



ELSEVIER

Available at  
www.ComputerScienceWeb.com  
POWERED BY SCIENCE @ DIRECT®

Pattern Recognition Letters 24 (2003) 1081–1087

Pattern Recognition  
Letters

www.elsevier.com/locate/patrec

# Using attributed plex grammars for the generation of image and graph databases

Markus Hagenbuchner <sup>a,\*</sup>, Marco Gori <sup>b</sup>, Horst Bunke <sup>c</sup>, Ah Chung Tsoi <sup>a</sup>,  
Christophe Irniger <sup>c</sup>

<sup>a</sup> Faculty of Informatics, University of Wollongong, Wollongong, 2500 Australia

<sup>b</sup> Dip. di Ingegneria dell'Informazione, Università di Siena, Siena, 53100 Italy

<sup>c</sup> Institute of Computer Science and Mathematics, Universität Bern, Bern, 3012 Switzerland

## Abstract

In this paper, a methodology for the generation of benchmarks in pattern recognition is described. The patterns are represented by means of an attributed plex language, which are based on plex grammars augmented by attributes. It is shown that the generated patterns are particularly suitable for the extraction of graph-based representations. As a result, databases of artificial pictures and correspondent graphs can be generated. These collections of graphs are very appropriate for benchmarks in the area of structural pattern recognition, since they are originated from a grammar and not from random distributions. The tools for creating the databases are public domain and have been already used for benchmarking artificial neural networks operating on structured domains.

© 2002 Elsevier Science B.V. All rights reserved.

*Keywords:* Benchmarking; Attribute grammars; Graph-based representations; Graphical databases; Structural pattern recognition; Plex grammars

## 1. Introduction

The acquisition and distribution of standard databases has become an important issue in pattern recognition. On the one hand, large amounts of training data are needed in many applications to build reliable and accurate systems. On the other

hand, large testing sets widely available in the scientific community are important to evaluate system performance, and to compare different approaches. Examples of well known and widely used databases include NIST (Wilkinson et al., 1992) and CEDAR (Hull, 1994) for off-line handwriting recognition, UNIPEN (Guyon et al., 1994) for on-line handwriting recognition, (Kim et al., 1993; Saito et al., 1985) for oriental character recognition, the IAM-database (Marti and Bunke, 1999) for handwritten English sentence recognition, (Achermann and Bunke, 2000) for face recognition, and (Hoover et al., 1996) for range image segmentation.

\* Corresponding author.

*E-mail addresses:* markus@uow.edu.au, markus@artificial-neural.net (M. Hagenbuchner), marco@ing.unisi.it (M. Gori), bunke@iam.unibe.ch (H. Bunke), tsoi@artificial-neural.net (A.C. Tsoi), irniger@iam.unibe.ch (C. Irniger).