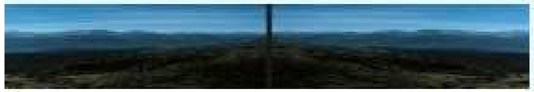
Project acronym	SIMBAD
Project full title	Beyond Features: Similarity-Based Pattern Analysis and Recognition
Deliverable Responsible	Dipartimento di Informatica Università Ca' Foscari Venezia Via Torino 155 30172 Venezia Mestre Italy http://www.dsi.unive.it/
Project web site	http://simbad-fp7.eu
EC project officer	Teresa De Martino
Document title	Project leaflet
Deliverable n.	D8.3
Document type	Other
Dissemination level	Public
Contractual date of delivery	M 7
Project reference number	213250
Status & version	Definitive
Work package	WP 8
Deliverable responsible (SHORT NAME)	UNIVE
Contributing Partners (SHORT NAME)	UNIVE, UNIYORK, TUD, IST, UNIVR, ETH Zurich
Author(s)	Marcello Pelillo
Additional contributor(s)	Veronica Giove

Starting date: 01 April 2008

Duration: 36 months

http://simbad-fp7.eu

contact: info@simbad-fp7.eu



Università Ca' Foscari Venezia - Italy
http://www.dsi.unive.it

Marcello Pelillo - pelillo@dsi.unive.it

University of York - United Kingdom
http://www.york.ac.uk

Edwin Robert Hancock - erh@cs.york.ac.uk

Technische Universiteit Delft - The Netherlands
http://ict.ewi.tudelft.nl/

Robert Duin - r.duin@ieee.org

Instituto Superior Técnico - Portugal
http://www.ist.utl.pt

Mário Figueiredo - mario.figueiredo@lx.it.pt

Università degli Studi Di Verona – Italy
http://vips.sci.univr.it

Vittorio Murino - vittorio.murino@univr.it

Eidgenössische Technische
Hochschule Zürich - Switzerland
http://www.ethz.ch

Joachim Buhmann - jbuhmann@inf.ethz.ch

COMMISSION OF THE EUROPEAN COMMUNITIES

INFORMATION SOCIETY AND MEDIA
DIRECTORATE-GENERAL

Seventh Framework Programme
Information and Communication Technologies

Collaborative Project

FET Open

SIMBAD

Beyond Features

Similarity-based pattern analysis and recognition

BEYOND FEATURES

Traditional pattern recognition techniques are centered on the notion of "feature". According to this view, each object is described in terms of a vector of numerical attributes and is therefore mapped to a point in a **Euclidean (geometric)** vector space so that the distances between the points reflect the observed (dis)similarities between the respective objects.

Despite its potential, the geometric approach suffers from a major intrinsic limitation, which concerns the representational power of feature-based descriptions. In fact, there are numerous application domains where either it is not possible to find satisfactory features or they are inefficient for learning purposes. Most commonly, this is typically the case when objects are described in terms of **structural properties**, such as parts and relations between parts, as is the case in shape recognition.

This project aims at bringing to full maturation a paradigm shift that is currently just emerging within the pattern recognition and machine learning domains, where researchers are becoming increasingly aware of the importance of **similarity information** *per se*, as opposed to the classical feature-based approach. Indeed, the notion of similarity (which appears under different names such as proximity, resemblance, and psychological distance) has long been recognized to lie at the very heart of human cognitive processes and can be considered as a connection between perception and higher-level knowledge, a crucial factor in the process of human recognition and categorization.

WHY SIMBAD?

By departing from vector-space representations one is confronted with the challenging problem of dealing with (dis)similarities that do not necessarily possess the Euclidean behavior or not even obey the requirements of a metric. The lack of the Euclidean and/or metric properties undermines the very foundations of traditional pattern recognition theories and algorithms, and poses totally new theoretical/computational questions and challenges.

We aim at undertaking a thorough study of several aspects of similarity-based pattern analysis and recognition methods, from the theoretical, computational, and applicative perspective, with a view to substantially advance the state of the art in the field, and contribute towards the long-term goal of organizing this emerging field into a more coherent whole.

The whole project will revolve around two main themes, which basically correspond to the two fundamental questions that arise when abandoning the realm of feature-based representations:

1. How can one **obtain** suitable similarity information from object representations that are more powerful than, or simply different from, the vectorial?
2. How can one **use** similarity information in order to perform learning and classification tasks?

According to this perspective, the very notion of similarity becomes the pivot of non-vectorial pattern recognition in the same way as the notion of feature-vector plays the role of the pivot in the classical (geometric) paradigm.

SIMBAD AT A GLANCE

From a methodological perspective, SIMBAD will be structured around the following strands:

Deriving similarities for non-vectorial data (structural kernels), to develop computational models for generating similarities for non-vectorial data, with particular emphasis on structured and semi-structured descriptions

Foundations of non-(geo)metric similarities, to study both the causes of the lack of (geo)metricity in the similarity data and its effects over traditional machine learning algorithms

Imposing geometricity on non-geometric similarities (embedding), to develop algorithms for transforming the original similarity data into proper vectorial representations suitable for traditional learning algorithms

Learning with non-(geo)metric similarities, to develop novel, general learning models which do not require the (geo)metric assumption

An important part of SIMBAD will concern the validation of the developed techniques, focusing mainly on biomedical problems.

Analysis of tissue micro-array (TMA) images of renal cell carcinoma, to validate the techniques developed by applying them to the analysis of Tissue Micro Array (TMA) images of renal cell carcinoma

Analysis of brain magnetic resonance (MR) scans for the diagnosis of mental illness, to validate the techniques developed by applying them to the analysis of brain MR scans in the context of mental health research