Project acronym	SIMBAD
Project full title	Beyond Features: Similarity-Based Pattern Analysis and Recognition
Deliverable Responsible	Dipartimento di Informatica Università Ca' Foscari di Venezia Via Torino 155 30172 Venezia Mestre Italy http://www.dsi.unive.it/~pelillo
Project web site	http://simbad-fp7.eu
EC project officer	Teresa De Martino
Document title	Final project report
Deliverable n.	D1.6
Document type	Report
Dissemination level	Public
Contractual date of delivery	36
Project reference number	213250
Status & version	Final version
Work package	WP1
Deliverable responsible (SHORT NAME)	UNIVE
Contributing Partners (SHORT NAME)	
Author(s)	Marcello Pelillo
Additional contributor(s)	

SIMBAD

Deliverable D1.6

PROJECT FINAL REPORT

26 September 2011

Grant Agreement number: 213250

Project acronym: SIMBAD

Project title: Beyond Features: Similarity-Based Pattern Analysis and Recognition

Funding Scheme: Collaborative Project

Period covered: from 1/4/2008 to 30/9/2011

Name, title and organisation of the scientific representative of the project's coordinator:
Marcello Pelillo, Professor of Computer Science, Università Ca' Foscari Venezia, Italy

Tel: +39 041 2348440

Fax: +39 041 2348 419

E-mail: pelillo@dsi.unive.it

Project website7 address: http://simbad-fp7.eu

4.1 Final publishable summary report

4.1.1 Executive summary

Traditional pattern recognition techniques are centered around the notion of "feature". According to this view, the objects to be classified are represented in terms of properties that are intrinsic to the object itself. Hence, a typical pattern recognition system makes its decisions by simply looking at one or more feature vectors provided as input. The strength of this approach is that it can leverage a wide range of mathematical tools ranging from statistics, to geometry, to optimization. However, in many real-world applications a feasible feature-based description of objects might be difficult to obtain or inefficient for learning purposes. In these cases, it is often possible to obtain a measure of the (dis)similarity of the objects to be classified, and in some applications the use of dissimilarities (rather than features) makes the problem more viable.

In the last few years, researchers in pattern recognition and machine learning are becoming increasingly aware of the importance of similarity information per se. Indeed, by abandoning the realm of vectorial representations one is confronted with the challenging problem of dealing with (dis)similarities that do not necessarily obey the requirements of a metric. This undermines the very foundations of traditional pattern recognition theories and algorithms, and poses totally new theoretical and computational questions. In this project we aim at undertaking a thorough study of several aspects of purely similarity-based pattern analysis and recognition methods, from the theoretical, computational, and applicative perspective. We aim at covering a wide range of problems and perspectives. We shall consider both supervised and unsupervised learning paradigms, generative and discriminative models, and our interest will range from purely theoretical problems to real-world practical applications.

4.1.2 Summary description of project context and objectives

The challenge of automatic pattern analysis and recognition (or machine learning) is to develop computational methods which learn, from examples, to distinguish among a number of classes, with a view to endow artificial systems with the ability to improve their own performance in the light of new external stimuli. This ability is widely recognized to be instrumental in building next-generation artificial cognitive systems (ACS's) which, as opposed to traditional machine or computer systems, can be characterized "as systems which cope with *novel* or *indeterminate* situations, which aim to achieve *general goals* as opposed to solving specific problems, and which integrate *capabilities* normally associated with people or animals" (from: *Artificial Cognitive Systems in FP7: A Report on Expert Consultations for the EU Seventh Framework Programme 2007-2013 for Research and Technology Development*). The socio-economic implications of this scientific endeavor are enormous, as ACS's will have applications in a wide variety of real-world scenarios ranging from industrial manufacturing to vehicle control and traffic safety, to remote and on-site (environmental) sensing and monitoring, and to medical diagnostics and therapeutics.

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 (or vectorial) 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.

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. This kind of representation is attractive because geometric spaces offer powerful analytical as well as computational tools that are simply not available in other representations. Indeed, classical pattern recognition methods are tightly related to geometrical concepts and numerous powerful tools have been developed during the last few decades, starting from linear discriminant analysis in the 1920's, to perceptrons in the 1960's, to kernel machines in the 1990's.

The geometric approach suffers from a major intrinsic limitation, which concerns the representational power of vectorial, 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. This is typically the case when experts cannot define features in a straightforward way, when data are high dimensional, when features consist of both numerical and categorical variables, and in the presence of missing or inhomogeneous data. But, probably, this situation arises most commonly when objects are described in terms of structural properties, such as parts and relations between parts, as is the case in shape recognition. This led in 1960's to the development of the structural pattern recognition approach, which uses symbolic data structures, such as strings, trees, and graphs for the representation of individual patterns, thereby, reformulating the recognition problem as a pattern-matching problem.

In the last few years, interest around purely similarity-based techniques has grown considerably. For example, within the supervised learning paradigm (where expert-labeled training data is assumed to be available) the now famous "kernel trick" shifts the focus from the choice of an appropriate set of features to the choice of a suitable kernel, which is related to object similarities. However, this shift of focus is only partial, as the classical interpretation of the notion of a kernel is that it provides an implicit transformation of the feature space rather than a purely similarity-based representation. Similarly, in the unsupervised domain, there has been an increasing interest around pairwise algorithms, such as spectral and graph-theoretic clustering methods, which avoid the use of features altogether.

Despite its potential, presently the similarity-based approach is far from seriously challenging the traditional paradigm. This is due mainly to the sporadicity and heterogeneity of the techniques proposed so far and the lack of a unifying perspective. On the other hand, classical approaches are inherently unable to deal satisfactorily with the complexity and richness arising in many real-world situations. This state of affairs hinders the application of machine learning techniques to a whole variety of relevant, real-world problems. Hence, progress in similarity-based approaches will surely be beneficial for machine learning as a whole and, consequently, for the long-term enterprise of building ACS's.

However, 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[1] 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 that we shall endeavor to address with this project. In fact, this situation arises frequently in practice. For example, non-Euclidean or non-metric (dis)similarity measures are naturally derived when images, shapes or sequences are aligned in a template matching process. In computer vision, non-metric measures are preferred in the presence of partially occluded objects. Other non-metric examples include pairwise structural alignments of proteins that focus on local similarity, variants of Hausdorff distance,

[1] A set of distances D is said to be *Euclidean* (or *geometric*) if there exists a configuration of points in some Euclidean space whose interpoint distances are given by D.

normalized edit-distances, and also some probabilistic measures such as the Kullback-Leibler divergence. Asis well known, the violation of the triangle inequality is often not an artifact of poor choice of features or algorithms, and it is inherent in the problem of robust matching when different parts of objects (shapes) are matched to different images. The same argument may hold for any type of local alignments. Corrections or simplifications may therefore destroy essential information.

With this project, we undertook 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 revolved around two main themes, which basically correspond to the two fundamental questions that arise when abandoning the realm of vectorial, feature-based 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?

Although the two issues are clearly interrelated, it was advantageous to keep them apart as this allows one to separate the *similarity generation* process (a data modeling issue) from the *learning and classification* processes (a task modeling issue). According to this perspective, the very notion of similarity becomes the pivot of non-vectorial pattern recognition in much the same way as the notion of feature-vector plays the role of the pivot in the classical (geometric) paradigm. This results in a useful modularity, which means that all interactions between the object representation and the learning algorithm are mediated by the similarities, which is where the domain knowledge comes into the scene.

As for the first question, we devised suitable similarity measures for non-vectorial data, specifically tailored to a given task. We focused primarily on structured data (e.g., graphs), because of their expressive power and ubiquity, and on geometric measures as they allow one to employ the whole arsenal of powerful techniques available in the (classical) pattern recognition literature. We have also explored an alternative to this “tailoring” approach, which consists in learning similarities directly from training data. As concerns the second question, we addressed both foundational issues related to similarity information and developed practical similarity-based algorithms that do not depend on the actual object representation. In particular, as concerns the latter objective, we distinguished between the situation where the informational content associated with the violation of the geometric properties is limited, or is simply an artifact of the measurement process, and that where this is not the case. This distinction is important as, depending on the actual situation, two complementary strategies can be pursued: the first attempts to impose geometricity by somehow transforming or re-interpreting the similarity data, the second does not and works directly on the original similarities.

An important part of the project concerned the validation of the developed techniques. To this end, we focused mainly on biomedical problems, which lend themselves particularly well to similarity-based approaches. Specifically, we applied the new methods developed within the project to inference tasks in the field of medical image analysis, i.e., to Tissue Micro Array (TMA) analysis and to Magnetic Resonance (MR) brain imaging. We analyzed the effectiveness of different approaches and critically assessed their individual advantages and shortcomings when compared to existing methods.

4.1.3. Description of the main S&T results/foregrounds

In this project we aimed at advancing the state of the art in similarity-based pattern analysis and recognition from the theoretical, computational, and applicative perspective. As outlined in the previous section, the project revolved around two main themes, which concerns the issues of how to obtain suitable similarity information from non-vectorial representations, and how to use them, irrespective of the way in which they are obtained. In addition to these two basic themes, a third one was addressed which concerns the validation of the proposed techniques and their applicability to real-world problems. These were used to quantitatively evaluate the success of the proposed research on large-scale applications with clear social impact.

Accordingly, the project was structured around the following three strands.

1. Deriving similarities for non-vectorial data. The goal here was to develop kernels and more general similarity measures for non-vectorial data. We focused primarily on structured data (e.g., graphs), because of their expressive power and ubiquity, and on geometric measures as they allow one to employ the whole arsenal of powerful techniques available in the geometric pattern recognition literature. We pursued our goal by developing suitable Mercer kernels, which are known to be in correspondence with geometric (dis)similarities. In particular, during the course of the project we worked on:

- Generative kernels
- Compression kernels
- Learning and combining similarities

2. Learning and classification with non-(geo)metric similarities. Within this research strand we aim at both addressing foundational issues related to similarity information and developing practical algorithms that do not depend on the actual object representation. In particular, as concerns the latter objective, we have distinguished two cases, which in turn lead to two complementary approaches. On the one hand, we have considered the case where the informational content of non-(geo)metricity is limited or caused by measurement error. In this case it is a plausible strategy to perform some correction on the similarity data in an attempt to impose (geo)metricity, and then use classical geometric techniques. On the other hand, when the information content of non-(geo)metricity is relevant one needs brand new tools, as standard techniques would not work in this case. More specifically, the activity within these themes has been organized around three main areas:

- Foundations of non (geo)metric similarities.
- Imposing geometricity on non-geometric similarities (embedding).
- Learning with non-(geo)metric similarities.

3. Validation. Given the heterogeneity of different approaches that we are pursuing in this project, it is of particular importance to build a real-world testbed that specifically addresses the various difficulties involved with non-metric data. To this end, we focused on biomedical datasets that nicely combine high practical relevance of the underlying learning tasks and intrinsically non-metric dissimilarity data. In particular, we applied our algorithms to (i) the analysis of Tissue Micro Array (TMA) images of renal cell carcinoma (RCC) and (ii) the analysis of brain magnetic resonance (MR) images in the context of mental health research (e.g., schizophrenia).

Following the outline set forth above, the three major themes form the basis for structuring the project's work plan into coherent work packages (WP's). Specifically, WP2 covered the topics of deriving similarities for non-vectorial data (theme 1), while the second theme concerning learning and classification with non-(geo)metric similarities was addressed by WP3, WP4 and WP5. Finally, the validation phase was undertaken in WP6 and WP7. An additional work package (WP1) dealt with management issues, while WP8 dealt with dissemination strategies.

More specifically, as far as the scientific part is concerned, the project was articulated in the following work packages:

WP2. Deriving similarities for non-vectorial data (structural kernels)
WP3. Foundations of non-(geo)metric similarities
WP4. Imposing geometricity on non-geometric similarities (embedding)
WP5. Learning with non-(geo)metric similarities
WP6. Analysis of tissue micro-array (TMA) images of renal cell carcinoma
WP7. Analysis of brain magnetic resonance (MR) scans for the diagnosis of mental illness

The results achieved within the various work packages are described in the corresponding final reports: D2.4 (for WP2), D3.4 (for WP3), D4.4 (for WP4), D5.3 (for WP3), D6.2 (for WP6), and D7.2 (for WP7).

4.1.4 Impact and main dissemination activities

The increasing complexity of our society and economy places great emphasis on artificial systems such as robots, smart devices and machines which can deal autonomously with our needs and with the peculiarities of the environments we inhabit and construct. Accordingly, the goal of building artificial cognitive systems (ACS's) has become a worldwide research challenge, and one of the seven key challenges that the European Commission has proposed in order for Europe to be among the world leaders in next-generation information and communication technologies (ICT) and their applications (challenge 2 of work programme 2007-08: "Cognitive systems, interaction, robotics"). Research in this area will in fact enable significant progress in many key applications domains with relevant economic and social impact such as, for example, robotics and other types of assistive devices, human-machine interaction, vehicle control and traffic safety, management and control of transport, energy and communication networks, remote and on-site (environmental) sensing and monitoring, and medical diagnostics and therapeutics, thereby contributing "to improve the competitiveness of European industry – as well as to enable Europe to master and shape the future developments of these technologies so that the demands of its society and economy are met," which is the main objective of ICT research under the EU's Seventh Framework Programme (FP7).

With the FP7 programme in mind, the Commission's Cognition Unit of Directorate General "Information Society and Media" consulted, during the period December 2005–March 2006, a number of leading researchers in different disciplines to explore the potential for making progress toward the creation of a scientific foundation for engineering artificial cognitive systems, and the positioning of artificial cognition as an enabling technology in many areas of applied systems engineering. Five thematic workshops were held with a view to soliciting recommendations on cutting edge research, longer and medium-term R&D goals, scope for interdisciplinary co-operation and impact (http://cordis.europa.eu.int/ist/cognition/presentations.htm). A major conclusion of these consultations was that

> "machine learning or, more generally, learning in artificial systems, comprises a set of methods and techniques that are extremely relevant for ACS. It extends the remit of the latter from immediate natural environments (as for example of robots, robotic devices and other appliances and machines) to more generalised notions of environment, including all sorts of digital spaces."

(from: *Artificial Cognitive Systems in FP7: A Report on Expert Consultations for the EU Seventh Framework Programme 2007-2013 for Research and Technology Development*).

Machine learning is also thought to be of crucial importance in developing next-generation robots. Indeed, according to the 2006 Strategic Research Agenda of EUROP (European Robotics Platform) "there is a need to endow the systems with higher cognitive functions that allows recognition of context, reasoning about actions and a higher degree of error diagnostics and failure recovery. Such flexibility can only be achieved through use of advanced cognitive skills and requires elements of perception, decision making, machine learning and other intelligent systems." The relevance of machine learning in "cognitive robotics" was stressed already at the thematic Workshop for EU Seventh Research Framework Programme held in Luxembourg, on 20th December 2005, where it was pointed out that "learning theory can contribute useful concepts to the scientific foundations of cognitive robotics, with its strong maths and formalisation methods. Machine learning methods can also help in specific scenarios when dealing e.g. with very fast or one-shot learning or with very large or high-dimensional data-sets," and "research should try to avoid a too narrow, specific data-driven approach."

Our project was positioned precisely within this context. We contributed to the theory and application of learning in artificial systems by bringing to substantial development a paradigm shift that is just emerging. Indeed, there is an increasing awareness of the importance of similarity-based approaches to pattern recognition and machine learning and research in this area has gone past the proof-of-concept phase and is now spreading rapidly. On the other hand, traditional techniques are inherently unable to deal satisfactorily with the complexity and richness arising in many real-world situations, thereby hindering the application of machine learning techniques to a whole variety of relevant, real-world problems. Hence, progress in similarity-based approaches will surely be beneficial for machine learning as a whole and, consequently, for the long-term enterprise of building ACS's.

Given the substantial effort that we devoted to the two large-scale biomedical imaging applications (work packages WP6 and WP7) and the direct involvement of leading pathologists and neuroscientists in the project (from the University Hospital Zurich and the Verona-Udine Brain Imaging and Neuropsychology Program) we also contributed towards the more concrete objective of providing effective, advanced techniques to assist in the diagnosis of renal cell carcinoma, one of the ten most frequent malignancies in Western countries, as well as of major psychoses such as schizophrenia and bipolar disorder. Indeed, these problems are not amenable to be tackled with traditional machine learning techniques due to the difficulty of deriving suitable feature-based descriptions as well as the intrinsic non-metric behavior of any meaningful similarity function. In fact, many biomedical applications exhibit precisely the same characteristics. The successful outcome of our experimentation provides evidence as to the practical applicability of our approach in biomedicine, thereby fostering further research along the lines set up by SIMBAD, both at the methodological and at the practical level. This would potentially open new opportunities in health and disease management and bring radical improvements to the quality and efficiency of our healthcare systems, which is one of the priorities of FP7. It seems that the EC had precisely this objective in mind in writing the ICT work programme 2007-08, where it is stressed that "the new capabilities of modelling, simulation and biomedical imaging, combined with knowledge about diseases that ranges from molecular to organ and system levels, give rise to a new generation of predictive medicine." Indeed, the fifth challenge formulated in the FP7 ICT work programme 2007-08 ("Towards sustainable and personalised healthcare") identifies in the biomedical context a major application field of machine learning and pattern recognition techniques. The expected outcome of such methods includes "the development of environments for predictive, individualised medicine based on tools for patient-specific computational modelling and simulation which target specific clinical needs such as prediction of diseases, early diagnosis, surgery planning, treatment and training." Research related to this challenge should "develop Europe's excellence in biomedical informatics and molecular medicine by bringing ICT, medical device, medical imaging, pharmaceutical and biotech companies more closely together."

Our research project responds perfectly to the *desiderata* expressed in *A Preparatory Workshop for EU Seventh Research Framework Programme* held in Luxembourg on December 2005 on Machine "Learning and Cognition," where the participants brought forward several research topics, among which:

> "exploring the limits of vector representations and possible alternative methods, including how to choose the best representation, the possibility of adaptive representations etc."

Further, it was strongly felt that:

> "mathematical foundations are important in this area of research. Machine learning has useful tools to bring to the table, but new work is needed. One avenue could be to link game theory and/or agent modelling more closely to statistical learning methods. Research could also address the use of optimisation methods to assist in the analyses of new representations, or manifold learning techniques for reducing the complexity of data in high-dimensional spaces or over time"

which is precisely what we did with this project.

By trying to overcome the limits of vector representations, we therefore contribute towards the goal of building next generation machine learning algorithms and hence to the long-term vision of building more sophisticated ACS's.

The dissemination of the project's results has taken place mainly through publications in the top-level specialized technical journals and presentations at the leading international conferences, workshops, summer schools, etc. We also organized a series of satellite scientific events, and launched a new series of international workshops devoted precisely to the project's theme. Further, will publish a book containing the major results achieved by the consortium, which will be published by Springer in the "Advances in Computer Vision and Pattern Recognition" series.

More details on our dissemination activities can be found in Section 4.2 of this document.

4.1.5 Address of the project public website and relevant contact details

The project's website is: http://simbad-fp7.eu

The scientific coordinator is:

Prof. Marcello Pelillo
DAIS - Università Ca' Foscari Venezia
Via Torino 155, 30172 Venezia Mestre, Italy
Tel: (39) 041 2348.440
Fax: (39) 041 2348.419
E-mail: marcello.pelillo@gmail.com
URL: http://www.dsi.unive.it/~pelillo

4.2 Use and dissemination of foreground

The consortium put a lot of effort to ensuring the highest diffusion of the research results, both inside and outside SIMBAD.

In order to strengthen the communication among the partners, we set up a blog on the project's site. We also made extensive use of the *SIMBAD Technical Report Series* as a tool to provide a timely access of information within the consortium, and increase interactions among the SIMBAD partners (overall, we produced 109 technical reports).

To strengthen internal collaborations, during the course of the project we had many exchange visits among partners which resulted in several joint publications. We run several project meetings (Venice, April 2008; York, November 2008; Zurich, June 2009; Verona, November 2009; Lisbon, May 2010; Delft, November 2010) and a one-week preparatory meeting focused on our biomedical applications (Zurich, February 2010). Also, in July 2010 we run a highly successful "hands-on" internal workshop in Castelbrando, Treviso, Italy.

The external dissemination of the project's results took place mainly through publications in the top-level conferences and journals in the fields of machine learning, pattern recognition and computer vision. We also disseminated the results related to the project's main applications to the medical and the chemometrics communities.

During the 42 months of the project the consortium has produced about 150 peer-reviewed scientific papers, of which 21 were published (or accepted for publication) in journals, and 6 are under review. The following table lists the most relevant ones for each work package (see the individual deliverables for a more comprehensive list).

Table A1
SELECTED SCIENTIFIC PUBLICATIONS

No.	Title	Authors	Venue	Issue	Pages	Year	WP
1	A hybrid generative/discriminative classification framework based on free-energy terms	A. Perina, M. Cristani, U. Castellani, V. Murino, N. Jojic	ICCV		2058-2065	2009	WP2
2	Pairwise probabilistic clustering using evidence accumulation	S. Rota Bulò, A. Lourenço, A. Fred, M. Pelillo	SSPR		395-404	2010	WP2
3	2D shape recognition using information theoretic kernels	M. Bicego, A. Martins, V. Murino, P. Aguiar, M. Figueiredo	ICPR			2010	WP2
4	Free energy score space	A. Perina, M. Cristani, U. Castellani, V. Murino, N. Jojic	NIPS 2009		1428–1436	2009	WP2
5	Nonextensive information theoretic kernels on measures	A. Martins, M. Figueiredo, P. Aguiar, N. Smith, E. P. Xing	Journal of Machine Learning Research	10	935-975	2009	WP2
6	One-lead ECG-based personal identification using Ziv-Merhav cross parsing	D. Coutinho, A. Fred, M. Figueiredo	ICPR			2010	WP2
7	On the scalability of evidence accumulation clustering	A. Lourenço, A. Fred, A. K. Jain	ICPR			2010	WP2
8	Revisiting complex moments for 2-D shape representation and image normalization	J. Crespo and P. Aguiar	IEEE Transactions on Image Processing	20	2896-2911	2011	WP2
9	Online learning of structured predictors with multiple kernels	A. Martins, N. Smith, E. Xing, P. Aguiar, M. Figueiredo	AISTATS		507-515	2011	WP2
10	Nonextensive entropic kernels	A. Martins, M. Figueiredo, P. Aguiar, N. Smith, E. P. Xing	ICML		640-647	2008	WP2
11	Towards weakly supervised semantic segmentation by means of multiple	A. Vezhnevets, J. M. Buhmann	CVPR		3249-3256	2010	WP3

	instance and multitask learning						
12	Image dissimilarity-based quantification of pathology	L. Sørensen, M. de Bruijne, R. Duin, M. Loog, P. Lo, A. Dirksen	MICCAI		37-44	2010	WP3
13	The dissimilarity representation as a tool for three-way data classification using 2D measures	D. Porro-Muñoz, R. Duin, M. Orozco-Alzate, I. Talavera, J. Londono-Bonilla	Signal Processing	91	2520-2529	2011	WP3
14	A study on combining sets of differently measured dissimilarities	A. Ibba, R. Duin, W. J. Lee	ICPR		3360-3363	2010	WP3
15	Non-Euclidean dissimilarities: Causes and informativeness	R. Duin, E. Pekalska	S+SSPR		324-333	2010	WP3
16	Dissimilarity representation on functional spectral data for classification	D. Porro-Muñoz, I. Talavera, R. Duin, N. Hernandez, M. Orozco-Alzate	Journal of Chemometrics		---	2011	WP3
17	An empirical comparison of kernel-based and dissimilarity-based feature spaces	S.W. Kim, R. Duin	S+SSPR		559-568	2010	WP3
18	The dissimilarity space: Between structural and statistical pattern recognition	R. Duin, E. Pekalska	Pattern Recognition Letters		---	in press	WP3
19	The minimum transfer cost Principle for model-order selection	M. Frank, M. H. Chehreghani, J. M. Buhmann	ECML PKDD		---	2011	WP3
20	Information theoretic model validation for clustering	J. M. Buhmann	ISIT		1398-1402	2010	WP3
21	Selecting the rank of truncated SVD by maximum approximation capacity	M. Frank, J. M. Buhmann	ISIT		---	2011	WP4
22	Bayesian partitioning of large-scale distance data	D. Adametz, V. Roth	NIPS		---	2011	WP4
23	The translation-invariant Wishart-Dirichlet process for clustering distance data	J. E. Vogt, S. Prabhakaran, J. Fuchs, V. Roth	ICML		1111-1118	2010	WP4
24	Spherical embeddings for non-Euclidean dissimilarities	R. C. Wilson, E. R. Hancock, E. Pekalska, R. Duin	CVPR		1903-1910	2010	WP4
25	Bridging structure and feature representations in graph matching	W. Lee, V. Cheplygina, D. Tax, M. Loog, R. Duin	International Journal of Pattern Recognition and Artificial Intelligence		---	in press	WP4
26	Coupled prediction-classification for robust visual tracking	I. Patras, E. R. Hancock	IEEE Transactions on Pattern Analysis and Machine Intelligence	32	1537-1552	2010	WP4
27	Graph characterization via Ihara coefficients	P. Ren, R.C. Wilson, E.R. Hancock	IEEE Transactions of Neural Networks	22	233-245	2011	WP4
28	Geometric characterization and clustering of graphs using heat kernel embeddings	B. Xiao, R.C. Wilson, E.R. Hancock	Image and Vision Computing	28	1003-1021	2010	WP4
29	Manifold embedding for shape analysis	B. Xiao, H. Yu, E.R. Hancock	Neurocomputing	73	1606--1613	2010	WP4
30	A polynomial characterization of hypergraphs using the Ihara zeta function	P. Ren, T. Aleksic, R.C. Wilson, E. R. Hancock	Pattern Recognition	44	1941-1957	2011	WP4
31	Supervised relevance maps for increasing the distinctiveness of facial images	M. Kawulok, J. Wu, E. R. Hancock	Pattern Recognition	44	929-939	2011	WP4
32	Ihara zeta functions, quantum walks and cospectrality in strongly regular graphs	P. Ren, T. Aleksic, R.C. Wilson, E. R. Hancock	Quantum Information Processing	10	405--417	2011	WP4
33	Efficient computation of Ihara coefficients using the Bell polynomial recursion	S Rota Bulò, E. R. Hancock, F. Aziz, M. Pelillo	Linear Algebra and Applications	--	---	in press	WP4
34	Graph matching through entropic manifold alignment	F. Escolano, E.R. Hancock, M. A. Lozano	CVPR		2417-242	2011	WP4
35	Imposing semi-local geometric constraints for accurate correspondence selection in structure from motion: A game- theoretic perspective.	A. Albarelli, E. Rodolà, A. Torsello	International Journal of Computer Vision	---	---	in press	WP5
36	Graph-based quadratic optimization: A fast evolutionary approach	S. Rota Bulò, M. Pelillo, I. M. Bomze	Computer Vision and Image Understanding	115	984-995	2011	WP5
37	A game theoretic approach to partial clique	S. Rota Bulò, A. Torsello, M.	Image and Vision	27	911-922	2009	WP5

	enumeration	Pelillo	Computing				
38	Infection and immunization: A new class of evolutionary game dynamics	S. Rota Bulò, I. M. Bomze	Games and Economic Behaviour	71	193-211	2011	WP5
39	Graph transduction as a non-cooperative game	A. Erdem, M. Pelillo	Neural Computation	---	---	in press	WP5
40	A generalization of the Motzkin-Straus theorem to hypergraphs	S. Rota Bulò, M. Pelillo	Optimization Letters	3	287-295	2009	WP5
41	A game-theoretic approach to fine surface registration without initial motion estimation	A. Albarelli, E. Rodolà, A. Torsello	CVPR		430 -437	2010	WP5
42	Structured class-labels in random forests for semantic image labelling	P. Kontschieder, S. Rota Bulò, H. Bischof, M. Pelillo	ICCV		---	2011	WP5
43	Matching as a non-cooperative game	A. Albarelli, S Rota Bulò, A. Torsello, M. Pelillo	ICCV		1319-1326	2009	WP5
44	A game-theoretic approach to hypergraph clustering	S. Rota Bulò, M. Pelillo	NIPS	22	1571-1579	2009	WP5
45	Loosely distinctive features for robust surface alignment.	A. Albarelli, E. Rodolà, A. Torsello	ECCV		519-532	2010	WP5
46	Probabilistic clustering using the Baum-Eagon inequality	S. Rota Bulò, M. Pelillo	ICPR		1429 - 1432	2010	WP5
47	Computational pathology: Challenges and promises for tissue analysis	T. J. Fuchs, J. M. Buhmann	Computerized Medical Imaging and Graphics	35	515-530	2011	WP6
48	Renal cancer cell classification using generative embeddings and information theoretic kernels	M. Bicego, A. Ulas, P. Schüffler, U. Castellani, V. Murino, A. Martins, P. Aguiar, M. Figueiredo	IAPR-PRIB		---	2011	WP6
49	A multiple kernel Learning algorithm for cell nucleus classification of renal cell carcinoma	P. Schüffler, A. Ulas, U. Castellani, V. Murino	ICIAP		---	2011	WP6
50	Hybrid generative-discriminative nucleus classification of renal cell carcinoma	A. Ulas, P. Schüffler, M. Bicego, U. Castellani, V. Murino	SIMBAD		77-89	2011	WP6
51	Combining data sources nonlinearly for cell nucleus classification of renal cell carcinoma	M. Gönen, A. Ulas, P. Schüffler, U. Castellani, V. Murino	SIMBAD		250-260	2011	WP6
52	Brain morphometry by probabilistic latent semantic analysis	U. Castellani, A. Perina, V. Murino, M. Bellani, G. Rambaldelli, M. Tansella, P. Brambilla	MICCAI		177-184	2010	WP7
53	Dissimilarity-based detection of schizophrenia	A. Ulaş, R. Duin, U. Castellani, M. Loog, M. Bicego, V. Murino, M. Bellani, S. Cerruti, M. Tansella, P. Brambilla	International Journal of Imaging Systems and Technology	21	179-192	2011	WP7
54	A hybrid generative/discriminative method for classification of regions of interest in schizophrenia brain MRI	D. S. Cheng, M. Bicego, U. Castellani, M. Cristani, S. Cerruti, M. Bellani, G. Rambaldelli, M. Atzori, P. Brambilla, V. Murino	MICCAI Workshop		174-184	2009	WP7
55	Schizophrenia classification using regions of interest in brain MRI	D. S. Cheng, M. Bicego, U. Castellani, S. Cerruti, M. Bellani, G. Rambaldelli, M. Atzori, P. Brambilla, V. Murino	IDAMAP		47-52	2009	WP7
56	Heat diffusion based dissimilarity analysis for schizophrenia classification	A. Ulaş, U. Castellani, V. Murino, M. Bellani, M. Tansella, P. Brambilla	IAPR-PRIB		306-317	2011	WP7
57	Selecting scales by multiple kernel learning for shape diffusion analysis	U. Castellani, A. Ulaş, V. Murino, M. Bellani, M. Tansella, P. Brambilla	MICCAI Workshop		148-158	2011	WP7
58	A new shape diffusion descriptor for brain classification	U. Castellani, P. Mirtuono, V. Murino, M. Bellani, M. Tansella, P. Brambilla	MICCAI		---	2011	WP7

Besided the scientific publications, the consortium has put a lot of effort towards disseminating its achievements in a variety of ways. The following table summarizes the main activities carried out during the 42 months of the project.

Table A2:
LIST OF MAIN DISSEMINATION ACTIVITIES

No.	Type of activity	Main leader	Title	Date	Place
1	Book	UNIVE	*Similarity-Based Pattern Analysis and Recognition* Springer's Series "Advances in Computer Vision and Pattern Recognition"	mid-2012 (to appear)	---
2	Workshop	UNIVE	SIMBAD 2011 -- First International Workshop on Similarity-Based Pattern Analysis and Recognition (Springer's LNCS proceedings; videolectures coverage; sponsored by PASCAL 2 and IAPR)	September 2011	Venice, Italy
3	Workshop	UNIYORK	SIMBAD 2013 -- Second International Workshop on Similarity-Based Pattern Analysis and Recognition	Summer 2013	York, UK
4	Workshop	UNIVE	ICML 2010 Workshop on "Learning in non-(geo)metric spaces" (sponsored by PASCAL 2; videolectures coverage)	June 2010	Haifa, Israel
5	Special Session	UNIVE UNIYORK	Similarity-Based Pattern Analysis and Recognition Special session at S+SSPR 2010 (videolectures coverage)	August 2010	Cesme, Izmir, Turkey
6	Special Session	UNIVE	"Learning and Intelligent OptimizatioN in Structured domains" Special session at LION 2010	January 2010	Venice, Italy
7	Tutorial	UNIVE	"Game theory in computer vision and pattern recognition" CVPR 2011 Tutorial	June 2011	Colorado Springs, CO
8	Tutorial	UNIVE UNIYORK UNIVR	"Beyond features: Similarity-based pattern analysis and recognition" ICIAP 2011 Tutorial	September 2011	Ravenna, Italy
9	Tutorial	UNIVE	"Game theory in pattern recognition and machine learning" ICPR 2010 Tutorial	August 2010	Istanbul, Turkey
10	Tutorial	UNIVE	"Game theory in pattern recognition" PRIA 2010 Tutorial	December 2010	St. Petersburg, Russia
11	Tutorial	TUD	"Issue of non-Euclidean data" CIARP 2011 Tutorial	November 2011	Pucon, Chile
12	Summer School Course	UNIVE TUD	Summer School on "Graphs in computer graphics, image and signal analysis"	August 2011	Bornholm, Denmark
13	Summer School Course	UNIYORK	CVPR Summer School 2010	January 2010	Kioloa, Australia
14	Summer School Course	UNIVE	Summer School VISMAC 2010	November 2010	Catania, Italy
15	Magazine Article	UNIVE	"Artificial intelligence on a learning curve" In *Projects: Science, Technology and Innovation*, Insight Publisher	December 2010	
16	Newsletter Article	UNIVE	"Extending the frontiers of artificial intelligence" Cordis News	November 2008	
17	Newsletter Entry	UNIVE	FET Through the Keyhole:	January	

			Future and Emerging Technologies in Europe	2011	
18	Talk	UNIVE	Project Exhibition at ECML PKDD 2009	September 2009	Bled, Slovenia

Other activities include: invited/keynote talks by all PI's at several international conferences as well as seminars given in various research labs all over the world. Further, the activities related to the project have been advertised via several interviews which are available at the project's website.

4.3 Report on societal implications

Replies to the following questions will assist the Commission to obtain statistics and indicators on societal and socio-economic issues addressed by projects. The questions are arranged in a number of key themes. As well as producing certain statistics, the replies will also help identify those projects that have shown a real engagement with wider societal issues, and thereby identify interesting approaches to these issues and best practices. The replies for individual projects will not be made public.

A General Information *(completed automatically when **Grant Agreement number** is entered.*	
Grant Agreement Number:	
Title of Project:	
Name and Title of Coordinator:	

B Ethics	
1. Did your project undergo an Ethics Review (and/or Screening)? • If Yes: have you described the progress of compliance with the relevant Ethics Review/Screening Requirements in the frame of the periodic/final project reports? Special Reminder: the progress of compliance with the Ethics Review/Screening Requirements should be described in the Period/Final Project Reports under the Section 3.2.2 *'Work Progress and Achievements'*	***No***
2. Please indicate whether your project involved any of the following issues (tick box) :	***YES***
RESEARCH ON HUMANS	
• Did the project involve children?	
• Did the project involve patients?	
• Did the project involve persons not able to give consent?	
• Did the project involve adult healthy volunteers?	
• Did the project involve Human genetic material?	
• Did the project involve Human biological samples?	
• Did the project involve Human data collection?	yes
RESEARCH ON HUMAN EMBRYO/FOETUS	
• Did the project involve Human Embryos?	
• Did the project involve Human Foetal Tissue / Cells?	
• Did the project involve Human Embryonic Stem Cells (hESCs)?	
• Did the project on human Embryonic Stem Cells involve cells in culture?	
• Did the project on human Embryonic Stem Cells involve the derivation of cells from Embryos?	
PRIVACY	
• Did the project involve processing of genetic information or personal data (eg. health, sexual lifestyle, ethnicity, political opinion, religious or philosophical conviction)?	
• Did the project involve tracking the location or observation of people?	
RESEARCH ON ANIMALS	
• Did the project involve research on animals?	
• Were those animals transgenic small laboratory animals?	
• Were those animals transgenic farm animals?	

• Were those animals cloned farm animals?	
• Were those animals non-human primates?	
RESEARCH INVOLVING DEVELOPING COUNTRIES	
• Did the project involve the use of local resources (genetic, animal, plant etc)?	
• Was the project of benefit to local community (capacity building, access to healthcare, education etc)?	
DUAL USE	
• Research having direct military use	no
• Research having the potential for terrorist abuse	no

C Workforce Statistics

3. Workforce statistics for the project: Please indicate in the table below the number of people who worked on the project (on a headcount basis).

Type of Position	Number of Women	Number of Men
Scientific Coordinator		
Work package leaders		
Experienced researchers (i.e. PhD holders)		
PhD Students		
Other		

4. How many additional researchers (in companies and universities) were recruited specifically for this project?	
Of which, indicate the number of men:	

D Gender Aspects

5. Did you carry out specific Gender Equality Actions under the project?	☒ ○	Yes No

6. Which of the following actions did you carry out and how effective were they?

		Not at all effective				Very effective
☒	Design and implement an equal opportunity policy	○	○	x	○	○
❑	Set targets to achieve a gender balance in the workforce	○	○	○	○	○
❑	Organise conferences and workshops on gender	○	○	○	○	○
❑	Actions to improve work-life balance	○	○	○	○	○
○	Other:					

7. Was there a gender dimension associated with the research content – i.e. wherever people were the focus of the research as, for example, consumers, users, patients or in trials, was the issue of gender considered and addressed?

○ Yes- please specify

☒ No

E Synergies with Science Education

8. Did your project involve working with students and/or school pupils (e.g. open days, participation in science festivals and events, prizes/competitions or joint projects)?

○ Yes- please specify

☒ No

9. Did the project generate any science education material (e.g. kits, websites, explanatory booklets, DVDs)?

○ Yes- please specify

☒ No

F Interdisciplinarity

10. Which disciplines (see list below) are involved in your project?

○ Main discipline[21]: 1.1

○ Associated discipline[21]: 1.5 | ○ Associated discipline[21]:

G Engaging with Civil society and policy makers

11a Did your project engage with societal actors beyond the research community? *(if 'No', go to Question 14)*	○ ☒	Yes No

11b If yes, did you engage with citizens (citizens' panels / juries) or organised civil society (NGOs, patients' groups etc.)?

○ No

○ Yes- in determining what research should be performed

○ Yes - in implementing the research

○ Yes, in communicating /disseminating / using the results of the project

[21] Insert number from list below (Frascati Manual).

11c In doing so, did your project involve actors whose role is mainly to organise the dialogue with citizens and organised civil society (e.g. professional mediator; communication company, science museums)?	○ ○	Yes No

12. Did you engage with government / public bodies or policy makers (including international organisations)

- ○ No
- ○ Yes- in framing the research agenda
- ○ Yes - in implementing the research agenda
- ○ Yes, in communicating /disseminating / using the results of the project

13a Will the project generate outputs (expertise or scientific advice) which could be used by policy makers?

- ○ Yes – as a **primary** objective (please indicate areas below- multiple answers possible)
- ○ Yes – as a **secondary** objective (please indicate areas below - multiple answer possible)
- ○ No

13b If Yes, in which fields?

Agriculture Audiovisual and Media Budget Competition Consumers Culture Customs Development Economic and Monetary Affairs Education, Training, Youth Employment and Social Affairs		Energy Enlargement Enterprise Environment External Relations External Trade Fisheries and Maritime Affairs Food Safety Foreign and Security Policy Fraud Humanitarian aid		Human rights Information Society Institutional affairs Internal Market Justice, freedom and security Public Health Regional Policy Research and Innovation Space Taxation Transport	

13c If Yes, at which level? ○ Local / regional levels ○ National level ○ European level ○ International level	

H Use and dissemination

14. How many Articles were published/accepted for publication in peer-reviewed journals?	**21**
To how many of these is open access[22] provided?	
How many of these are published in open access journals?	**1**
How many of these are published in open repositories?	
To how many of these is open access not provided?	**20**
Please check all applicable reasons for not providing open access:	
☒ publisher's licensing agreement would not permit publishing in a repository ❑ no suitable repository available ❑ no suitable open access journal available ❑ no funds available to publish in an open access journal ❑ lack of time and resources ❑ lack of information on open access ❑ other[23]: ……………	

15. How many new patent applications ('priority filings') have been made? *("Technologically unique": multiple applications for the same invention in different jurisdictions should be counted as just one application of grant).*		**none**
16. Indicate how many of the following Intellectual Property Rights were applied for (give number in each box).	Trademark	
	Registered design	
	Other	
17. How many spin-off companies were created / are planned as a direct result of the project?		**none**
Indicate the approximate number of additional jobs in these companies:		

18. Please indicate whether your project has a potential impact on employment, in comparison with the situation before your project:

❑	Increase in employment, or	❑	In small & medium-sized enterprises
❑	Safeguard employment, or	❑	In large companies
❑	Decrease in employment,	☒	None of the above / not relevant to the project
❑	Difficult to estimate / not possible to quantify		

[22] Open Access is defined as free of charge access for anyone via Internet.
[23] For instance: classification for security project.

19. For your project partnership please estimate the employment effect resulting directly from your participation in Full Time Equivalent (*FTE = one person working fulltime for a year*) jobs:	*Indicate figure:*
Difficult to estimate / not possible to quantify	☒

I Media and Communication to the general public

20. As part of the project, were any of the beneficiaries professionals in communication or media relations?

○ Yes ☒ No

21. As part of the project, have any beneficiaries received professional media / communication training / advice to improve communication with the general public?

○ Yes ☒ No

22 Which of the following have been used to communicate information about your project to the general public, or have resulted from your project?

❑	Press Release	x	Coverage in specialist press
❑	Media briefing	❑	Coverage in general (non-specialist) press
❑	TV coverage / report	❑	Coverage in national press
❑	Radio coverage / report	❑	Coverage in international press
❑	Brochures /posters / flyers	x	Website for the general public / internet
❑	DVD /Film /Multimedia	❑	Event targeting general public (festival, conference, exhibition, science café)

23 In which languages are the information products for the general public produced?

❑	Language of the coordinator	☒	English
❑	Other language(s)		

Question F-10: Classification of Scientific Disciplines according to the Frascati Manual 2002 (Proposed Standard Practice for Surveys on Research and Experimental Development, OECD 2002):

FIELDS OF SCIENCE AND TECHNOLOGY

1. NATURAL SCIENCES
1.1 Mathematics and computer sciences [mathematics and other allied fields: computer sciences and other allied subjects (software development only; hardware development should be classified in the engineering fields)]
1.2 Physical sciences (astronomy and space sciences, physics and other allied subjects)
1.3 Chemical sciences (chemistry, other allied subjects)
1.4 Earth and related environmental sciences (geology, geophysics, mineralogy, physical geography and other geosciences, meteorology and other atmospheric sciences including climatic research, oceanography, vulcanology, palaeoecology, other allied sciences)
1.5 Biological sciences (biology, botany, bacteriology, microbiology, zoology, entomology, genetics, biochemistry, biophysics, other allied sciences, excluding clinical and veterinary sciences)

2 ENGINEERING AND TECHNOLOGY
2.1 Civil engineering (architecture engineering, building science and engineering, construction engineering, municipal and structural engineering and other allied subjects)

2.2 Electrical engineering, electronics [electrical engineering, electronics, communication engineering and systems, computer engineering (hardware only) and other allied subjects]

2.3. Other engineering sciences (such as chemical, aeronautical and space, mechanical, metallurgical and materials engineering, and their specialised subdivisions; forest products; applied sciences such as geodesy, industrial chemistry, etc.; the science and technology of food production; specialised technologies of interdisciplinary fields, e.g. systems analysis, metallurgy, mining, textile technology and other applied subjects)

3. MEDICAL SCIENCES

3.1 Basic medicine (anatomy, cytology, physiology, genetics, pharmacy, pharmacology, toxicology, immunology and immunohaematology, clinical chemistry, clinical microbiology, pathology)

3.2 Clinical medicine (anaesthesiology, paediatrics, obstetrics and gynaecology, internal medicine, surgery, dentistry, neurology, psychiatry, radiology, therapeutics, otorhinolaryngology, ophthalmology)

3.3 Health sciences (public health services, social medicine, hygiene, nursing, epidemiology)

4. AGRICULTURAL SCIENCES

4.1 Agriculture, forestry, fisheries and allied sciences (agronomy, animal husbandry, fisheries, forestry, horticulture, other allied subjects)

4.2 Veterinary medicine

5. SOCIAL SCIENCES

5.1 Psychology

5.2 Economics

5.3 Educational sciences (education and training and other allied subjects)

5.4 Other social sciences [anthropology (social and cultural) and ethnology, demography, geography (human, economic and social), town and country planning, management, law, linguistics, political sciences, sociology, organisation and methods, miscellaneous social sciences and interdisciplinary , methodological and historical S1T activities relating to subjects in this group. Physical anthropology, physical geography and psychophysiology should normally be classified with the natural sciences].

6. HUMANITIES

6.1 History (history, prehistory and history, together with auxiliary historical disciplines such as archaeology, numismatics, palaeography, genealogy, etc.)

6.2 Languages and literature (ancient and modern)

6.3 Other humanities [philosophy (including the history of science and technology) arts, history of art, art criticism, painting, sculpture, musicology, dramatic art excluding artistic "research" of any kind, religion, theology, other fields and subjects pertaining to the humanities, methodological, historical and other S1T activities relating to the subjects in this group]