9	71.4	68
D12	71.8	73.4
D19	77	77.2
D68	86.4	90.8
D112	78.8	74.8
Avg. Classification	77.08%	76.84%

REFERENCES

[1] K. Ghuman and V. DeBrunner, "Hirschman uncertainty using rényi, instead of shannon, entropy is invariant to the rényi entropy order," in *Asilomar Conference on Signals, Systems, and Computers*, November 2012.

[2] W. Heisenberg, "Translation of Über den anschaulichen inhalt der quantentheoretischen Kinematik und Mechanik (The actual content of quantum theoretical kinematics and mechanics, NASA TM-77379 Dec 1983)," *Zeitschrift für Physik*, vol. 43, 1927.

[3] I. I. Hirschman, "A note on entropy," *Amer. J. Math.*, vol. 79, pp. 152–156, 1957.

[4] V. E. DeBrunner, J. P. Havlicek, T. Przebinda, and M. Özaydın, "Entropy-based uncertainty measures for $\ell^2(\mathbb{R}^n)$, $\ell^2(\mathbb{Z})$, and $\ell^2(\mathbb{Z}/N\mathbb{Z})$ with a hirschman optimal transform for $\ell^2(\mathbb{Z}/N\mathbb{Z})$," *IEEE Trans. Signal Process.*, vol. 53, pp. 2690–2699, Aug. 2005.

[5] K. Ghuman and V. DeBrunner, "Hirschman uncertainty using discrete fractional fourier transform," in *Asilomar Conference on Signals, Systems, and Computers*, November 2013.

[6] V. DeBrunner, M. Özaydın, and T. Przebinda, "Resolution in time-frequency," *IEEE Trans. on Sig. Process.*, vol. 47, No. 3, pp. 783–789, Mar. 1999.

[7] T. Przebinda, V. DeBrunner, and M. Özaydın, "Using a new uncertainty measure to determine optimal bases for signal representations," *ICASSP'99*, vol. 3, pp. 1365–1368, Mar. 1999.

[8] T. Przebinda, V. DeBrunner, and M. Özaydın, "The optimal transform for the discrete Hirschman uncertainty principle," *IEEE Trans. Inform. Theory*, vol. 47, No. 5, pp. 2086–2090, Jul 2001.

[9] M. Tuceryan and A. K. Jain, "Texture analysis," *The handbook of patt. recog and vis.,2nd ed.*, pp. 207–248, 1998.

[10] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Trans. on Sys., Man. and Cyber.*, vol. 3, No. 6, pp. 610–621, Nov. 1973.