BANDO *COOPERINT 2010*

SCHEDA PROGETTO

Soggetto proponente	Manuele Bicego
Dipartimento di afferenza	Dipartimento di Informatica
Tipologia di riferimento (nell'ambito della partecipazione al Bando CooperInt 2010)	B3
Università/Istituto di destinazione	Universidad Nacional de Colombia - sede Manizales
Paese	Colombia

DESCRIZIONE DETTAGLIATA DEL PROGETTO DI MOBILITÀ
(almeno 3000 caratteri)

TITLE
"Investigation of advanced Hidden Markov Models-related techniques for the analysis of seismic signals from multiple volcano"

MOTIVATIONS and OBJECTIVES
The project is aimed at investigating advanced Pattern Recognition techniques for the analysis of seismic signals, coming from different volcanoes. In particular the main interest is in the classification of seismic events: actually, in spite of the availability of fully digitized seismic networks and automated information processing techniques, classification of volcanic earthquakes at many volcano observatories is still fully undertaken by human expert judgments. Such a task implies a heavy workload and, therefore, is subject to mistakes related to tedious evaluations and changes or undertraining of personnel. Therefore automatic Pattern Recognition techniques may be very helpful in this context; actually, the development of automatic classification system for volcanic earthquakes has been explored in a number of previous studies. Techniques applied range from neural networks to dissimilarity-based classifiers, from Bayesian networks to hidden Markov models (HMM). In particular, Hidden Markov Models seem to be a very appealing choice, due to the sequential nature of the analyzed signal. In recent years, many advances have been proposed in this context, which have been never applied to this seismic scenario.

The project is motivated by two important facts: from one hand, M. Bicego has gained a relevant knowledge and expertise on HMM (both from the theoretical and the applicative point of view), being very active in the field since 2001. On

the other hand, the Pattern Recognition and Signal Processing group of the University of Colombia (Manizales) has one of the largest database in the world of seismic signals, which contains more than 16 years of daily recordings, taken from at least 5 stations and relative to 4 volcanoes. They have been developing Pattern Recognition systems based on these data since 2005, gaining an invaluable context-related expertise. The analysis of such peculiar database is a real challenge also from the theoretical point of view, but hardly leading to sensible results without face to face interaction with the experts in the field.

METHODOLOGICAL APPROACH
The problem may be faced following different directions; the goal of the visit will be to identify the most feasible ways, possibly starting a long-term cooperation. In particular, the idea is to follow the recent trends developed by M. Bicego and colleagues in the generative embeddings field (also referred to as hybrid generative-discriminative schemes), which currently represents one of the hottest topics in the Machine Learning / Pattern Recognition community. Even if different versions have been recently proposed, the underlying basic idea is to exploit a generative model (like a Hidden Markov Model) to project entities of interest in a feature space, where a discriminative classifier may be used. This scheme seems to be particularly suited in the above referenced scenario due to the presence of different length events (which can be dealt with HMM) and of multiple sources (which can be dealt with feature-level fusion in the generative embeddings). During the visit, different embeddings will be analyzed, starting from simple up to very recent ones. Novel embeddings, derived from the specific application domain, may also be proposed. All methodologies will be tested using (a part of) the database of the Colombia University, possibly comparing them with already developed methodologies.

IMPLICATIONS and EXPECTED RESULTS
Clearly outcomes of this research field may have many practical implications: preventive responses and pre/post-eruption management strategies, e.g. evacuations, can be better planned if volcanic activity is continuously monitored and automatically analyzed. This is especially crucial in the Colombia area, where the Nevado del Ruiz volcano is very active -- in the 1985 eruption more than 20 thousands of people perished in the city of Armero. In response to the disaster, the government founded the National Volcanological Observatory, later known as the Volcanological and Seismological Observatory at Manizales , which currently monitors Nevado del Ruiz volcano as well as other volcanoes along the so-called Cerro Machín – Cerro Bravo volcanic complex (and from which the data come).
From a scientific point of view, the project may of course lead to publication of papers which are relevant for the seismologic (or hybrid) community. On the other side (and most interestingly), the challenging nature of the problem may require a tailoring of general purpose techniques, which may lead to the ideation of novel and peculiar methodologies to be published in relevant Pattern Recognition conferences or journals.

RELATION BETWEEN THE TWO UNIVERSITIES

The project starts from communications and interactions between M. Bicego (University of Vereona) and M. Orozco-Alzate (University of Colombia), indirectly originated through the FP7 European project SIMBAD (Similarity-based Pattern Analysis and Recognition FET project number 213250), and in particular through the University of Delft (both M. Bicego and M. Orozco-Alzate visited the Bob Duin's Lab for a while). As a result, two BS students (Bioinformatics degree) in the University of Verona are currently working on the problem (stage + thesis). The goal of the visit is also to plan activities leading to a long-term contact and academic cooperation. The visit will coincide with the calendar of PhD and Master's students at the Columbia University, where M. Bicego will give a couple of seminars related on his expertise and interests.

Data	Firma del soggetto proponente
……………… ………	………………………………………………………………