Entropy Quad-Trees for High Complexity Regions Detection

Rosanne Vetro, University of Massachusetts Boston, USA
Dan A. Simovici, University of Massachusetts Boston, USA
Wei Ding, University of Massachusetts Boston, USA

ABSTRACT

This paper introduces entropy quad-trees, which are structures derived from quad-trees by allowing nodes to split only when those correspond to sufficiently complex sub-domains of a data domain. Complexity is evaluated using an information-theoretic measure based on the analysis of the entropy associated to sets of objects designated by nodes. An alternative measure related to the concept of box-counting dimension is also explored. Experimental results demonstrate the efficiency of entropy quad-trees to mine complex regions. As an application, the proposed technique is used in the initial stage of a crater detection algorithm using digital images taken from the surface of Mars. Additional experimental results are provided that demonstrate the crater detection performance and analyze the effectiveness of entropy quad-trees for high-complexity regions detection in the pixel space with significant presence of noise. This work focuses on 2-dimensional image domains, but can be generalized to higher dimensional data.

Keywords: Box-Counting Dimension, Circular Hough Transform, Entropy, Nodes, Quad-Trees

INTRODUCTION

The concept of complexity relates to the presence of variation. In science there are many approaches that characterize complexity. A variety of scientific fields have dealt with complex mechanisms, simulations, systems, behavior and data complexity as those have always been a part of our environment. In this work, we focus on the topic of data complexity which is studied in information theory. While randomness is not considered complexity in certain areas such as those related to the study of complex systems, information theory tends to assign high values of complexity to random noise. Many fields benefit from the identification of content or noise related complex areas. In data hiding, adaptive steganography takes advantage of high concentration of self-information on high complexity areas originated from both content and noise to embed data. Solanki, Dabeer, Madhow, Manjunath, and Chandrasekaran (2003) describe the benefits of selective embedding related to the reduction of perceptual degradation for transform domain steganographic techniques. Bio diversity is another area where complexity can be used for identification and localization of different species. In this case, the complexity

DOI: 10.4018/jssci.2011010102

Copyright © 2011, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.
originated from content is more important than the one originated from noise.

Our goal in this paper is to introduce a variant of quad-trees for mining high complexity sub-domains of a data domain. A quad-tree is a tree structure defined on a finite set of nodes that either contains no nodes or is comprised of a root node and 4 quad-subtrees. In a full quad-tree, each node is either a leaf or has degree exactly 4. Our variant of quad-trees requires that each node that has descendants corresponds to a region that has a sufficient level of diversity as assessed by the value of an information-theoretical measure. We also present an alternative measure that has its roots in fractal geometry where the so-called box-counting dimension (BCD) is used to determine the fractal dimension of a set $S$ in a Euclidean space $\mathbb{R}^n$. We then provide an algorithm to capture high complexity areas of 2-dimensional images domains and observe that diversity originated from both data content and noise are mined.

As an application, we used our proposed data structure in the initial stage of a crater detection algorithm. The algorithm is composed by two methods. The first method uses an information-theoretical approach with entropy quad-trees to create an edge filter that generates a binary image from complex areas which may contain edges. The second method applies a Circle Hough Transform (CHT) with modified threshold to detect the presence of circular shapes in complex areas. The new threshold is imposed to increase the quality of the results given the lack of prior knowledge about the number of craters in an image and the difficulty to estimate a good threshold for the minimum number of votes required in the parameter space to indicate true center points. Efficient methods for crater detection have been proposed (Bue & Stepinksi, 2007; Urbach & Stepinksi, 2009; Salamunicar & Loncaric, 2008). We provide a distinct approach where no external pre-processing of the original image other than conversion to the JPEG format and resizing is needed. Likewise, no external image filters are used.

In the next section we introduce a framework for the rest of the paper. The notion of entropy associated to a partition is presented as well as its usefulness in measuring diversity. In section 2 we introduce the proposed data structures, an algorithm for high complexity detection and explain the searching process. In subsection 2.1 we describe the information-theoretic method used for mining complex sub-domains. An alternative method uses the concept of box-counting dimension and is introduced in subsection 2.2. We provide a brief description about implementation details in subsection 2.3. In subsection 2.4 we discuss the experiments and compare the results generated by both methods. In Section 3 we introduce the crater detection algorithm. In subsection 3.1 we describe the information theoretic method used for mining complex subareas that may contain edges. The CHT method with modified threshold is described in subsection 3.2. Subsection 3.3 contains a description of the experiments and major challenges we faced. Finally, Section 4 contains our conclusions and ideas for future work.

1. PARTITIONS, ENTROPY, AND TREES

The notion of entropy quantifies the uncertainty associated with probability distributions. Let $S$ be a finite set. A partition on $S$ is a non-empty collection of non-empty subsets of $S$, $\pi = \{B_1,\ldots,B_n\}$ such that:

(i) $B_i \cap B_j = \emptyset$ for $1 \leq i, j \leq n$ and $i \neq j$;
(ii) $\bigcup \{B_i \mid 1 \leq i \leq n\} = S$.

The sets $B_1,\ldots,B_n$ are referred to as the blocks of $\pi$.

We denote by Part($S$) the set of partitions of $S$. For $\pi, \sigma \in$ Part($S$) define $\pi \geq \sigma$ if each
Formal Rules for Fuzzy Causal Analyses and Fuzzy Inferences
[www.igi-global.com/article/formal-rules-for-fuzzy-causal-analyses-and-fuzzy-inferences/88928?camid=4v1a](www.igi-global.com/article/formal-rules-for-fuzzy-causal-analyses-and-fuzzy-inferences/88928?camid=4v1a)

Fuzzy-Controlled Energy-Conservation Technique (FET) for Mobile ad hoc Networks
[www.igi-global.com/chapter/fuzzy-controlled-energy-conservation-technique/72507?camid=4v1a](www.igi-global.com/chapter/fuzzy-controlled-energy-conservation-technique/72507?camid=4v1a)