

Survey and Comparative Analysis on Entropy Usage for Several Applications in Computer Vision

Nitin Chamoli

Graphic Era University,
566/6, Bell Road, Dehradun,
Uttarakhand, India

Sneh Kukreja

Graphic Era University,
566/6, Bell Road, Dehradun,
Uttarakhand, India

Monika Semwal

Graphic Era University,
566/6, Bell Road, Dehradun,
Uttarakhand, India

ABSTRACT

This paper presents a thorough study of different types of entropies. Application and comparison of various entropies have been considered with their effectiveness and suitability in different applications being explored. The usage of entropy in the fields of image thresholding, image reconstruction, image segmentation, and incorporation of entropy in tackling real life problems have been mentioned categorically. A comparative analysis of different forms of entropy accordingly to their suitability for various applications has been discussed.

Keywords

Entropy, computer vision, thresholding, segmentation, restoration, registration

1. INTRODUCTION

From a conventional point of view, the entropy is a basic thermodynamic concept that is associated with the order of irreversible processes in the universe. Entropy has already been known in thermodynamics and statistical mechanics. In statistical thermodynamics the most general formula for the thermodynamic entropy S of a thermodynamic system is the Gibbs entropy.

$$S = -Kb \sum P_i \ln P_i$$

Where, Kb is the Boltzmann constant and P_i is the probability of microstate.

Motivated by the problem of efficient transmission of information over a noisy communication channel; Claude Shannon redefined the entropy concept of Boltzmann/Gibbs as a measure of uncertainty regarding the information content of a system. He introduced a new revolutionary probabilistic way of thinking about communication and simultaneously created the first truly mathematical theory of entropy in 1948. He defined an expression for measuring quantitatively the amount of information produced by a process.

His line of research was based upon two categories: (a) Information theory, (b) Coding theory; [1]

$$H = -K \sum_{i=1}^n P_i \ln P_i$$

Where, K is a positive constant. The constant K merely amounts to a choice of a unit of measure. P_i is the probability of a system being in cell I of its phase space. Consider there be a finite number of events and the probability of their occurrence is $p_1, p_2, p_3, p_4 \dots p_n$ which give a measure $H(p_1,$

$p_2, p_3 \dots p_n)$. H should be continuous in the p_i , also, H should be a monotonic function of n if all p_i are equal, $p_i=1/n$, with equal likely events there is more choice, or uncertainty when there are more possible events. If a choice be broken down into two successive choices, the original H should be the weighted sum of the individual values of H .

Logarithmic measure is used in entropy as most parameters of engineering importance such as time, bandwidth, number of relays, etc, tend to vary linearly with the algorithm of the number of possibilities. One feels that two punched cards should have twice the capacity of one for information storage, and two identical channels twice the capacity of one for transmitting information, therefore, logarithmic measure is closer to our intuitive feeling as to the proper measure. Many of the limiting operations are simple in terms of the logarithm but would require clumsy restatement in terms of the number of possibilities which makes it mathematically more suitable.

In this paper we investigate different forms of entropy that have been utilized for various applications in computer vision. In particular we try to relate, and compare different forms of entropy on the basis of their application and conclude the suitability of a form of entropy for a particular application. The paper is organized as follows, in section 2; it brief states a literature survey on the usage of various entropies in different applications in computer vision. Section 3, presents a comparison between entropies, following it section 4 gives applications, and finally, conclusion is given in section 5.

2. LITERATURE SURVEY

Entropy is used in image processing majorly in image thresholding. Image thresholding often represents a first step in image understanding. The use of maximum entropy in image thresholding was done by Pun [2,3] which was later improved by P.K. Sahoo et al. [4] this concept was then generalized to Renyi entropy by Wilkins, Yeager et al. [5]. Relative entropy (also, known as the Kullback-Leibler divergence, cross entropy) has primarily been used for thresholding purpose. In Kittler et al.[6] problem of minimum error thresholding has been solved assuming that object and pixel gray level values are normally distributed, it has been supposed that the thresholded image can be modeled by a mixture of two Gaussian Distributions, with one distribution describing the image background, and another describing the foreground of image. The relative entropy is used as a matching criterion and the desired optimal threshold that produce a two member Gaussian mixture will best match the original 1-D image histogram. The focus is to minimize the relative entropy which is like finding a two member Gaussian mixture with minimum discrepancy between the original image and the thresholded image. Compared to Kittler et al. [6] for thresholding, the relative entropy and Gaussian

mixture model are replaced by the divergence and a Poisson model. Pal and Pal [8] used co-occurrence matrix to define second order local and joint entropy. The local entropy is defined in two quadrants, background to background (BB) and object to object (OO) only, whereas the joint entropy is defined in two quadrants only, that is, background to object (BO) and object to background (OB). Based on these two definitions two algorithms were developed which maximizes local and joint entropy. Rather than looking into entropies of background and object individually Chang Chen et al. [9] further introduced the concept of relative entropy which was different from the two known thresholding techniques known then. The method implemented is to calculate the grey-level transition probability distributions of the co-occurrence matrices for an image and a thresholded bi-level image, and then their relative entropy, which is to find a threshold which minimizes the discrepancy between these two transition probability distributions. The smaller the discrepancy, the better the matching between the original image and the thresholded image will be. The threshold rendering the smallest relative entropy will be selected to segment the image. Further, the work of Chang et al. has been improved by Jianwei Wang et al. [10] by introducing local relative entropy thresholding (LRE) and joint relative entropy thresholding (JRE). Relative entropic thresholding is generally sensitive to sparse histograms, a histogram compression and translation (HCT) is proposed to improve and enhance their performance. JRE and LRE generally perform better than Chang et al.'s [9] approach with the HCT when histograms are sparse. Abutaleb [11] uses 1-dimensional histogram of an image, to separate, effectively, the desired objects from the background. This method doesn't take into consideration the spatial correlation between the pixels in an image. Two-dimensional entropy is used, where the gray-level value of each pixel as well as the average value of its immediate neighborhood is studied. Thus, the threshold is a vector and has two entries: the gray level of the pixel and the average gray level of its neighborhood. The vector that maximizes the 2-dimensional entropy is used as the 2-dimensional threshold.

Next, for image segmentation the work has been done by N.R. Pal et al. [7] where an ideal image model for gray tone digital image has been done. Entropy maximization and X^2 minimization are the principles adopted in the formulation of algorithms. The algorithm based on the Poisson distribution produces better segmentation result than the method of Otsu. Others that involved entropy for image segmentation include Lothar Hermes et al. [12] which involve minimum entropy. An algorithm based on an intuitive generative model for pixel intensities, with its associated cost function which can be effectively optimized by a hierarchical triangulation algorithm is formulated. The algorithm involves images by polygon segments. Sahoo, Arora [13] developed a method of image segmentation that employs Renyi entropy of order α , and a two dimensional histogram. This method of thresholding incorporates Abutaleb's method [11], and the two-dimensional version of the entropic correlation method. In almost every image used, the proposed method yielded a good threshold value for fractional α , that is, when $0 < \alpha < 1$. Tsallis entropy is utilized for thresholding by Albuquerque et al. [14]. The advantage of the method is the use of a global and objective property of the histogram and this method is easily implemented. The Tsallis q coefficient can be used as an adjustable value and can play an important role as a tuning parameter in the image processing chain for the same class of

images. This can be an advantage when the image processing tasks depend on an automatic thresholding.

Entropy has been utilized in tackling various real life problems. Beane et al. [15] has utilized maximum entropy in identifying area of suitable red spruce habitat, analyzing a total of 32 variables like warmest month, minimum temperature of the coldest month. The Maximum Entropy Method, as introduced by Alhassid et al. [16] (1977), is a means for deconvolving "truth" from an image and PSF. Hence, maximum entropy has been used most spectacularly in radio astronomical interferometry where it deals routinely with images of up to a million pixels, and high dynamic range. Also, the maximum entropy method allows incorporating extra, prior knowledge about the object being imaged, and this technique being used in spectroscopy. Information about the cell from a microscope is gathered using Shannon and Renyi entropy by Tomas Nahlik [17]. In an interesting approach Haidar Sharif et al. [18] use Shannon entropy to detect occurrence of abnormal activities in a video stream, here a classic example of accidents in an escalator is used as data set. Entropy here can be calculated by considering the error function for different frames extracted from the videos. The frames resulting in a higher error function will have higher entropy. Lothar Hermes [12] has large potential in applying their method to voxel data, as they occur in tomography. Modified C4.5 decision trees based on Tsallis and Renyi entropies have been tested on several high-dimensional microarray datasets with interesting results by Maszczyk et al [19]. The data sets include DLBCL (Diffuse Large B-cell Lymphoma), Leukemia [training dataset consists of 38 bone marrow samples (27 ALL and 11 AML), over 7129 probes from 6817 human genes. Also, 34 samples testing data is provided, with 20 ALL and 14 AML cases. According to Katiyar [20] in the field of remote sensing it was found that as Tsallis entropy is preferable for thresholding and clustering. Renyi is the simplest entropy method and is computationally simple as it avoids parametric estimation. The various experiments revealed that Shannon is computationally complex entropy variation, and hence, is less preferable in the above discussed operations. Since Renyi and Tsallis are a limiting case of Shannon for a better accurate estimation Shannon may be adopted. An entropy based method for human motion detection is described in Jing et al. [21] using spatial entropy. The difference image between consecutive images are calculated and at each particular pixel, and a spatio-temporal histogram is generated by accumulating pixels in difference image. This histogram is then normalized to calculate entropy and the magnitude of entropy is used to denote the significance of motion.

In the field of image restoration Frieden et al. [22] were the first ones to use entropy, and applied maximum entropy in astronomy imagery. The maximum entropy method occupies a privileged position among methods for reconstructing images from incomplete and noisy data. It is the only way of consistently combining different data into a single image. Nonlinear data may be inherently ambiguous, so that there may be several local maxima of entropy. In any such case of grossly inadequate data, extra information must be introduced. Extending their work further Frieden [23] used maximum entropy, and the restoring method was applied to experimental data consisting of line spectra. Results were consistent with those of the computer simulations. They took pragmatic examples where noise in the signal is poisson, that is, dominated by single photon statistics, with noise no longer being additive, particularly the work focus on astronomy. Dominikus [24] propose an alternative for inverse weiner

filtering for restoration of degraded images due to system imperfections using maximum entropy. Two problems were encountered and solved- first, to identify blur and second, to restore the image as close as possible to the not degraded version. Michael Gary Grotenhuis [25] uses a maximum entropy maximization algorithm (MEM) which is iterative and with each iteration making an improvement upon the deconvolved image. The end result of using the MEM is smoothest image possible, equivalent to the greatest entropy. Further, Hong et al. [26] extend the use of the classical EM algorithm for image recovery from randomly sampled data and sensor field estimation from randomly scattered sensor networks. Also, further to use this approach in density estimation, image recovery and sensor field estimation.

Mondal, Rajan [27] in their work implement conditional entropy in image reconstruction for positron image tomography. It involves maximization of conditional entropy as a cost function rather than the likelihood or posterior function, ensuring the positivity of reconstructed image pixels since entropy is not defined for negative pixels. There is inclusion of prior distribution knowledge of the object during image reconstruction process. For the modeling of prior Gibbs distribution function is used.

In other applications, entropy has been used as a solution to the missing short-baseline problem in the synthesis observations in [28]. Using relative entropy it is shown in the article that a measure of distance between the prior model and the image in the plane of pixel brightness is an adequate measure of relative entropy. Minimum-relative-entropy method (MREM) has been presented as a possible scheme for reconstructing extended sources mapped with sparsely sampled short baselines. Next, the work of Fotinos et al. [29] demonstrates a new color edge detection operator that uses the shape information provided by the local probability distribution of the image which is investigated and evaluated. The proposed approach is based on the Relative Entropy index (E_{rel}), which has long been used as a distance measure between two distributions. The unknown density is locally estimated and this estimate is again sampled to compute the relative entropy. The above study indicates that the Relative Entropy index can be used for the analysis of color and other multichannel images. Entropy is used in image compression in Sukanesh et al. [30]. A two stage compression process is performed through homogenous MRE method, and heterogeneous MRE. The performance of image compression is analyzed by the estimation of parametric values such as Mean Square Error (MSE) and Peak Signal to Noise Ratio (PSNR). Wang et al. [31] investigates the potential of spatial entropy in measuring image information and proposes a new mutual information (MI) band selection method based on the spatial entropy. Then selected band images are validated for supervised classification via Support Vector Machine (SVM). Further, Urban et al. [32] developed a method specifically for use on microscopy images captured in phase-contrast mode. However, the approach is generally applicable to any image. The basic concept is that the background remains informatively poor, whereas the objects carry relevant information. This method preserves the details, highlights edges, and decreases random noise; all of this is done in one calculation.

3. COMPARISONS

In this section entropies have been compared on various parameters. The strong points of an entropy, categorically, for a particular application enable it to give the best solution while the limitations of other entropies make them less

preferable. The applicability of entropy for many remote sensing operations is considered. For good image registration operation least RMSE value is preferred as the RMSE value indicates the error in registration, also, entropy with least execution time is preferred. The strong point of Renyi entropy is that, it produces the least RMSE value followed by Tsallis entropy, Shannon entropy is the least preferred for registration. For thresholding the average score, correlation coefficients were used for accuracy estimation. Greater value of correlation and average score corresponds to better thresholding technique. Tsallis entropy for this particular application gives the most accurate outcome. For segmentation purpose, the increase in value of kappa statistics (maximum = 1) and overall accuracy (max=100%) indicates segmentation is Tsallis entropy followed by Renyi entropy then Shannon entropy. Performances of different entropies for the discussed operations have been summarized in table 1.

Table 1

Application	Most suitable Entropy
Image registration	Renyi then Tsallis then Shannon entropy
Thresholding and clustering	Tsallis entropy
Image segmentation	Tsallis then Renyi then Shannon entropy

For automatic thresholding of gray level images the positive of 2-dimensional entropy is that it performs better than 1-dimensional entropy especially when the signal to noise ratio (SNR) is decreased. Both entropies yield comparable outcomes when the SNR was high (more than 12 dB). The SNR is defined as the 10 log of the ratio of the noise-free image power to the noise power. For a Gaussian noise with different values of the variance, at each pixel the noise-free signal power is calculated as the squared value of the gray level. The noise-free image power is then the summation of these values over the whole image. The noise power is calculated in a similar manner. The result has been briefed in table 2.

Table 2

SNR	Most suitable entropy
High SNR	2 Dimensional and 1 Dimensional entropy are comparable
Low SNR	2-dimensional entropy

In hyper spectral band selection in remote sensing the weak point of Shannon entropy for not being accurate in measuring image information is it neglects the spatial distribution of pixels and is computed only from a histogram, thereby, making spatial entropy better than Shannon entropy. For another application for identification of Choquet Integral's parameter and tomographic images the values of the relative entropy for different subsets of attributes are considered as the values of Choquet capacity coefficients. Kojadinovic entropy gives importance to the attributes which allow differentiating

many classes; the relative entropy based method proposed focuses on the researched classes. Therefore, Relative entropy performs better than Kojadinovic entropy for this application. Further, for system exhibiting q-exponential behavior two parameters have been analyzed- first is the sensitivity to initial conditions and the entropy increase rates, and second is the generalized bit cumulant approach. The qualifying parameters for Tsallis entropy being better than Renyi entropy are that Tsallis entropy as function of time t, reproduces the correct $q \rightarrow 1$ limit of BG entropy proportional to t, and Tsallis entropy is a much more sensitive function than Renyi entropy with respect to changes in q value. For image segmentation the approving factor for maximum entropy is that it can robustly separate face image from background image better than Otsu approach. The performance of the discussed entropies for the above mentioned applications is presented in a concise way in table 3.

Table 3

Application	Most suitable entropy
Hyper spectral band selection in remote sensing	Spatial entropy better than Shannon's entropy
Identification of choquet Integral's parameter and tomographic images	Relative entropy better than Kojadinovic entropy
System exhibiting q-exponential behaviour	Tsallis's entropy better than renyi's
Face image segmentation	Maximum entropy better than Otsu's method

4. APPLICATIONS

The application of entropy in computer vision has been vast and dynamic. Based on the study, a summarized overview of different entropies being implemented for several applications is presented. A comprehensive and precise description of entropy versus their applications has been briefed in table 4.

Table 4

Entropy	Application	
Maximum entropy	Most suited for image reconstruction	
	Radio interferometry	
	Optical deconvolutions	
	Tomography	
	Spectroscopy	
	Used to detect suitable red spruce forest habitat	
Renvi entropy	Used in image thresholding	
	A	Threshold value
	$0 < \alpha < 1$	Good threshold value

	$\alpha > 1$	Not good threshold value
	$\alpha = 1$	Good threshold in some cases
Tsallis entropy	Image thresholding	
	Best suitable when applied to noisy gray-level distribution composed of unimodal histogram	
	Best suitable for system exhibiting q-exponential behavior.	
Spatial entropy	Human motion detection	
	Hyper spectral band selection in remote sensing.	
Minimum entropy	Natural photographic imagery segmentation	
	Visual object recognition	
Conditional entropy	Visual saliency detection	
	Image reconstruction for PET system	
Shannon's entropy	Preprocessing of microscopy images	
	Edge detection in gray level images	
Cross entropy	Image segmentation	
	Image thresholding	
	Image reconstruction	
Relative entropy	Used in color edge detection technique	
	Image thresholding	
Minimum relative entropy	Missing short baseline problems	
	Image compression	

5. CONCLUSION

In this paper, a comparative and comprehensive study of different types of entropy and their usage in computer vision has been done. It has been attempted to cover all the major entropies that have been used in computer vision. Majorly, entropy has been used for the purpose of image thresholding. The suitability of entropy for various operations enhances accuracy with a shift from Shannon entropy to Renyi or Tsallis entropy or to other derived forms. The specific use of entropies along with their suitability for a particular application has been discussed, and comparison with other forms of entropies wherever possible has been made. This paper contributes to plainly get an idea of the status of entropy usage in the fields of image thresholding, image reconstruction, image segmentation, and real life usage.

6. ACKNOWLEDGMENT

Our sincere gratitude and thanks to Mr. Vikas Tripathi for his unconditional support and guidance.

7. REFERENCES

- [1] C. E. Shannon 1948. A Mathematical Theory of Communication.
- [2] Pun, T. 1980 A new method for grey-level picture thresholding using the entropy of the histogram.
- [3] Pun, T. 1981 Entropic thresholding: a new approach.
- [4] Kapur, J.N., Sahoo, P.K., and Wong, A.K.C. 1985 A new method for grey-level picture thresholding using the entropy of the histogram.
- [5] Sahoo, P., Wilkins, C., and Yeager, J. 1997 Threshold selection using Renyi's entropy.
- [6] Kittler, J., and Illingworth, J. 1986 Minimum error thresholding.
- [7] Pal, N.R., and Pal, S.K. 1991 Image model, Poisson distribution and object extraction.
- [8] N.R. Pal and S. K. Pal. 1989 Entropy thresholding, Signal Process.
- [9] Chang, C.-I, Chen, K., Wang, J., and Althouse, M.L.G. 1994. A relative entropy-based approach to image thresholding.
- [10] Jianwei Wang Eliza Yingzi Du Chein-I Chang. Relative entropy-based methods for image thresholding. 0- 7803-7448-7/02/\$17.00 02002 IEEE
- [11] Ahmed S. Abljtaleb. 1989 Automatic Thresholding of Gray-Level Pictures Using Two-Dimensional Entropy
- [12] Lothar Hermes, and Joachim M. Buhmann 2003 A Minimum Entropy Approach to Adaptive Image Polygonization
- [13] Prasanna K. Sahoo; Gurdial Arora. 2003 A thresholding method based on two-dimensional Renyi's entropy.
- [14] M. Portes de Albuquerque a,*, I.A. Esquef b, A.R. Gesualdi Mello a, M. Portes de Albuquerque. 2004 Image thresholding using Tsallis entropy.
- [15] Nathan R. Beane James S. Rentch Thomas M. Schuler. 2013 Using maximum entropy modeling to identify and prioritize red spruce forest habitat in West Virginia.
- [16] N. Agmon, Y Alhassid, R.D. Levine 1977. An algorithm for finding the distribution of maximum entropy.
- [17] Tomas Nahlik. Using Shannon and renyi entropy in microscopic image processing. University of South Bohemia, Institute of physical bioogy, Nove Hradý.
- [18] Md. HaidarSharif , ChabaneDjeraba, 2012. An entropy approach for abnormal activities detection in video streams.
- [19] Tomasz Maszczyk and W_lodzis_law Duch. Comparison of Shannon, Renyi and Tsallis Entropy used in Decision Trees", Department of Informatics, Nicolaus Copernicus University Grudzi,adzka 5, 87-100 Toru'n, Poland.
- [20] S.K. Katiyar Arun P.V1. A Comparative Analysis on the Applicability of Entropy in remote sensing. Dept. Of Civil MANIT-Bhopal, India Ph: +914828149999.
- [21] Guo Jing, Chng Eng Siong and Deepu Rajan. Foreground motion detection by difference-based spatial temporal entropy image. School of Computer Engineering Nanyang Technological University, Singapore 639798 fguoj0005, aseschn, asdrajan g@ntu.edu.sg
- [22] Frieden, B.R. 1972. Restoring with maximum likelihood and maximum entropy.
- [23] Frieden, B.R., and Wells, D.C. 1978 Restoring with maximum entropy: Poisson sources and backgrounds.
- [24] Dominikus Noll 1997, Restoration of degraded images with maximum entropy.
- [25] Michael Gary Grotenhuis. An Overview of the Maximum Entropy Method of Image Deconvolution. A University of Minnesota – Twin Cities "Plan B" Master's paper.
- [26] Hunsop Hong and Dan Schonfeld.2008 Maximum-Entropy Expectation-Maximization Algorithm for Image Reconstruction and Sensor Field Estimation.
- [27] Mondal, Pratha Pratim, 2003. Conditional entropy maximization for PET.
- [28] J. Astrophys. Astr., R. K. Shevgaonkar. 1986 minimum-relative-entropy method—solution to missing short-baseline problem.
- [29] Fotinos, George Economou and Spiros Fotopoulos. Using the relative entropy as a color edge detector. University of Patras, Patras 26100, Greece. E-mail: spiros@physics.upatras.gr.
- [30] Dr. (Mrs.) R.Sukanesh, R.Harikumar, N.S.Balaji and S.R.Balasubramaniam, 2007. Analysis of Image Compression by Minimum Relative Entropy (MRE) and Restoration through Weighted Region Growing Techniques for Medical Images.
- [31] Baijie Wang, Xin Wang and Zhanxin Chen. 2012 Spatial entropy based mutual information in hyper spectral band selection for supervised classification.
- [32] Jan Urban, Jan Vaneřk, and Dalibor Štys. Preprocessing of microscopy images via Shannon's entropy", Department of Bioengineering, Institute of Physical Biology, University of South Bohemia, Zařmek 136, NoveřHradý, 37333, Czech republic urban@ufb.jcu.cz.