

PhD topic - Image Team - 2017

# Estimation of the Graph Edit distance: Sequential and parallel strategies

**Objectives:** The Graph edit distance measures the amount of distortions required to transform one graph into another. Such a distance allows to set a metric in the space of graphs and constitutes the main connection between the graph world and Pattern Recognition or Machine learning methods. Applications of the graph edit distance are numerous but include:

- Chemo and bio informatics (prediction of the properties of molecules defined by graphs),
- Malware detection (characterization and recognition of a malware by its graph of functions calls),
- Biometry (of fingers or eyes), Video analysis (re identify a person which leave and reenter into the scene or pass within the point of view of an other camera through a graph describing his appearance),
- Handwritten document analysis (keyword spotting, clustering of documents described by a region adjacency graph),
- Image analysis (Indexation of images described by a graph of interest points or a region adjacency graph)...

Unfortunately, the exact computation of the Graph edit distance is an  $\mathcal{NP}$ -hard problem and current computers do not allow to compute this value for graphs exceding about 10 nodes. The aim of this PhD thesis consists to design new heuristics allowing to find close approximation of this exact value with execution times which should allow the computation of the graph edit distance on large graphs (over 500 nodes) or the processing of datasets of thousands of medium graphs (between 20 and 100) nodes.

The teams involved in this PhD have demonstrated that the computation of the Graph edit distance is equivalent to the minimization of a quadratic functional on the space of permutation matrices. The PhD student will have to follow two lines of research. On one hand, he will have to study the litterature and try to adapt to the graph edit distance problem different minimization schemes proposed in other frameworks. This line of research should be concluded by the proposal by the PhD of original minimization algorithms.

On an other hand, the student will have to study different parallel implementation of the existing minimization schemes. Within this context two types of parallelisation are planed: Design of full parallel algorithms on GPU and distributed algorithms. Let us stress that both parallelisation strategies should not be opposed since one software may be distributed on several machines, with several threads, each machine having its own graphical cards.

## PhD's organization and required skills

The student will be mainly registered at the University of Salerno (IT). Nevertheless, a co-tutorship<sup>1</sup> for this PhD is planed with the University of Caen (FR). The student will spend half of his PhD in each institution. The extra fees related to this organization will be taken in charge by the French/Italian University. This PhD will lead to two PhD diploma (one French, one Italian) together with an European label for the PhD. The two director for this PhD are: Mario Vento (MIVIA, University of Salerno) and Luc Brun (GREYC, ENSICAEN).

The candidate should have good skills in mathematics and programming languages (C/C++). Some skills on GPU and distributed programming would be welcome.

## Contacts

Interested candidates should contact : luc.brun@ensicaen.fr. Please include in your email one CV and one letter explaining your interest and your skills for this PhD.

Further documents such as your marks of master may be required latter.

<sup>1</sup>[http://ressources.campusfrance.org/catalogues\\_recherche/diplomes/fr/cotutelle\\_fr.pdf](http://ressources.campusfrance.org/catalogues_recherche/diplomes/fr/cotutelle_fr.pdf) (en français)