International Journal of Engineering Inventions
ISSN: 2278-7461, ISBN: 2319-6491, www.ijeijournal.com
Volume 1, Issue 10 (November2012) PP: 14-18

Information Theory and its Application to Pattern Recognition in Image Processing

Yahia Sabri Al-Halabi
Department of Computer Science, Princess Sumaya University for Technology

Abstract:—Pattern recognition is a very growing field in computer science. This research describes a technique for the assessment of information contents of pattern. It discusses the use of information theory measures in pattern recognition problems. Also, an iterative techniques such as fixed point iterative method or binary search method are ideal to be used to find pattern. Information theory is closely associated with a collection of pure and applied disciplines that have been investigated and reduced to engineering practice under a variety of topics throughout the world over the past half century or more: adaptive systems, anticipatory systems, artificial intelligence, complex systems, complexity science, cybernetics, informatics, machine learning, along with systems sciences of many descriptions. Information theory is a broad and deep computer and mathematical theory, with equally broad and deep applications, amongst which is the vital field of coding theory. This research applies a technique used to assess the information to the recognition capability and to increase the efficiency of the pattern recognition related with information theory. Entropy and conditional entropy have been used for learning models and designing inference algorithms. Many techniques and algorithms are developed over the last decades. Most of them involve extraction some of the information that describes a pattern. Results obtained that the computed orientation information contents agree with the theory, which is in the limit goes to zero, in case of orientation pattern and that the information contends depend strongly on size and information. Using fixed point iteration technique is new method to be used to find area of pattern depending on shifting and matching. Application of the entropy, fundamentals of information theory, assessment of the translational and rotational information contents of patterns, and assessment of the total information contents used in this technique prove that this technique is suitable for recognition problems and that information theory measures, are an important tool for pattern recognition using iterative fixed foint application.

Keywords:—Information Theory, Entropy, Fixed Point, Rotational and Translational, Pattern.

I. INTRODUCTION

Information theory is closely associated with a collection of pure and applied disciplines. Information theory is a broad and deep computer and mathematical theory, with equally broad and deep applications, amongst which is the vital field of coding theory. Information in the information theory sense of the word should not be confused with information as we commonly understand it. According to Shannon and Weaver, information is defined as "a measure of one's freedom of choice when one selects a message". In information theory, information and uncertainty are closely related. Information refers to the degree of uncertainty present in a situation. The larger the uncertainty removed by a message, the stronger the correlation between the input and output of a communication channel, the more detailed particular instructions are the more information is transmitted. Uncertainty also relates to the concept of predictability. When something is completely predictable, it is completely certain. A related term, entropy, is also very important in information theory. Entropy refers to the degree of randomness, lack of organization, or disorder in a situation. Information theory measures the quantities of all kinds of information in terms of bits (binary digit). Redundancy is another concept which has emerged from the information theory to communication. Redundancy is the opposite of information. Something that is redundant adds little, if any, information to a message. Redundancy is important because it helps combat noise in a communicating system (e.g. in repeating the message). Noise is any factor in the process that works against the predictability of the outcome of the communication process. Information theory has contributed to the clarification of certain concepts such as noise, redundancy and entropy. These concepts are inherently part of the communication process [1].

One of the most and very growing fields in computer science is the field of pattern recognition. Computer Vision and Pattern Recognition are extremely important research fields with an enormous range of applications. They are also extremely difficult. It is involved heavily in many practical fields such as; medical and other scientific problem. Neural network is also used widely in the last twenty years as a major tool for pattern recognition. Standard information theory [6, 7] also specifies how to encode data which obeys a probability distribution and encode it so that it can be transmitted and then decoded. The connections between information theory and computer vision are deep. Vision can be considered to be a decoding problem where the encoding of the information is performed by the physics of the world — by light rays striking objects and being reflected to cameras or eyes. Ideal observer theories were pioneered by scientists like Barlow [1] to compute the amount of information available in the visual stimuli and to see how efficient humans are at exploiting it. This research discusses the use of information theory measures in pattern recognition problems. We will also describe how concepts, measures, and techniques from information theory can be applied to vision.

Applications of fundamental topics of information theory include lossless data compression (e.g. ZIP files), lossy data compression (e.g. MP3s and JPGs), and channel coding (e.g. for Digital Subscriber Line (DSL)). The field is at the intersection of mathematics, statistics, computer science, physics, neurobiology, and electrical engineering

Additionally, information theory is closely associated with a collection of pure and applied disciplines that have been investigated and reduced to engineering practice. The fundamental problem of communication is that of reproducing at one point, either exactly or approximately, a message selected at another point. New ideas related are the information entropy and redundancy of a source, and its relevance through the source coding theorem, the mutual information, and the channel capacity of a noisy channel, including the promise of perfect loss-free communication given by the noisy-channel coding theorem and finally the practical result of the Shannon–Hartley law for the channel capacity of a Gaussian channel [1].
Entropy and conditional entropy have been used for learning models and designing inference algorithms. Many techniques and algorithms are developed over the last decades. Most of them involve extraction some of the information that describes a pattern.

A huge amount of information is available and it is a difficult tasks to analyses such redundant information. Most available techniques for pattern recognition can be as in template matching approach, syntactic and structure approach or decision approach.

These techniques are not general and concentrate on algorithmic development and not on information assessment [3, 5,9]. Mutual information [6] has been used as a measure for finding the correspondence between two images. This is particularly useful for applications in medical images where the task is to match images that were taken under different parameter settings which cause non-linear transformations to the images. The intensities of corresponding points in the images may differ greatly but mutual information is largely invariant to.

One of the main problems faced in the development of pattern recognition algorithms is assessment of their performance. This paper will discuss the use of information theory measures in pattern recognition. Will describe how techniques from information theory can be used to analyze vision models and measure the effectiveness of different view. The work here concentrates on entropy topic. Initially, one has to find the information content of the set of the input binary image then assess the information that could be used to define different patterns in order to find a means of recognition, and finally to assess recognition capability of pattern recognition system. In this work we first used an information model applicable to any pattern, and its elaboration to measure recognition performance, and second we used this model to derive parameters such as the resolution required to distinguish between the patterns. This has resulted in a powerful method for assessing the performance of any pattern recognition system. This paper describes the development of a technique for the assessment of information content of 2-D patterns encountered in practical pattern recognition problems which is the product of shift in actual y-direction multiplied by the number of matches found in x-coordinates for that fixed point.

Fixed point iteration technique will be used to generate solutions of root finding, and then using such results to build an area which is the product of shift in actual y-direction multiplied by the number of matches found in x-coordinates for that fixed point, after applying orientation and translation which defines its pattern. In the recognition phase, a hypothesis indicates a presumed pattern and, furthermore, indicates where such further prominent areas should be located in the pattern to be recognized if the presumption information correct. Patterns are learned through their location information, orientation, translational and through their spatial relationship to each other, on the other hand, stored as an information and then re-recognized.

II. THEORETICAL BACKGROUND

One of the main branches of mathematical theory of probability and statistics is the subject of information theory. Its main interest is in studying the collection and manipulation of information any probabilistic. Three types of measures for information are required, namely: Total information, Rotational information, and Translational information.

The information related with input pattern of the binary image while all its orientations and locations considered by mean of the most important pixel, is the total information. The information obtained by rotating the pattern of the binary image around fixed position, which is the center of gravity related to matrix of sensors, is also related to rotational information. Image is rotated by many degrees on the right or left and produces many patterns detected by the sensors. Translational information and rotational information are the same except that the point in which the image rotates around it is flexible. Point will be changed around the image and in each new location it will be rotated

The basic concept of entropy in information theory has to do with how much randomness is in a signal or in a random event. An alternative way to look at this is to talk about how much information is carried by the signal. Entropy is a measure of this randomness. Shannon derives[7] his definition of entropy from the assumptions that:
1-The measure should be continuous - i.e. changing the value of one of the probabilities by a very small amount should only change the entropy by a small amount.
2-If all the outcomes are equally likely then increasing the number of event should always increase the entropy.
3-We should be able to make the procedure in two steps. Entropy of the final result should be a weighted sum of the entropies of the two steps.
If we define a random variable $X=(x_1, x_2, x_3, \ldots, x_n)$ and the probability measure P defined for event x_i as:

$$P= p(x_i) \qquad \ldots\ldots(1)$$

then the information contents of event (x_i) are computed by $[\log_2]$ of reciprocal of the probability measure of $p(x_i)$, which can be written as:

$$K_X(x_i) = \log(1/p(x_i)) \qquad \ldots\ldots(2)$$

where $K_X(x_i)$ is the information contents of event x_i.
It is clear that the information contents $K_x(x_i)$ goes to 1 when the probability measure $p(x_i)$ goes to 0, and it goes to 0 when p(xi) goes to 1. In other words,

$$\text{Lim } K_X(x_i) = 0 \quad \text{as} \quad p(x_i) \rightarrow 0 \qquad \ldots\ldots(3)$$

and

$$\text{Lim } K_X(x_i) = 1 \quad \text{as} \quad p(x_i) \rightarrow 0 \qquad \ldots\ldots(4)$$

Such information for event x_i is increased as $p(x_i)$ decreased, and one can conclude that no more information for an event x_i, if $p(x_i)$ goes to 1. Accordingly, the term entropy (€) which is the key item of our work, related to the average of information per message, and generally, € (X) , can be defined for any discrete random variable X, as a measure of the amount of uncertainty associated with the value of x_i. If X is the set of all messages $\{x_1, \ldots, x_n\}$ that X could be, and P(X) is the probability of given some x_i , then the entropy of X defined as: .

$$€(X) = \sum_{n=1}^{\infty} p(xi) \ \log \ (i/p(xi)) \qquad \ldots\ldots\ldots(5)$$

Similarly, within the context of information theory, self-information is defined as the amount of information that knowledge about (the outcome of) a certain event, adds to someone's overall knowledge. The amount of self-information is expressed in the unit of information: a bit. By definition, the amount of self-information contained in a probabilistic event depends only on the probability p of that event. More specifically: the smaller this probability is, the larger is the self-information associated with receiving information that the event indeed occurred [8].

Further, by definition, the measure of self-information has the following property. If an event *C* is composed of two mutually independent events *A* and *B*, then the amount of information at the announcement that *C* has happened, equals the sum of the amounts of information at announcement of event *A* and event *B* respectively. Taking into account these properties, the self-information $I(A_n)$ (measured in bits which is the expected information) associated with outcome A_n whose outcome has probability p is defined as:

$$I(A_n) = \log_2\left(\frac{1}{p(A_n)}\right) = -\log_2(p(A_n)) \qquad \ldots\ldots\ldots(6)$$

This means that information contents for an event A_n also decrease as p (A_n) increases and no information exist as p (A_n) reaches one. As a sequence of this, and similarly, one can define the entropy € (X) as:

$$€(X) = k_X[I(x)] = -\sum_{x\in X} p(x)\log p(x). \qquad \ldots\ldots\ldots(7)$$

For a given message x_i through x_n, the term entropy is related to the average of information per message. Equation (7) is used to measure the amount of information obtained when the output of the random experiment is noticed and the particular message $\{x_1, \ldots, x_n\}$ is expected to occur.

In this research we will apply: the entropy, fundamentals of information theory, assess the translational, application of fixed points and rotational information contents of patterns, and assessment of the total information contents. This research will show that information theory measures are an important tool for pattern recognition, and can be modified and applied for complicated cases with higher dimension images.

III. THE SIMULATION AND IMPLENTATION

Different measurement can be defined and used as information related to a specific input pattern. One of the most interesting information measures is called the total information. For such measure, it is necessary to look also at the all possible orientations and locations related to such particular input pattern in addition to other measurements called shared measurement and net quantity of information measure. Other measurements can be included as information to the total information for particular input pattern [2]. By such input pattern and measurement, one has to find all possible orientation and location measures.

Define the following:

® be the average rotational information,
Θ be the average orientation information,
Ψ be the average shared information,
Π be the average translational information,
© be the average net quantity of information,
T be the average total information.

The above measures used to assess the information with orientation and location, then to be extended to assess the recognition capability and to increase the efficiency of pattern recognition related to information theory. Since total information (T) includes average net quantity of information (©) and also shared information (Ψ), then one can relate average translational information (Π) to the following measures:

Π = © + Ψ .... (8)

Moving the input pattern image in any direction one pixel, result a configuration output similar in structure and shape, but moved to a new set of pattern displaced one location used to evaluate generally average translational information (Π) [8]. For the purpose of implementation, the average rotational information (®) can be calculated by mean of the sum of average orientation (Θ) and the average shared information (Ψ) as follows:

® = Θ + Ψ ……………………………………………..………(9)

which shows that the basic measurements of the average of total information is related to basic information about average orientation (Θ) and average shared information (Ψ). Since the average total information includes also average net quantity of information (©) and average shared information Ψ, then one can write the following:

® = Θ + Π - ©
where Π is the average translational information given as:
Π = © + Ψ …………………………………..………………….(10)
which relates average transitional Π and average net quantity © of information.
During implementation, three measures are essentials to be computed in order to use it as an important tool for pattern reconditions. These measures to be computed are the orientation, then location of a particular input followed by the positional information. Other information called translational information should be computed too, and it is a result of moving input image in any direction only one pixel similar of output configuration in structure and shape. The output is new set of pattern displacement one location (pixel). Center of gravity of input image should be evaluated and accordingly, positional information obtained directly by knowing the positional of center of gravity. Fixed point iteration application will be used to generate an area which is the product of shift in actual y-direction multiplied by the number of matches found in x-coordinates for that fixed point. Finally, the proposed technique is simply completed by calculating Average rotational information using equation (7).

This technique is proposed by using basics of information theory and its application in image recognition. It is used to determine the amount of information generated by either rotational or translational or total information. Center of gravity of the image is the first element selected and its value should be acceptable value from origin of coordinate system.
The orientation yields more accurate information, and consequently good recognition. Fixed point iteration technique is used for this purpose in implementation. An entropy function is implemented to produce the final output, which is the information of the pattern recognition.
The use of fixed point iteration technique of the functions generated between:
$F(X=\{x_1,x_2,x_3\ldots..,x_n\})=(0,0,0,\ldots\ldots0)$
where X is the set of coordinates $\{x_1, \ldots, x_n\}$ and
$Y\{y_1 ,y_2 ,y_3 ,\ldots..,yn\} =X\{x_1,x_2,\ldots..,x_n\}$ ……………………………….….(11)
yield better understanding of information about the area of the pattern.
In general, we can write the general n-dimensional non-linear fixed point system as:

$$\vec{\mathbf{X}} = \vec{\mathbf{G}}(\vec{\mathbf{X}}) \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(12)$$

given one initial approximation $\vec{\mathbf{P}}_0$, and generating a sequence $\{\vec{\mathbf{P}}_k\}$ which converges to the solution $\vec{\mathbf{P}}$,

$$\vec{\mathbf{G}}(\vec{\mathbf{P}}) = \vec{\mathbf{P}} \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (13)$$

The fixed point is calculated iteratively using the values of pixels at coordinates x_i and y_i after orientation and translation, The result, together with the center of gravity, is shifted or rotational or translational, is used for the purpose of calculating the area of pattern matching and location of new pattern for matching. This area, is the product of shift in actual y-direction multiplied by the number of matches found in x-coordinates for that fixed point at location (x_i,y_i) after orientation and translation. By applying this technique, center of gravity of image shifted from the starting coordinates with initial guess or any guessed value for fixed point iteration technique between center of gravity and the maximum location, is iterated until approximate iterative result obtained. This value is used to calculate the area, which is the expected information for pattern. Different methods can be used, such as binary search, which is unconditionally converges to an approximate root, and gives similar results.

IV. CONCLUSIONS

The previous technique based on the concept of rotational, translational and total information, is introduced and applied on a sample of images representing a set of characters for the purpose of character recognition. This is the basic of information theory and its application in image recognition. Such technique can be further modified to be able to recognize more complicated patterns such as hand written character recognition and three-dimensional face recognition. The application of fixed point iteration and binary search give better results to build area of pattern depending on matching and fetching. Different resolutions are used, namely: $(1/2^n)$, n=3 to any known standard power for such resolution and a 32 * 32 pixel grid. Results obtained for sample of squares of different sizes and for different resolutions. Also, the boundary points of the characters with respect to their center of gravity, and all possible patterns found are defined. Probability of occurrences of each pattern is computed and then translational, orientation using fixed point iteration method, and finally, total information are obtained.

Table 1, shows the results obtained for sample of squares of different sizes and for different resolutions. Using such technique, yields to better understanding and complete determination to the information content of a set of representation of a given input pattern, to allocate different pattern for different classes in order to find the mean of recognition. Results obtained shows that information theory measures and simulation numerical techniques, such as fixed point iterative technique or binary search are an important tool for pattern recognition. Excellent results are obtained in the assessment of information content of patterns. Results of this research will open a new hot topic which at least apply the entropy, fundamentals of information theory, assess the translational, application of fixed points and rotational information contents of patterns, and assessment of the total information contents, in addition to the application of approximate solutions using iterative methods for simulation problems. This research proves that information theory measures are an important tool

for pattern recognition, and can be modified and applied for complicated cases with higher dimension images. Pattern data set in information theory can be in any orientation, twisted and tilted, transferred or others, and information theory can still do such matching and pattern recognition can still do comparisons with adding more information to your model.
In information theory and its applications in pattern, pattern recognition technology easily answers increasingly common, yet critical questions: how good is the information, do differences exist, and where are those differences? It also relies on the proper amount of information.
In conclusion, Information Theory and its Application to Pattern Recognition in Image Processing is practical method for matching information set.

Table 1. The results obtained for sample of squares of different sizes and for different resolutions

Size	Res.	Trans	Rot.	Total
4x2	1/1024	4.5101	3.6212	7.1121
4x4	1/1024	4.5911	3.9212	7.2899
4x2	1/512	4.5991	3.7001	7.2221
4x2	1/128	4.5723	3.6990	7.2212
4x2	1/32	4.5711	3.6691	7.2202
4x2	1/8	4.5770	3.6690	7.2200

REFERENCES

1. Yeh A. G. and Li. Xia,. 2001 "Measurement and Monitoring of Photogrammetric Engineering and Remote Sensing", Vol. 67(1): pp 83.
2. Wu Z. S, Y. and Mumford D., "Filters, Random Fields and Maximum Entropy", 1998. Journal of Computer Vision, 27 (2),
3. I. Talavera, M.D. P., R. Duin P. W. , and Orozco-Alzate M, 2011, "Combining dissimilarities for three-way data classification", Computacióny Sistemas", vol. 15, issue 1, pp. 117-127.
4. Pekalska D. R. P. W. , , 2011,"The Dissimilarity Representation for Structural Pattern Recognition", Progress in Pattern Recognition, Image Analysis, Compute Vision, and Applications - 16th Iberoamerican Congress, {CIARP} 2011, Pucón, Chile, November 15-18. Proceedings, vol. 7042: Springer, pp. 1–24, 2011.
5. Merhav, N., Serouss G. i, and Weinberger M. J.. September 2005."Inform. Theory (ISIT'05)", Proceedings of the 2005 IEEE Intl. Symp. On , Adelaide, Australia.
6. Erik T. E. Ordentlich. " HP Labs Tech. Report HPL", 2005-68, Proceedings of IS&T/SPIE Symp. on Electronic Imaging, Jan. 15-19, 2006.
7. E. Ordentlich, T. Weissman IEEE Intl. "Symp. on Inform. Theory 2005", Proceedings on Infprmation Theory, Sept. 4-9, 2005, pp. 2198 – 2002.
8. LeffH. S. and Rex A. F., |"Entropy, Information, Computing, Princeton University Press", Princeton, NJ (1990). ISBN 069108727X.
9. L. Guan.R. D. Green, "Tracking HumanMovement PatternsUsing Particle Filtering". Proceedings of the 2003 International Conference on Multimedia and Expo, Vol. 3, 117-120, 2006.
10. Sun C., D., Stirling F. ,Naghdy c . "Human Behaviour, Recognition with Segmented Inertial Data". In ARAAA ustralasian Conference on Robotics and Automation, 1–9,2006.
11. Chen N. and D., "A Survey of Document Image Classification: Problem Statement, Classifier Architecture and Performance Evaluation", *Int'l Journal on Document Analysis and Recognition,* Vol. 10, No. 1, June 2007, pp. 1-16
12. D. Blostein and Blostein A. Grbavec, "Recognition of Mathematical Notation," in *Handbook of Character Recognition and Document Image Analysis*, 1997 Eds. H. Bunke and P. Wang, World Scientific, 1997, pp. 557-582.
13. D. Blostein, "General Diagram-Recognition Methodologies," in *Graphics Recognition: Methods and Applications*, Eds. R. Kasturi and K. Tombre, Lecture Notes in Computer Science, Vol. 1072, Springer Verlag, 1996, pp. 106-122.

BOUT THE AUTHOR

Profesor Dr. Yahia ALHalabi Received his Ph.D Degree in Math and Computer Science in 1982, fromClarkson University, Potsdam,NY, USA and his Ms.c degree in Numerical Simulation from the University of Jordan - Hashemite Kingdom of Jordan . Since 1982, he served as professor of computer Science, Chairman of the computer, Director of Computer Center Department at the University of Jordan, Dean Of faculty of Art and Sciences at Amman Private University, Dean of Faculty of Computer Studies at Arab Open University – Kuwait headquarter office.

He joined Princess Sumaya University for Technology in 2005 and continues his research in image processing and simulation. Since September 2008-2010, he is appointed as Dean, King Hussein School for Information Technology at Princess Sumaya University for Technology. He is involved in many scientific organizations, locally and editor for many international research journals and member of accreditation and validation group for higher education in Jordan and abroad.