Summarizing Changes in Image Sequences Using Algorithmic Information Theory

by

Andrew R. Cohen

An Abstract of a Thesis Submitted to the Graduate

Faculty of Rensselaer Polytechnic Institute

in Partial Fulfillment of the

Requirements for the degree of

DOCTOR OF PHILOSOPHY

Major Subject: Electrical, Computer and Systems Engineering

The original of the complete thesis is on file
In the Rensselaer Polytechnic Institute Library

Examining Committee:

Prof. Badrinath Roysam, Thesis Adviser

Prof. Richard J. Radke, Member

Prof. Arthur C. Sanderson, Member

Prof. Deanna Thompson, Member

Prof. Mohammed J. Zaki, Member

Rensselaer Polytechnic Institute
Troy, New York

March, 2008
(For Graduation May, 2008)

ABSTRACT

This thesis presents a broadly applicable methodology based on algorithmic information theory and algorithmic statistics for generating a *concise and meaningful* summary of the changes occurring *within and across image sequences*. The notion of concise and meaningful summarization is captured by using an algorithmic information theoretic distance measure in conjunction with the gap statistic to estimate the randomness deficiency from algorithmic statistics. This is the first known technique for approximating the randomness deficiency; the lack of practical approximations for randomness deficiency have to date prevented any practical applications of algorithmic statistics. The methodology assumes the availability of object extraction and tracking algorithms. Beyond these requirements, it is parameter free and requires no training data. Our formulation is novel in its use of relative Kolmogorov complexity in the form of an algorithmic information theoretic distance capable of quantifying *any and all differences* between digital objects. The proposed Normalized Adaptive Information Distance (NAID) measure, is unique in its ability to compare vector or time-series data, including data of different dimensions. The image data is analyzed at multiple levels of precision using the proposed closed-form automatic multi-dimensional quantization algorithm. We propose a closed-form method for m ultidimensional quantization of numeric data capable of assigning any number of symbols to data of any dimension, such that the symbols are equiprobable under the assumption of normally distributed data. This dramatically improves the previously published method of Symbolic Aggregate Approximation (SAX), a widely used quantization technique developed in data mining community

The proposed method was evaluated on sets of image sequences from seven different applications domains from cell and tissue biology using a single implementation. For some of these data sets we reproduced and extended previously published results. In other cases *we discovered previously unknown behavioral differences which corresponded to biologically significant differences* in populations.