Similarity-Based Pattern Analysis and Recognition

Edwin R. Hancock
University of York, UK

Vittorio Murino
IIT & University of Verona, Italy

Marcello Pelillo
University of Venice, Italy

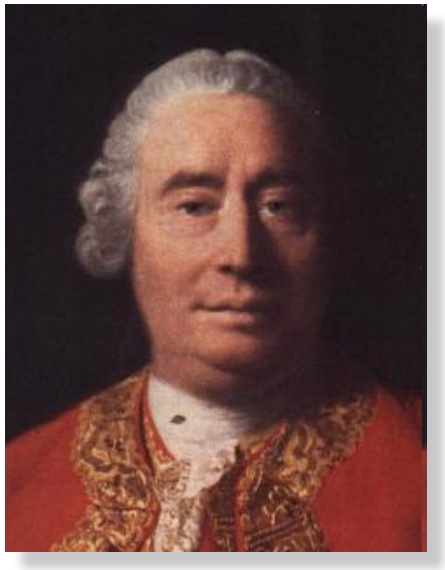
Richard Wilson
University of York, UK

Tutorial at *ECCV 2012*, Florence, Italy, 7 October 2012

http://simbad-fp7.eu

Pattern Recognition and Hume's Similarity Principle

« I have found that such an object has always been attended with such an effect, and I foresee, that other objects, which are, in appearance, similar, will be attended with similar effects. »



David Hume

An Enquiry Concerning Human Understanding

(1748)

See also the "homophily" principle in social network analysis.

The Classical "Feature-based" Approach and Its Limitations

Traditional pattern recognition techniques are centered on the notion of **feature-vector,** i.e. they *derive similarities from vector representations*.

But, there are variuos application domains where either it is not possible to find satisfactory feature vectors or they are inefficient for learning purposes.

This is typically the case, e.g.,

- ✓ when experts cannot define features in a straightforward way
- ✓ when data are high dimensional
- ✓ when features consist of both numerical and categorical variables,
- ✓ in the presence of missing or inhomogeneous data
- ✓ when objects are described in terms of structural properties, such as parts and relations between parts, as is the case in shape recognition

Beyond Features?

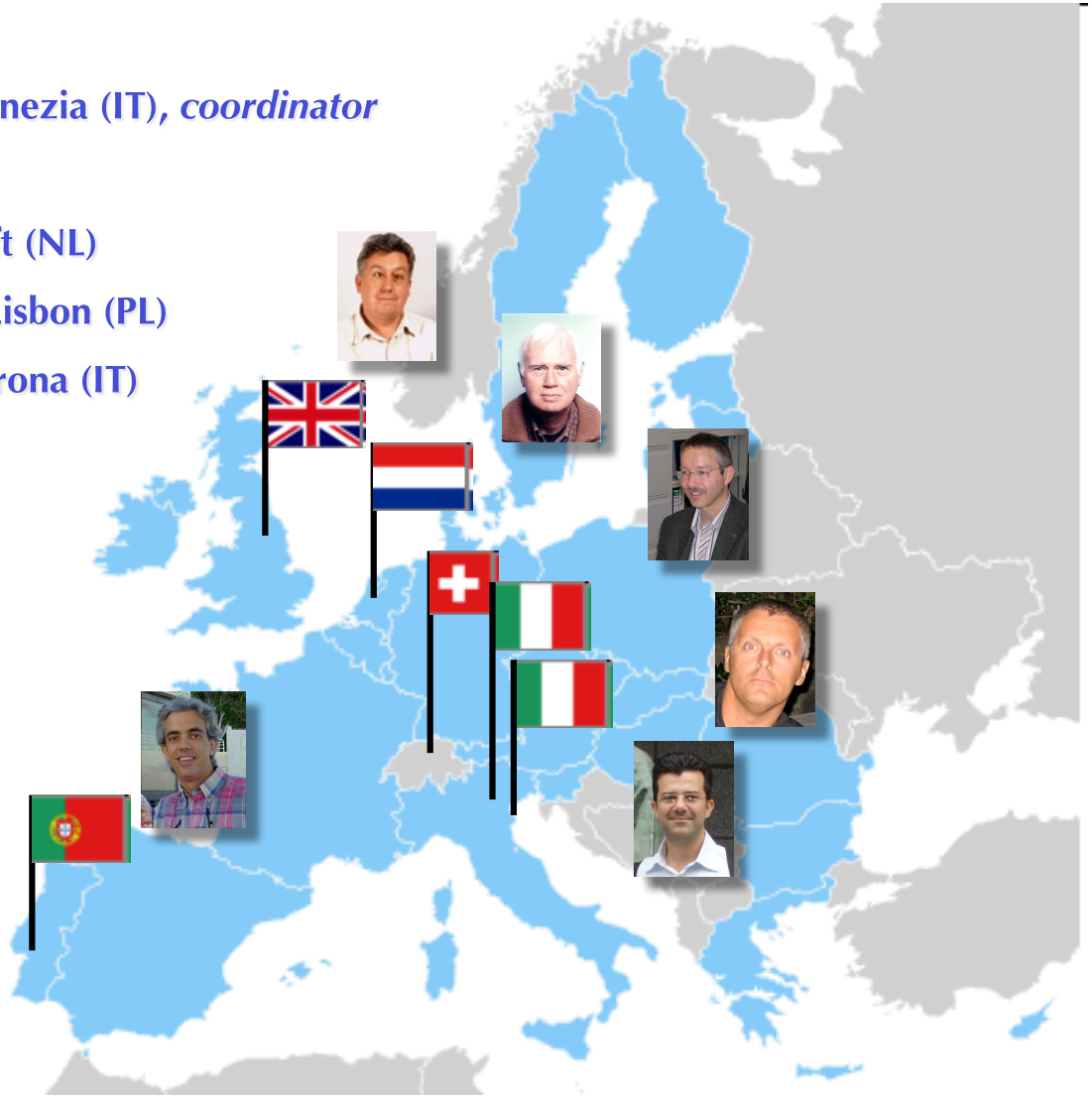
By departing from vector-space representations one is confronted with the challenging problem of dealing with (dis)similarities that do not necessarily possess the Euclidean behavior or not even obey the requirements of a metric.

The lack of the Euclidean and/or metric properties undermines the very foundations of traditional pattern recognition theories and algorithms!

The SIMBAD FP7 Project

Beyond Features:
Similarity-Based Pattern Analysis and Recognition

1. **Università Ca' Foscari di Venezia (IT), *coordinator***
2. **University of York (UK)**
3. **Technische Universität Delft (NL)**
4. **Insituto Superior Técnico, Lisbon (PL)**
5. **Università degli Studi di Verona (IT)**
6. **ETH Zürich (CH)**



Type: *FP7 FET-Open*

Starting date: April 2008

Closing date: September 2011

http://simbad-fp7.eu

Objectives of SIMBAD

SIMBAD 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 (see, e.g., the explosion of kernel methods).

The whole project revolves around two main themes, which basically correspond to the two fundamental questions that arise when abandoning the realm of vectorial representations, namely:

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

The Structure of SIMBAD

1. **Deriving similarities for non-vectorial data**
   - *Structural (generative/compression) kernels*
   - *Learning and combining similarities*

2. **Learning and classification with non-(geo)metric similarities**
   - *Foundations of non-(geo)metric similarities*
   - *Imposing geometricity on non-geometric similarities (embedding)*
   - *Learning with non-(geo)metric similarities (game theory)*

3. **Biomedical applications**
   - *Analysis of tissue micro-array (TMA) images of renal cell carcinoma*
   - *Analysis of brain magnetic resonance (MR) scans for the diagnosis of mental illness*

For more information:

http://simbad-fp7.eu

The SIMBAD Book

M. Pelillo (Ed.)

Similarity-Based Pattern Analysis and Recognition

2013 (to appear)

Springer

The ICML Workshop

Haifa, Israel, 21 July 2010

HAIFA ICML ISRAEL 2010

International Conference on Machine Learning

Haifa, Israel
June 21 - 24

Learning in Non-(geo)metric Spaces - Workshop @ ICML 2010

Invited speakers

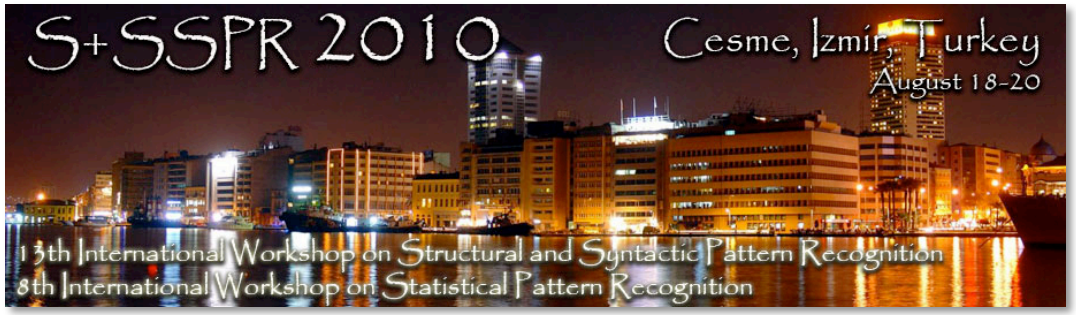
S. Ben David, U. von Luxburg
N. Balcan, A. Smola, N. Cesa-Bianchi,
R. Krauthgamer, A. Shashua

PASCAL2
Pattern Analysis, Statistical Modelling and Computational Learning

videolectures.net
exchange ideas & share knowledge

The S+SSPR Special Session

Cesme, Turkey, 25 August 2010



Special Session:

SIMILARITY-BASED PATTERN RECOGNITION: CHALLENGES AND PROSPECTS

Invited speakers:

N. Ahuja, E. Estrada, F. Porikili,
H. Bunke, J. Kitller

videolectures.net
exchange ideas & share knowledge

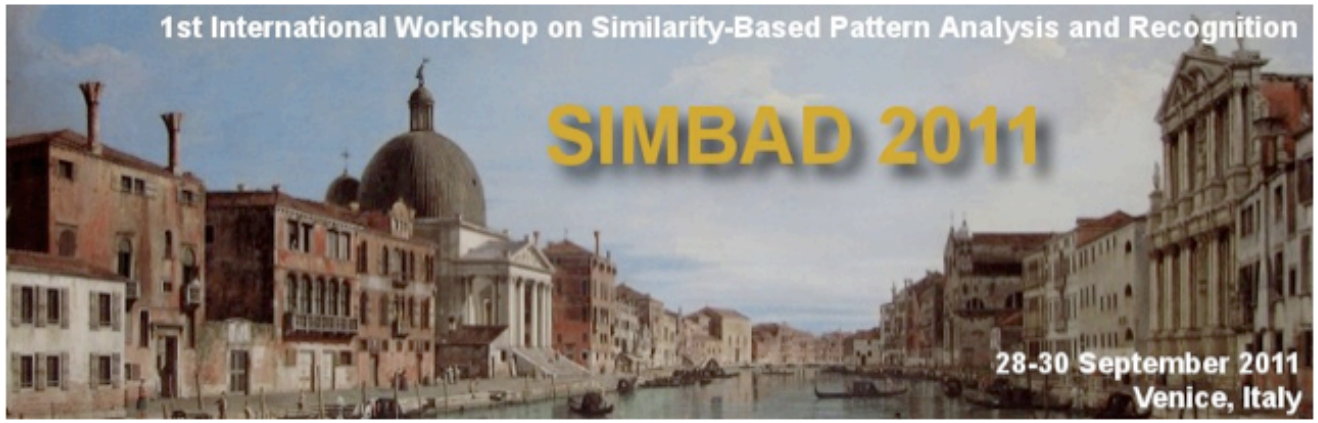
Sponsored by

SIMBAD

SIMBAD 2011

Venice, Italy, 28-30 September 2011

http://www.dais.unive.it/~simbad/2011/



- Home
- Motivations and objectives
- Format
- Venue
- Organization
- Important dates
- Paper submission
- Program
- Registration
- Social events
- Travel and accommodation

The aim of this workshop is to consolidate research efforts in the area of similarity-based pattern recognition and machine learning and to provide an informal discussion forum for researchers and practitioners interested in this important yet diverse subject.

We aim at covering a wide range of problems and perspectives, from supervised to unsupervised learning, from generative to discriminative models, and from theoretical issues to real-world practical applications.

The workshop will mark the end of the **EU FP7 Projects SIMBAD** and is a follow-up of the **ICML 2010 Workshop on Learning in non-(geo)metric spaces**.

videolectures.net
exchange ideas & share knowledge

SIMBAD

IAPR

Università Ca' Foscari
Venezia

Lecture Notes in Computer Science
LNCS LNAI LNBI

PASCAL2
Pattern Analysis, Statistical Modelling and Computational Learning

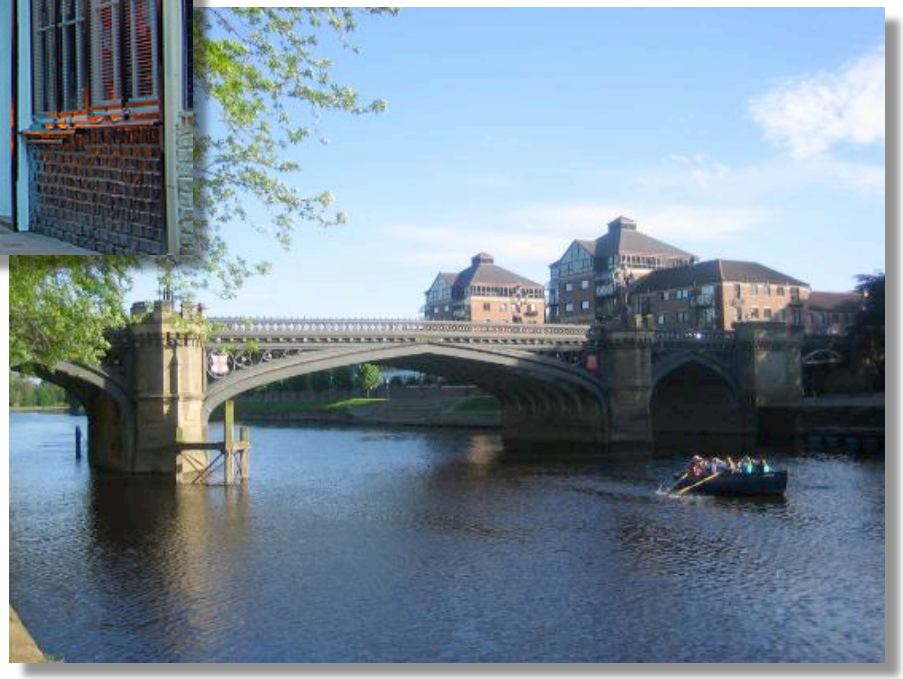
SIMBAD 2013

York, UK, July 3-5 2013

http://www.dais.unive.it/~simbad

IAPR

Lecture Notes in Computer Science
LNCS LNAI LNBI



Submission Deadline:

>>> February 1, 2013 <<<

Course Outline

09.30 – 11.00: Introduction to similarity-based pattern recognition (*R. Wilson*)

Vector-space, distance and similarity; Euclidean embedding techniques (standard methods, MDS, etc); Non-Euclidean data (causes, tests, corrections); Non-Euclidean embedding techniques (spherical embeddings).

Coffee break

11.30 – 13.00: Structure, complexity and learning (*E. R Hancock*)

Structure preserving embeddings (random walks, zeta functions, path and cycle based methods); Complexity level characterizations of relational structures; Algorithms on embedded dissimilarity data (graph and hypergraph matching, clustering and feature selection).

Lunch break

14.30 – 16.00: Learning with non-(geo)metric similarities (*M. Pelillo*)

Game-theoretic models of pattern recognition; Evolutionary games and data clustering; Polymatrix games and contextual pattern recognition.

Coffee break

16.30 – 18.15: Similarity-based techniques for medical image analysis (*V. Murino*)

The Classic approach; The similarity-based approach; A Case study: schizophrenia analysis by MRI brain classification.