TESI @
UNIVERSITY OF EDGE HILL (UK)

The **Visual Computing lab** focuses on the topics in the areas of 3D computer vision, image processing, pattern recognition, visual data analytics, computer graphics, virtual/augmented reality, human sensing and human-robot social interactions. The research is supported by the latest technologies such as the Computer Augmented Virtual Environment (CAVE) and a range of state-of-the-art devices for data acquisition, output, visualization, and full immersive experiences including: Tobii Eye X, Vive, Z-Space, Kinects, Hype Box, Emotive, Enobio BCI, Nao Robots, 3D Scanner and printer.

Marcello Trovati, is Director of PG Studies & Deputy Director for CDS. His research interest are in the areas of big data analytics and machine learning.

Peter Vangorp is technical manager of the Visual Computing Lab. Is research interest are in the areas of data visualization, creative VR/AR applications, game engine development.

Required Skills for the thesis:

Machine Learning e Big Data Analytics

Give a look to the Lab Website

.DIEM

Valentino Modestino Graziano

University of Edge Hill (UK)

BIG DATA VISUALIZATION: USING 3D + VR VISUALIZATION FOR GENOMICS DATA

Abstract

This research will mainly focus on visualising Metabolic Pathways map from KEGG and experimental data (that can be downloaded from NCBI Geo Datasets or created by the own experiment) all in one place. Is very important to underline that this research doesn't create something of new in terms of data analysis: all the study was driven from the fact that already exists something of similar, but it's "hard" to use. To explain what is meant, consider a classical use case of the tool: a biologist who wants to conduce a research on certain organism treating it in a particular way.

So after using some particular compound or exposing it to some particular conditions, the interested organism will naturally react to the treatment. In which way way? According with its species, his genome will change, and the result will be that some genes will be more expressed, some less expressed and some will remain at previous expression level. Is also reasonable to think that its enzymes production will change. In general is possible to say that all these changes will regard the organism in its pathways. The biologist at this time of the research needs to measure the gene expression for each gene, building a sort of table where to place them.

Considering the connection between Metabolic Pathways map and the organism is natural to think that the genes in the map are exactly the same genes of the organism, so is very natural to use KEGG maps after the treatment to investigate about the answer of the administration. Indeed in this kind of studies is important to understand what changes in the metabolism of the organism, and to understand better is necessary to consider all its reactions, all the enzyme productions and similar things.

For instance, consider that experimental results reveal that a gene "SCO330" involved in fatty acid biosynthesis (in the Lipid Metabolism) after the treatment is very highly expressed. In this case is very hard in the beginning to have a clear idea of the whole gene expression because there are over seven thousands of genes, but also considering that the poor biologist after a while finds something of interesting, in the table of the results, after is necessary to locate the gene in the map looking for its pathway, and looking for the involved genes of its same pathway.

This approach to the research (that is the current approach) has two main disadvantages: it's a waste of time and, relevant information can be easily not seen (considering the big size of the data) How the recent technological progress could speed up the approach to the research? Considering that this is a problem of analysis combined with visualization, the recent 3D and VR trend (that as we saw, is increasing in recent times) could be helpful, thanks to its immersion, to improve the current approach.