Nonextensive information-theoretic measure for image edge detection

A. Ben Hamza
Concordia University
Concordia Institute for Information Systems Engineering
1455 de Maisonneuve Boulevard West
CB-410-8 Montréal, Québec, Canada
E-mail: hamza@ciise.concordia.ca

Abstract. *We propose a nonextensive information-theoretic measure called Jensen-Tsallis divergence, which may be defined between any arbitrary number of probability distributions, and we analyze its main theoretical properties. Using the theory of majorization, we also derive its upper bounds performance. To gain further insight into the robustness and the application of the Jensen-Tsallis divergence measure in imaging, we provide some numerical experiments to show the power of this entopic measure in image edge detection. © 2006 SPIE and IS&T.* [DOI: 10.1117/1.2177638]

1 Introduction

Information-theoretic divergence measures[1] have been successfully applied in many areas including but not limited to statistical pattern recognition, neural networks, signal/image processing, speech processing, graph theory, computer vision, and optoelectronic systems.[2] Kulback-Liebler (or directed) divergence,[3] one of Shannon's entropy-based measures, has had success in many applications including image retrieval and indexing,[4] and performance analysis of image stochastic models.[5] A generalization of the directed divergence is the so-called α-divergence,[6] which is defined in terms of Rényi entropy.[7] The α-divergence measure has been applied in image registration and alignment as well as a variety of other problems.[6] Another entropy-based measure is called the Jensen-Shannon divergence.[8,9] This dissimilarity measure may be defined between any number of probability distributions, and due to this generalization, it may be used as a coherence measure between any number of distributions.

Inspired by the successful application of the mutual information measure,[10,11] and looking to address its limitations in often difficult imagery, we recently proposed an information-theoretic approach to ISAR image registration.[12] The objective of the proposed technique was to estimate the target motion during the imaging time, and was accomplished using a generalized Rényi's entropy-based similarity measure called the Jensen-Rényi divergence.[12] This divergence in fact measures the statistical dependence between an arbitrary number of consecutive ISAR image frames, which would be maximal if the images were geometrically aligned. In contrast to using mutual information,[10,11] one is able through the Jensen-Rényi divergence to adjust the weights, and also the exponential order of Rényi entropy to control the measurement sensitivity of the joint histogram, that is, the relative contributions of the histograms together with their order.[12] This flexibility ultimately results in better registration accuracy. The most fundamental and appealing characteristics of this divergence measure are its convexity and symmetry. In addition to its generality in involving an arbitrary number of probability distributions with possibly different weights, the Jensen-Rényi divergence enjoys appealing mathematical properties such as convexity and symmetry, affording great flexibility in a number of applications. Recently, it has been applied successfully to medical image segmentation[13] and detection of borders between coding and noncoding DNA regions.[14]

In this work, we propose a nonextensive information-theoretic divergence called Jensen-Tsallis divergence. Then, we investigate some of the main theoretical properties of the proposed measure as well as their implications. In particular, we derive its upper bounds, which are very useful for normalization purposes. Using the theory of majorization,[15] we also derive the maximum value of this divergence measure.

The remainder of this work is organized as follows. In the next section, we briefly recall some facts about Tsallis entropy prior to introducing the Jensen-Tsallis divergence. Section 3 is devoted to the theoretical properties of this divergence measure. Subject to some conditions, its convexity is subsequently established. By exploiting these properties and using the theory of majorization, we derive in Sec. 4 the maximum value of the Jensen-Tsallis divergence. Then, we derive the maximum value of the Jensen-Tsallis divergence for a weighted distribution. And finally, in Sec. 5, some experimental results are provided to show the much improved performance of the Jensen-Tsallis divergence in image edge detection.

2 Jensen-Tsallis Divergence Measure

Let $X=\{x_1,x_2,\ldots,x_k\}$ be a finite set with a probability distribution $\boldsymbol{p}=(p_1,p_2,\ldots,p_k)$, where $k>1$. Shannon's entropy is defined as $H(\boldsymbol{p})=-\Sigma_{j=1}^{k}p_j\log(p_j)$, and it is a measure of uncertainty, dispersion, information, and randomness. The maximum uncertainty or equivalently

Paper 05048R received Mar. 27, 2005; revised manuscript received Jul. 25, 2005; accepted for publication Aug. 18, 2005; published online Feb. 24, 2006.
1017-9909/2006/15(1)/013011/8/$22.00 © 2006 SPIE and IS&T.

minimum information is achieved by the uniform distribution. Hence, we can think of the entropy as a measure of uniformity of a probability distribution. Consequently, when uncertainty is higher, it becomes more difficult to predict the outcome of a draw from a probability distribution. A generalization of Shannon entropy is Rényi entropy,[7] given by

$$R_\alpha(\boldsymbol{p}) = \frac{1}{1-\alpha}\log\sum_{j=1}^{k} p_j^\alpha, \quad \alpha \in (0,1)\cup(1,\infty). \tag{1}$$

Another important generalization of Shannon entropy is Tsallis entropy,[16–18] given by

$$H_\alpha(\boldsymbol{p}) = \frac{1}{1-\alpha}\left(\sum_{j=1}^{k} p_j^\alpha - 1\right) = -\sum_{j=1}^{k} p_j^\alpha \log_\alpha(p_j), \tag{2}$$

where $\log_\alpha$ is the α-logarithm function defined as $\log_\alpha(x) = (1-\alpha)^{-1}(x^{1-\alpha}-1)$ for $x>0$. This generalized entropy was first introduced by Havrda and Charvát in Ref. 16, who were primarily interested in providing another measure of entropy. Tsallis, however, appears to have been principally responsible for investigating and popularizing the widespread physics applications of this entropy, which is referred to as Tsallis entropy.[17] Recently, in statistical physics there has been a concerted research effort to explore the properties of Tsallis entropy, leading to a statistical mechanics that satisfies many of the properties of the standard theory.[17] It is worth noting that for $\alpha \in (0,1]$, Rényi and Tsallis entropies are both concave functions; and for $\alpha>1$, Tsallis entropy is also concave, but Rényi entropy is neither concave nor convex. Furthermore, both entropies tend to Shannon entropy $H(\boldsymbol{p})$ as $\alpha \to 1$, and are related by

$$H_\alpha(\boldsymbol{p}) = \frac{1}{1-\alpha}\{\exp[(1-\alpha)R_\alpha(\boldsymbol{p})] - 1\}.$$

For $x,y>0$, the α-logarithm function satisfies the following property

$$\log_\alpha(xy) = \log_\alpha(x) + \log_\alpha(y) + (\alpha-1)\log_\alpha(x)\log_\alpha(y). \tag{3}$$

If we consider that a physical system can be decomposed in two statistical independent subsystems with probability distributions $\boldsymbol{p}$ and $\boldsymbol{q}$, then using Eq. (3) it can be shown that the joint Tsallis entropy is pseudo-additive

$$H_\alpha(\boldsymbol{p},\boldsymbol{q}) = H_\alpha(\boldsymbol{p}) + H_\alpha(\boldsymbol{q}) + (1-\alpha)H_\alpha(\boldsymbol{p})H_\alpha(\boldsymbol{q}),$$

whereas the joint Shannon and Rényi entropies satisfy the additivity property: $H(\boldsymbol{p},\boldsymbol{q}) = H(\boldsymbol{p}) + H(\boldsymbol{q})$, and $R_\alpha(\boldsymbol{p},\boldsymbol{q}) = R_\alpha(\boldsymbol{p}) + R_\alpha(\boldsymbol{q})$.

The pseudo-additivity property implies that Tsallis entropy has a nonextensive property for statistical independent systems, whereas Shannon and Rényi entropies have the extensive property (i.e., additivity). Furthermore, standard thermodynamics are extensive because of the short-range nature of the interaction between subsystems of a composite system. In other words, when a system is composed of two statistically independent subsystems, then the Boltzman-Gibbs entropy of the composite system is just the sum of entropies of the individual systems, and hence the correlations between the subsystems are not accounted for. Tsallis entropy, however, does take into account these correlations due to its pseudo-additivity property. Furthermore, many objects in nature interact through long-range interactions such as gravitational or unscreened Coulomb forces. Therefore, the property of additivity is very often violated, and consequently the use of a nonextensive entropy is more suitable for real-world applications. Figure 1 depicts Tsallis entropy of a Bernoulli distribution $\boldsymbol{p} = (p, 1-p)$, with different values of the parameter α. As illustrated in Fig. 1, the measure of uncertainty is at a minimum when Shannon entropy is used, and for $\alpha \geqslant 1$, it decreases as the parameter α increases. Furthermore, Tsallis entropy attains a maximum uncertainty when its exponential order α is equal to zero.

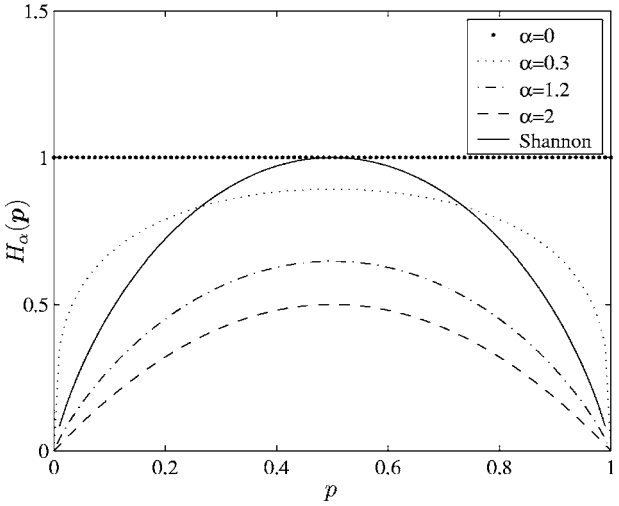

Fig. 1 Tsallis entropy $H_\alpha(\boldsymbol{p})$ of a Bernoulli distribution $\boldsymbol{p}=(p,1-p)$ for different values of α.

Definition 2.1. *Let $\boldsymbol{p}_1, \boldsymbol{p}_2, \ldots, \boldsymbol{p}_n$ be n probability distributions. The Jensen-Tsallis divergence is defined as*

$$D_\alpha^\omega(\boldsymbol{p}_1, \ldots, \boldsymbol{p}_n) = H_\alpha\left(\sum_{i=1}^{n} \omega_i \boldsymbol{p}_i\right) - \sum_{i=1}^{n} \omega_i H_\alpha(\boldsymbol{p}_i),$$

where $H_\alpha(\boldsymbol{p})$ *is Tsallis entropy, and* $\boldsymbol{\omega} = (\omega_1, \omega_2, \ldots, \omega_n)$ *is a weight vector such that* $\Sigma_{i=1}^n \omega_i = 1$ *and* $\omega_i \geqslant 0$.

Using the Jensen inequality, it is easy to check that the Jensen-Tsallis divergence is nonnegative for $\alpha>0$. It is also symmetric and vanishes if and only if the probability distributions $\boldsymbol{p}_1, \boldsymbol{p}_2, \ldots, \boldsymbol{p}_n$ are equal for all $\alpha>0$. Note that the Jensen-Shannon divergence[8] is a limiting case of the Jensen-Tsallis divergence when $\alpha \to 1$.

Unlike other entropy-based divergence measures such as the Kullback-Leibler divergence, the Jensen-Tsallis divergence has the advantage of being symmetric and generalizable to any arbitrary number of probability distributions or datasets, with a possibility of assigning weights to these distributions. Figure 2 shows 3-D representations and contour plots of the Jensen-Tsallis divergence with equal weights between two Bernoulli distributions $\boldsymbol{p} = (p, 1-p)$ and $\boldsymbol{q} = (q, 1-q)$ for $\alpha \in (0,1)$ and also for $\alpha \in (1,\infty)$.

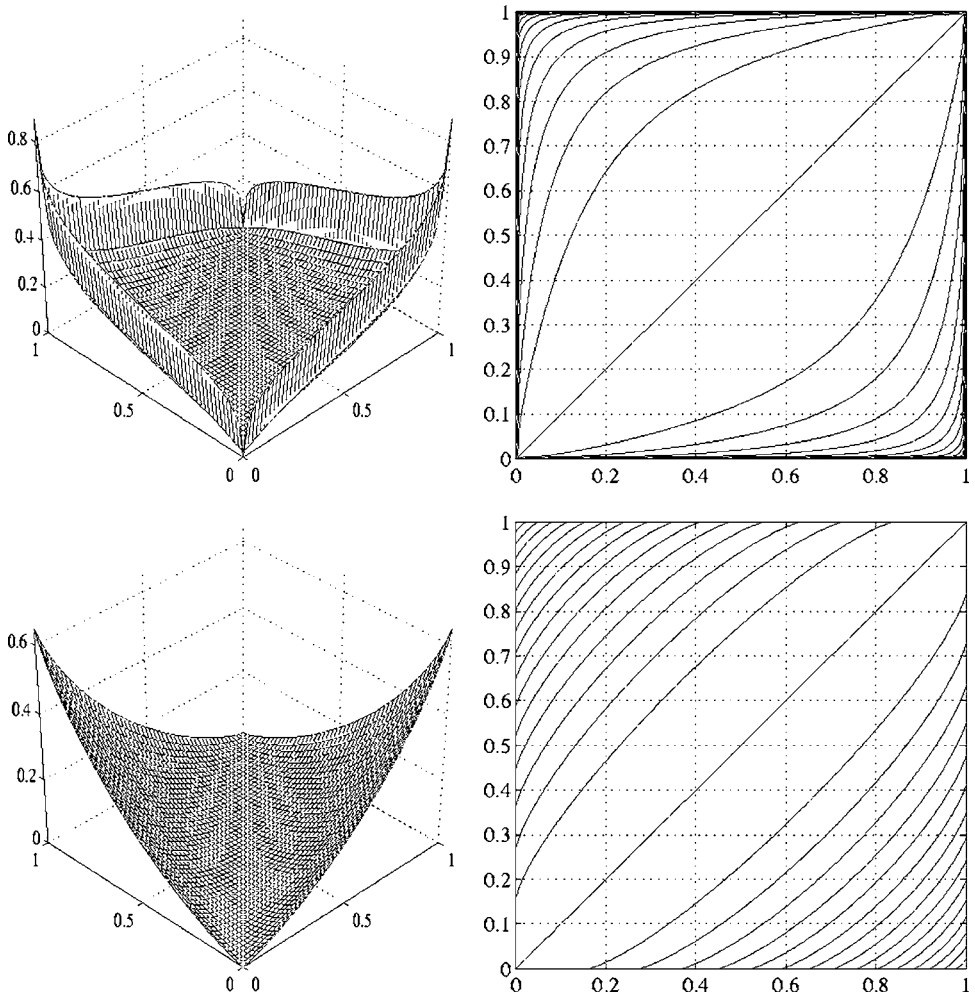

Fig. 2 Surface/contour plots of Jensen-Tsallis divergence between two Bernoulli distributions $\boldsymbol{p}=(p,1-p)$ and $\boldsymbol{q}=(q,1-q)$, and with equal weights $\omega_1=\omega_2=1/2$. First row: $\alpha=0.3$. Second row: $\alpha=1.2$.

3 Properties of the Jensen-Tsallis Divergence

The following result establishes the convexity of the Jensen-Tsallis divergence of a set of probability distributions.[18]

Proposition 1. For $\alpha\in[1,2]$, *the Jensen-Tsallis divergence* D_α^ω *is a convex function of* $\boldsymbol{p}_1,\boldsymbol{p}_2,\ldots,\boldsymbol{p}_n$.

In the sequel, we restrict $\alpha\in[1,2]$ unless specified otherwise. In addition to its convexity property, the Jensen-Tsallis divergence is an adapted measure of disparity among n probability distributions as shown in the next result.

Proposition 2. The Jensen-Tsallis divergence D_α^ω *achieves its maximum value when* $\boldsymbol{p}_1,\boldsymbol{p}_2,\ldots,\boldsymbol{p}_n$ *are degenerate distributions, that is,* $\boldsymbol{p}_i=(\delta_{ij})$, *where* $\delta_{ij}=1$ *if* $i=j$ *and* 0 *otherwise.*

Proof. The domain of the Jensen-Tsallis divergence is a convex polytope, in which the vertices are degenerate probability distributions. That is, the maximum value of the Jensen-Tsallis divergence occurs at one of the extreme points that are the degenerate distributions.

4 Performance Bounds of the Jensen-Tsallis Divergence

4.1 *Maximum Value of the Jensen-Tsallis Divergence*

Since the Jensen-Tsallis divergence is a convex function of $\boldsymbol{p}_1,\ldots,\boldsymbol{p}_n$, it achieves its maximum value when the Tsallis entropy function of the $\boldsymbol{\omega}$-weighted average of degenerate probability distributions achieves its maximum value as well. Assigning weights $\boldsymbol{\omega}_i$ to the degenerate distributions $\Delta_1,\Delta_2,\ldots,\Delta_n$, where $\Delta_i=(\delta_{ij})_{1\leqslant j\leqslant k}$, the following upper bound

$$D_\alpha^\omega(\boldsymbol{p}_1,\ldots,\boldsymbol{p}_n)\leqslant H_\alpha\left(\sum_{i=1}^{n}\omega_i\Delta_i\right), \tag{4}$$

which easily falls out of the Jensen-Tsallis divergence, may be used as a starting point. Without loss of generality, con-

sider the Jensen-Tsallis divergence with equal weights $\omega_i = 1/n$ for all i, and denote it simply by D_α, to write

$$
\begin{aligned}
D_\alpha(\boldsymbol{p}_1, \ldots, \boldsymbol{p}_n) &\leqslant H_\alpha\left[\sum_{i=1}^{n} (\Delta_i/n)\right] \\
&= \frac{1}{1-\alpha}\left\{\sum_{j=1}^{k}\left[\sum_{i=1}^{n} (\delta_{ij}/n)\right]^\alpha - 1\right\} \\
&= \frac{1}{1-\alpha}\left[\sum_{j=1}^{k} (a_j/n)^\alpha - 1\right] \\
&= \frac{1}{1-\alpha}\left(\frac{1}{n^\alpha}\sum_{j=1}^{k} a_j^\alpha - 1\right) \\
&= \frac{1}{(1-\alpha)n^\alpha}\sum_{j=1}^{k} a_j^\alpha - \frac{1}{1-\alpha} \\
&= \frac{H_\alpha(\boldsymbol{a})}{n^\alpha} + \frac{1}{(1-\alpha)n^\alpha} - \frac{1}{1-\alpha} \\
&= \frac{H_\alpha(\boldsymbol{a})}{n^\alpha} + \frac{1-n^\alpha}{(1-\alpha)n^\alpha}, \qquad (5)
\end{aligned}
$$

where

$$
\boldsymbol{a} = (a_1, a_2, \ldots, a_k) \quad \text{such that } a_j = \sum_{i=1}^{n} \delta_{ij}. \qquad (6)
$$

Since $\Delta_1, \Delta_2, \ldots, \Delta_n$ are degenerate distributions, it follows that $\Sigma_{j=1}^{k} a_j = n$. From Eq. (5), it is clear that the maximum value of D_α is also a maximum value of $H_\alpha(\boldsymbol{a})$.

To maximize $H_\alpha(\boldsymbol{a})$, the concept of majorization will be used.[15] Let $(x_{[1]}, x_{[2]}, \ldots, x_{[k]})$ denote the nonincreasing arrangement of the components of a vector $\boldsymbol{x} = (x_1, x_2, \ldots, x_k)$.

Definition 4.1. Let $\boldsymbol{a}$ and $\boldsymbol{b} \in \mathbb{N}^k$. $\boldsymbol{a}$ is said to be majorized by $\boldsymbol{b}$, written $\boldsymbol{a} \prec \boldsymbol{b}$, if

$$
\begin{cases}
\sum_{j=1}^{k} a_{[j]} = \sum_{j=1}^{k} b_{[j]} \\
\sum_{j=1}^{\ell} a_{[j]} \leqslant \sum_{j=1}^{\ell} b_{[j]}, \quad l = 1, 2, \ldots, k-1.
\end{cases}
$$

It can be easily shown that H_α is a Schur-concave function, and therefore $H_\alpha(\boldsymbol{a}) \geqslant H_\alpha(\boldsymbol{b})$ whenever $\boldsymbol{a} \prec \boldsymbol{b}$.

The following result establishes the maximum value of the Jensen-Tsallis divergence.

Proposition 3. Let $\boldsymbol{p}_1, \ldots, \boldsymbol{p}_n$ be n probability distributions. If $n \equiv r \pmod k$, $0 \leqslant r < k$, then

$$
D_\alpha(\boldsymbol{p}_1, \ldots, \boldsymbol{p}_n) \leqslant \frac{C}{(qk+r)^\alpha} + \frac{1-(qk+r)^\alpha}{(1-\alpha)(qk+r)^\alpha}, \qquad (7)
$$

where

$$
C = -r(q+1)^\alpha \log_\alpha(q+1) - (k-r)q^\alpha \log_\alpha(q), \quad \text{and}
$$

$$
q = (n-r)/k.
$$

Proof. It is clear that the vector

$$
\boldsymbol{q} = (\overbrace{q+1, \ldots, q+1}^{r}, \overbrace{q, \ldots, q}^{k-r})
$$

is majorized by the vector $\boldsymbol{a}$ defined in Eq. (6). Therefore, $H_\alpha(\boldsymbol{a}) \leqslant H_\alpha(\boldsymbol{q})$. This completes the proof using Eq. (5).

Corollary 1. If $n \equiv 0 \pmod k$, then

$$
D_\alpha(\boldsymbol{p}_1, \ldots, \boldsymbol{p}_n) \leqslant -k^{1-\alpha}\log_\alpha(q) + \frac{1-q^\alpha k^\alpha}{(1-\alpha)q^\alpha k^\alpha},
$$

where k is the number of components of each probability distribution.

4.2 Jensen-Tsallis Divergence and Mixture Models

Mixture probability distributions provide a multimodal distribution that model the data with greater flexibility and effectiveness, and are used extensively in signal/image processing and computer vision. It is therefore of great interest to evaluate the Jensen-Tsallis divergence between two probability distributions $\boldsymbol{p}$ and $\boldsymbol{q}$ with weights $\{\lambda, 1-\lambda\}$, where $\lambda \in [0,1]$. If we denote by $\boldsymbol{r}$ the weighted probability distribution defined by

$$
\boldsymbol{r} = (1-\lambda/2)\boldsymbol{p} + (\lambda/2)\boldsymbol{q},
$$

then the Jensen-Tsallis divergence can be expressed as a function of λ as follows

$$
D_\alpha(\lambda) = H_\alpha(\boldsymbol{r}) - \frac{H_\alpha[(1-\lambda)\boldsymbol{p} + \lambda\boldsymbol{q}] + H_\alpha(\boldsymbol{p})}{2}.
$$

Proposition 4. The Jensen-Tsallis divergence $D_\alpha(\lambda)$ achieves its maximum value when $\lambda = 1$.

Proof. Let $\boldsymbol{p} = (p_i)_{i=1}^k$ and $\boldsymbol{q} = (qi)_{i=1}^k$ be two distinct probability distributions. The Jensen-Tsallis divergence can then be written as

$$
\begin{aligned}
D_\alpha(\lambda) &= \frac{1}{1-\alpha}\left\{\sum_{i=1}^{k}\left[p_i + \frac{\lambda}{2}(q_i - p_i)\right]^\alpha - 1\right\} \\
&\quad - \frac{1}{2(1-\alpha)}\left\{\sum_{i=1}^{k}[p_i + \lambda(q_i - p_i)]^\alpha - 1\right\} \\
&\quad - \frac{1}{2(1-\alpha)}\left(\sum_{i=1}^{k} p_i^\alpha - 1\right).
\end{aligned}
$$

Using calculus, we can show that $\lambda = 0$ is a critical point of the Jensen-Tsallis divergence $D_\alpha(\lambda)$, i.e., the first derivative

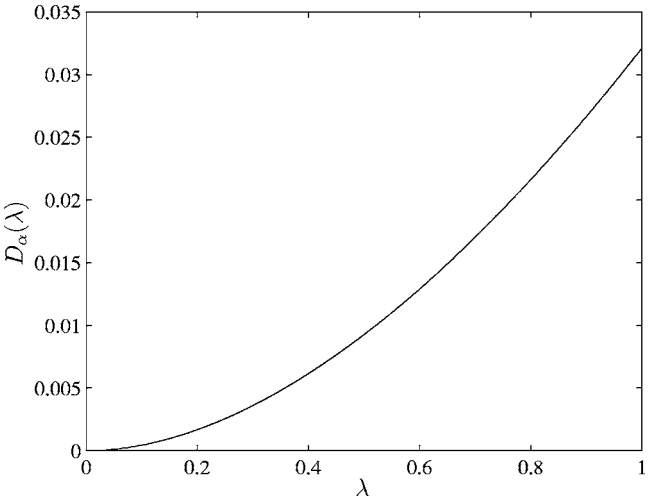

Fig. 3 Jensen-Tsallis divergence as a function of λ.

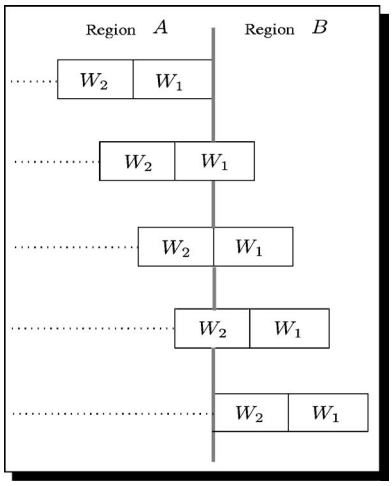

Fig. 4 Vertical edge between two image regions *A* and *B*.

$D'_\alpha(\lambda)$ vanishes at $\lambda=0$. Furthermore, it can be verified that the second derivative $D''_\alpha(\lambda)$ is always positive. Hence, the first derivative $D'_\alpha(\lambda)$ is an increasing function of λ, and therefore $D'_\alpha(\lambda)\geqslant 0$ for all $\lambda\in[0,1]$. Consequently, $D_\alpha(\lambda)$ is an increasing function of λ. This concludes the proof.

Figure 3 depicts the Jensen-Tsallis divergence as a function of λ when $\boldsymbol{p}=(.35,.12,.53)$, $\boldsymbol{q}=(.25,.34,.41)$, and $\alpha=1.2$.

5 Application to Image Edge Detection

5.1 *Description of the Proposed Method*

Edge detection is a fundamental problem in image processing.[9] The use of the Jensen-Tsallis divergence in image edge detection can be formulated as follows. An image sliding window W is split into two equal and adjacent subwindows W_1 and W_2. For each position of the sliding window, the histograms $\boldsymbol{p}_1$ and $\boldsymbol{p}_2$ of the subwindows W_1 and W_2 are computed, as well as the Jensen-Tsallis divergence between $\boldsymbol{p}_1$ and $\boldsymbol{p}_2$. We repeat this procedure for several orientations of the sliding window to ensure detection in all directions, and we then select the maximum divergence at each pixel of a given image. Consequently, we construct a divergence mapping matrix where the largest values in linear neighborhoods are identified. These largest pixel values correspond to edge points. It is worth noting that the window W can be split into any finite number of equal and adjacent subwindows, since the Jensen-Tsallis divergence is defined between any number of probability distributions. In the simulations, we consider without loss of generality the Jensen-Tsallis divergence between the histograms of two equal and adjacent subwindows that form the entire sliding window. To illustrate the behavior of the Jensen-Tsallis divergence, consider an image that consists of two regions A and B with respective histograms $\boldsymbol{p}_a$ and $\boldsymbol{p}_b$, the former histogram on the left and the latter on the right, with a vertical boundary between them as illustrated in Fig. 4. The window W slides from left to right, so that it passes gradually from region A to region B. Under a hypothesis of statistical homogeneity, the histograms $\boldsymbol{p}_1$ and $\boldsymbol{p}_2$ are constant and are equal when W is located in the same region, and hence the Jensen-Tsallis divergence between $\boldsymbol{p}_1$ and $\boldsymbol{p}_2$ vanishes. If the sliding window is located over an edge, one of the two subwindows will have some of its parts in the two regions A and B. Without loss of generality, we consider a subwindow W_2. Again, under the hypothesis of homogeneity, the parts of W_2 located in regions A and B have partial histograms $\boldsymbol{p}_a$ and $\boldsymbol{p}_b$, with weights proportional to the sizes of the subregions $A\cap W_2$ and $B\cap W_2$, respectively. More precisely, if $\lambda\in[0,1]$ represents the fraction of W_2 included in region B, then $\boldsymbol{p}_2=(1-\lambda)\boldsymbol{p}_a+\lambda\boldsymbol{p}_b$, giving rise, consequently, to a histogram that varies with respect to the position of W. Since the subwindow W_1 is located in region A, then it follows that $\boldsymbol{p}_1=\boldsymbol{p}_a$ and therefore $\boldsymbol{p}=(1-\lambda/2)\boldsymbol{p}a+(\lambda/2)\boldsymbol{p}b$. The corresponding Jensen-Tsallis divergence can then be expressed as a function of λ

$$D_\alpha(\lambda)=H_\alpha(\boldsymbol{p})-\frac{H_\alpha[(1-\lambda)\boldsymbol{p}_a+\lambda\boldsymbol{p}_b]+H_\alpha(\boldsymbol{p}_a)}{2},$$

where $\boldsymbol{p}=(1-\lambda/2)\boldsymbol{p}_a+(\lambda/2)\boldsymbol{p}_b$, and $\alpha\in[1,2]$.

It is worth noting that according to Proposition 4, the maximum value of $D_\alpha(\lambda)$ is achieved when $\lambda=1$.

5.2 *Experimental Results*

We conducted a number of experiments using the proposed information-theoretic measure in image edge detection. The performance of the proposed algorithm using the Jensen-Tsallis divergence for various values of α is illustrated in Fig. 5. The results clearly indicate that our proposed signature performs better when the Tsallis entropic exponent α belongs to the interval [1, 2], which validates the theoretical properties given in the previous sections. This better performance is in fact consistent with a variety of images used for experimentation. In particular, we performed additional simulations on images with more complicated structures, including textured images. These images are shown in Fig. 6 along with their corresponding histograms. The image edge detection results depicted in Figs. 7–10 show clearly that the proposed approach outperforms the Canny edge detector and the Jensen-Shannon divergence method. In all the numerical examples presented in this section, we estimated an optimal threshold based on the histogram of

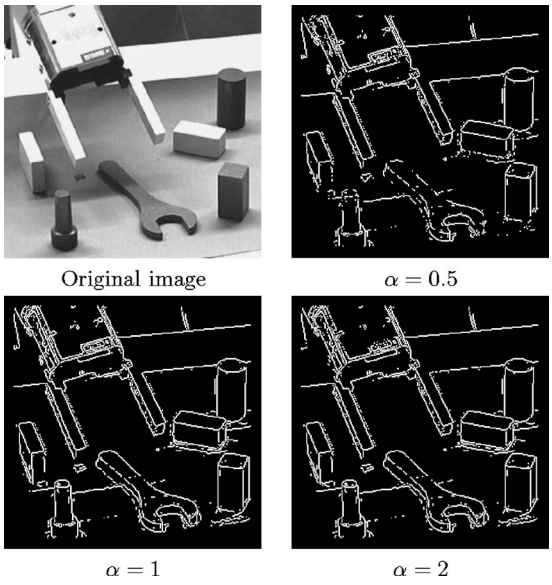

Fig. 5 Edge detection results using Jensen-Tsallis divergence for various values of α.

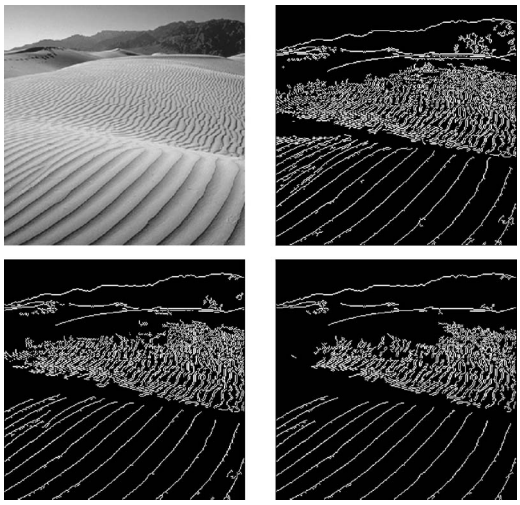

Fig. 7 Image edge detection results. Clockwise from upper left: original image, proposed technique with α=1.5, Jensen-Shannon method, and Canny edge detector.

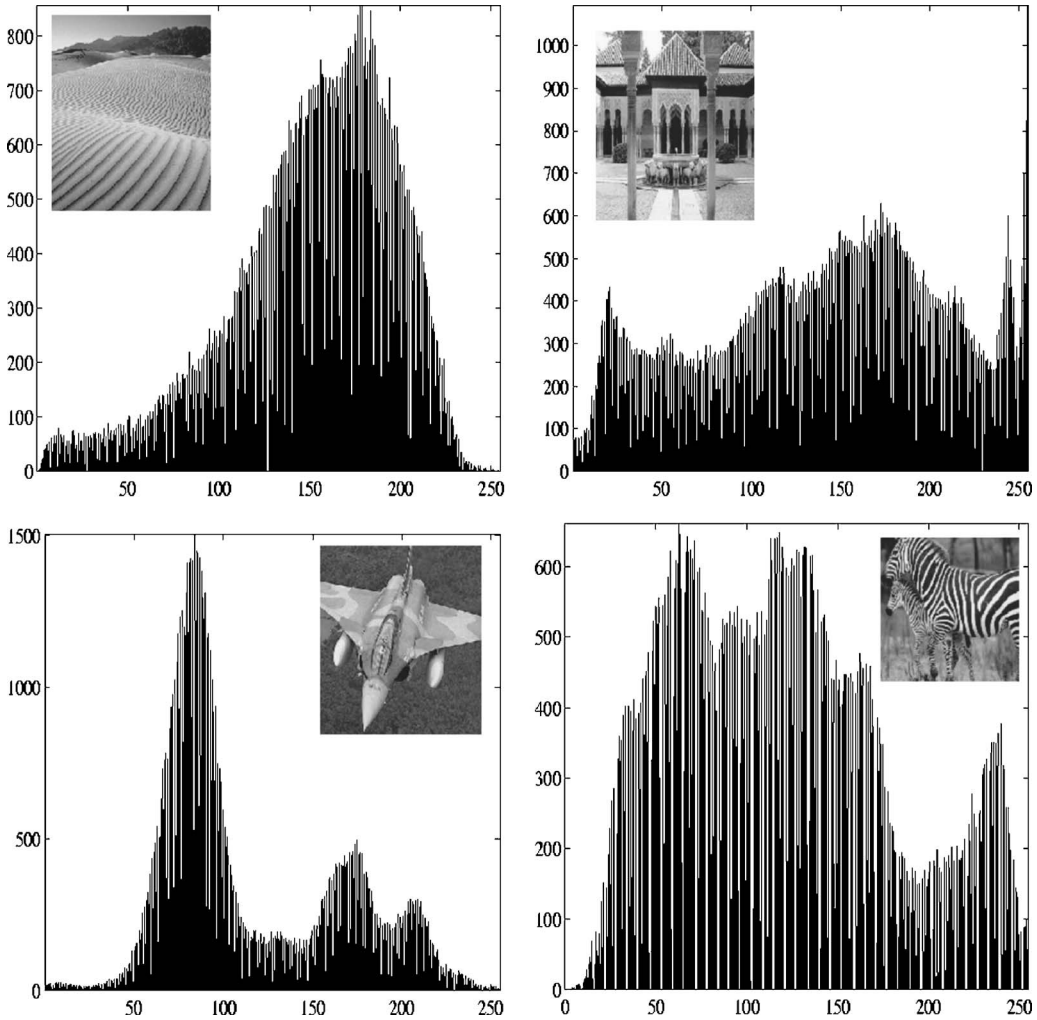

Fig. 6 Histograms of the test images used for experimentation.

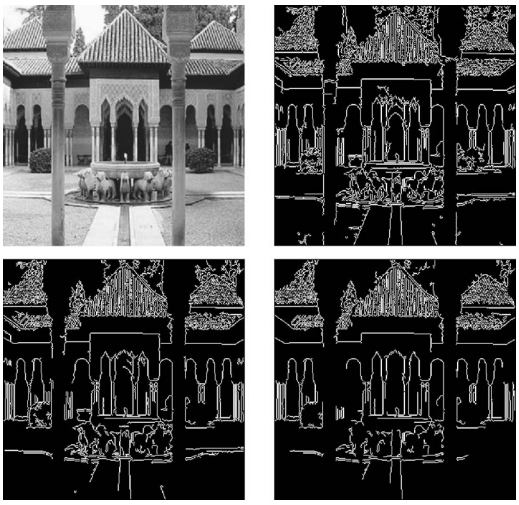

Fig. 8 Image edge detection results. Clockwise from upper left: original image, proposed technique with $\alpha=1.5$, Jensen-Shannon method, and Canny edge detector.

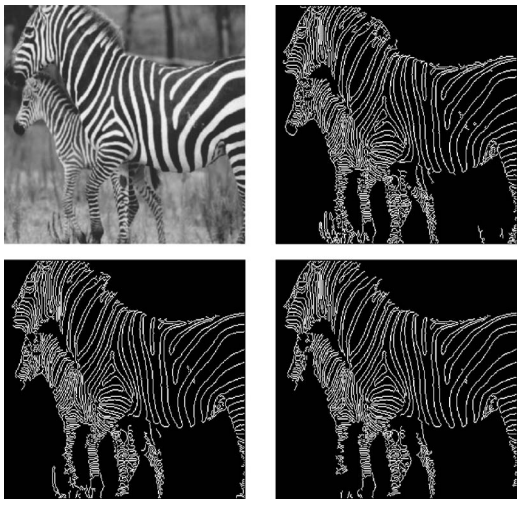

Fig. 10 Image edge detection results. Clockwise from upper left: original image, proposed technique with $\alpha=1.5$, Jensen-Shannon method, and Canny edge detector.

each image used for experimentation. To generate the edge map using the proposed approach, the algorithm takes about 45 sec on a Pentium 4 mobile processor (1.6 GHz, 1 GB of RAM) for a 512×512 image, but it is slower compared to the Canny edge detector. We also used four sliding window orientations: horizontal, vertical, and both diagonals.

On the other hand, our proposed technique overcomes the limitations of conventional gradient-based methods, including the Canny edge detector, which requires a smoothing operation to alleviate the effect of high spatial frequency in estimating the gradient. This smoothing step, however, tends to distort the edges.

6 Conclusions

We propose a nonextensive information-theoretic divergence called Jensen-Tsallis divergence, and we apply it to image edge detection. We consider the application to image segmentation a preliminary investigation. An important issue for further research in this direction is development of robust edge linking algorithms to be included in conjunction with the image edge detection based on the Jensen-Tsallis divergence measure. The proposed information-theoretic measure is very promising due to its simplicity and its potential applications, which may include medical image registration, clustering, classification, indexing, and retrieval. We are currently applying the Jensen-Tsallis divergence to register geographical digital elevation images.

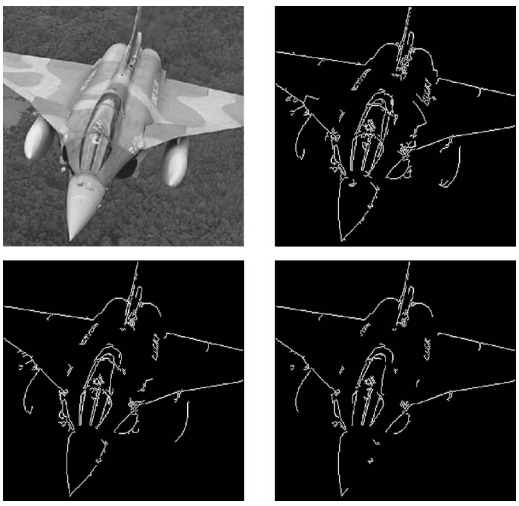

Fig. 9 Image edge detection results. Clockwise from upper left: original image, proposed technique with $\alpha=1.5$, Jensen-Shannon method, and Canny edge detector.

Acknowledgments

This work was supported in part by an NSERC Discovery Grant (No. 311656, NATEQ Nouveaux Chercheurs Grant (No. 2006–NC-106879), and by the Faculty of Engineering and Computer Science at Concordia University.

References

1. S. Ali and S. Silvey, "A general class of coefficients of divergence of one distribution from another," *J. R. Stat. Soc. Ser. B. Methodol.* **28**, 131–142 (1966).
2. D. Brady and M. A. Neifeld, "Information theory in optoelectronic systems: introduction to the feature," *Appl. Opt.* **39**(11), 1679-1680 (2000).
3. S. Kullback and R. Liebler, "On information and sufficiency," *Ann. Math. Stat.* **22**, 79–86 (1951).
4. R. Stoica, J. Zerubia, and J. M. Francos, "Image retrieval and indexing: A hierarchical approach in computing the distance between textured images," *Proc. IEEE Intl. Conf. Image Process.* (1998).
5. A. Srivastava, "Stochastic models for capturing image variability," *IEEE Signal Process. Mag.* **19**(5), 63–76(2002).
6. A. O. Hero, B. Ma, O. Michel, and J. Gorman, "Applications of entropic spanning graphs," *IEEE Signal Process. Mag.* **19**(5), 85–95 (2002).
7. A. Rényi, "On measures of entropy and information," *Sel. Papers Alfréd Rényi* **2**, 525–580 (1961).
8. J. Lin, "Divergence measures based on the shannon entropy," *IEEE Trans. Inf. Theory* **37**(1), 145–151 (1991).
9. J. F. Gomez, J. Martinez, A. M. Robles, and R. Roman, "An analysis of edge detection by using the Jensen-Shannon divergence," *J. Math. Imaging Vision* **13**(1), 35–56 (2000).
10. P. Viola and W. M. Wells, "Alignment by maximization of mutual information," *Int. J. Comput. Vis.* **24**(2), 173–154 (1997).
11. F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens, "Multimodality image registration by maximization of mutual information," *IEEE Trans. Med. Imaging* **16**(2), 187–198 (1997).

12. Y. He, A. Ben Hamza, and H. Krim, "A generalized divergence measure for robust image registration," *IEEE Trans. Signal Process.* **51**(5), 1211–1220 (2003).
13. L. S. Hibbard, "Region segmentation using information divergence measures," *Med. Image Anal* **8**, 233–244 (2004).
14. D. Corici and J. Astola, "Information divergence measures-for detection of borders between coding and noncoding DNA regions using recursive entropic segmentation," *Proc. IEEE Workshop Stat. Signal Process.*, pp. 577–580 (2003).
15. A. W. Marshall and I. Olkin, *Inequalities: Theory of Majorization and Its Applications*, Academic Press, San Diego, CA (1979).
16. M. E. Havrda and F. Charvát, "Quantification method of classification processes: concept of structural α-entropy," *Kybernetika* **3**, 30–35 (1967).
17. C. Tsallis, "Possible generalization of Boltzmann-Gibbs statistics," *J. Stat. Phys.* **52**, 479–487 (1988).
18. J. Burbea and C. R. Rao, "On the convexity of some divergence measures based on entropy functions," *IEEE Trans. Inf. Theory* **28**, 489–495 (1982).

A. Ben Hamza is an assistant professor in the Concordia Institute for Information Systems Engineering (CIISE) at Concordia University, Montreal, Canada. He received his PhD degree in electrical engineering from North Carolina State University in 2003, where he worked on computational imaging and 3-D object recognition. Prior to joining CIISE, he was a postdoctoral research associate at Duke University in North Carolina, affiliated with both the Department of Electrical and Computer Engineering and the Fitzpatrick Center for Photonics and Communications Systems. His research interests include geometric and topological modeling for 3-D object recognition, nonlinear image processing, information-theoretic measures, and multimedia watermarking.