DATA MINING, MACHINE LEARNING, COMPUTER VISION/IMAGE PROCESSING AND PATTERN RECOGNITION GROUP (DMC)

Group Leader: **Grace L. Samson (PhD Computer Science and Informatics)**

Group Focus: Mining Patterns from Large Complex Data Using Spatial Data Mining (SDM) and Machine Learning Techniques

Introduction:

Spatial data mining is the quantitative study of phenomena that are located in space. Studying spatial data management helps us to discover the relationship between spatial and non-spatial data and to be able to build and query an effective knowledgebase. Therefore, we continue to investigate methods for representing **BIG DATA** through mining patterns of complex datasets.

About the group:

The Data Mining, Machine Learning, Computer Vision/Image Processing and Pattern Recognition research Group is into multi-disciplinary, investigations. Traversing Expert systems, bioinformatics, knowledge representation, computational learning theory, computer vision/Image processing, Machine learning, spatial/classical data mining, informal reasoning, and spatial/statistical modelling.

We are in the era of ***big data***, and all big data are spatially referenced. As such, interpreting these data calls for highly automated approaches, which will rely on new machine learning and data mining approaches. A fusing/integrating theme for this group, is the understanding of multi-scale pattern recognition and rule identification problems by simply projecting underlying problems on an n-dimensional space (**Euclidean….**), using state-of-the-art powerful statistical and spatial algorithms, spatial data structures and Mathematical models, for modelling and solving the problems, and for learning from data.

We seek to find interactions amongst ideas based on ***statistics, artificial reasoning, data science, cognitive science, biology, GIS-Science, Computer vision, Data mining, Machine/deep learning, Image Processing, Remote sensing and Bioinformatics***.

We aim to develop practical applications from integrating one or more of these disciplines.

Applications:

- Big (spatial and non-spatial) Data mining
  - Object recognition, reconstruction, and classification at large scale
  - Machine Learning (Supervised and Unsupervised)
  - Location based image and text interpretation/retrieval.
  - Spatial Models for Computer vision, Pattern recognition Bioinformatics and Image Processing for object and pattern detection
- GIS (City Mapping, Hazardous zone mapping, Plant/Vegetation mapping, landscape mapping, Land-cover classification, Natural/Human induced phenomenon monitoring).
  - Data Mining (Feature/Pattern extraction and representation, Spatial Image processing, Spatial data modelling and management)
  - Expert/Prediction systems (Disease prediction---Bacteria, Symptoms, Diagnosis and treatments---- Market phenomenon prediction---Decision systems---- Forecasting Systems)
  - Complex spatial structure (Algorithms and Data structures)
  - Data/information retrieval and data management

Detailed description of group interest:

The inputs to the human eye and computers cameras are two dimensional, but the world is composed of three-dimensional objects. Though full computer vision requires three-dimensional information***, most real-life scenario can be simulated or modelled solely in a two-dimensional paradigm***.

We are interested in thoroughly investigating methods of mining patterns from datasets by projecting the data as a spatial data set (which generally describes any kind of data where the location in space of object holds importance), and then make predictions based on the outcome of our analyses. Our research is based on the analysis of some spatial characteristic of certain objects that exist in an ecosystem. The main idea is to extract spatial pattern from events or objects with respect to their attributes, in other words and most specifically, we look at how to describe them from a spatial (nature/characteristics of entities in an ecological environment) perspective with respect to their spatial and non-spatial attributes.

Our investigation is for predicting the likelihood of an object or an event, with a range of variables (using spatial quantity– distance, interpolation, overlay, raster creation, statistics, re-

class, multivariate analysis, maths, conditional and surface respectively) on a sample dataset and then we verify the model performance on the rest of the data.

With data acquired through dedicated imaging operations or collected through available public domains, we look to build prediction/suitability models for pattern prediction, detection, recognition and decision-making. The method of our research suggests the use of mapping statistical functions, machine learning algorithms, and AI in the prediction of large spatial and non-spatial database.

We try to achieve this by five (5) major stages –

First stage is modelling the system requirements,

Second stage is the spatial analysis,

Third stage is the statistical/mathematical analysis,

Fourth stage is the design (using ML and AI tools),

Finally, visualization.

Our ultimate goal is to integrate and further extend methods of traditional data mining in various fields for the analysis and management of large and complex data (spatial or not). The underlying concept is based on the fact that conventional database management system could be extended by modelling classical database items in form of spatial data types ***(e.g., points, lines, polygons and regions)***.

We base our research on the analysis of the spatial characteristics of certain objects following the steps outlined below:

- Describe the spatial pattern of events or objects with respect to their attributes;
- Examine modelling (predictive) methods for knowledge management of complex spatial systems
- Querying and implementing a complex spatial database (using spatial data structure and algorithms).

Our research group is particularly useful to researchers in the field of data mining especially ***spatial data mining, artificial intelligence, pattern recognition, data science and computer vision.*** Therefore, we invite researchers whose works applies to machine learning,

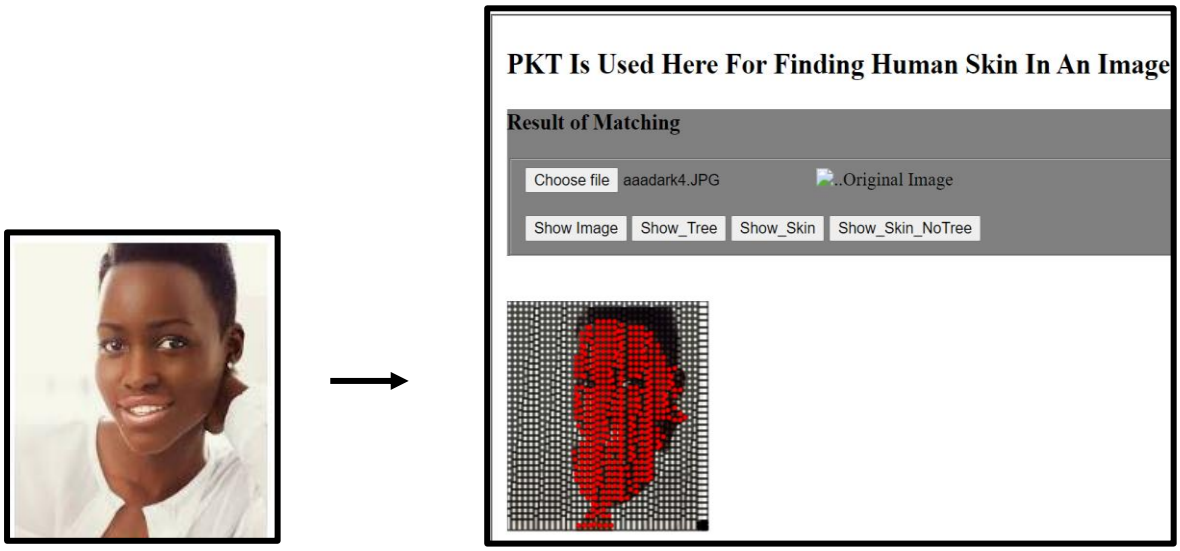
processing data for information extraction, GIS, Artificial intelligence, data mining and Computer vision.

Our ongoing projects include:

On-going research on critical/logical thinking and problem solving through computer vision, machine learning, pattern recognition and data mining, using both Python and JavaScript programming languages.

PKT, a new highly peer-reviewed detection for colour-based object detection that improves the performance of learning procedures in terms of speed, accuracy, and efficiency, using spatial inference, and algorithm. (PKT model (was applied to solving rea-world pattern recognition/pattern recognition problem and it scaled high …the article is now under publication production with Springer journals).

PKT: *Fast Color-Based Spatial Model for Human Skin Detection". ACCEPTED AND UNDER PRODUCTION FOR PUBLICATION (SPRINGER JOURNALS)*

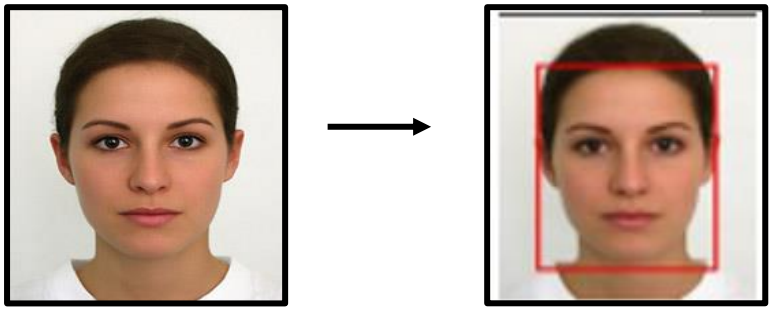


P_Clus, a robust and rapid density/proximity-based model for spatial clustering. Which will apply advanced filtering (machine learning) and data mining mechanism to achieve great efficiency on computer vision tasks.

This project requires using a large amount of data to manipulate, analyse, and find trends in activities and behaviours of consumers/customers/dataset, parameters, agents/events etc. The main target of this project is to create a NOVEL management information system that can influence operational performance and deliver continuous improvements in a growing dataset, business, organisation, or event.

P_Clus*: A robust and rapid density/proximity-based model for spatial clustering. UNPUBLISHED (Evaluation stage).*

P-Face: - Fast novel clustering algorithm for human face detection. Submitted to the 23rd int'l Conference on Image Processing, Computer vision & Pattern recognition [IPCV'19], in the 2019 World Congress in Computer Science, Computer Engineering, & Applied Computing (Accepted for the conference)…(requires improvement)



Others:

- ***aX_tree*** – A new packing techniques for static databases which performs better in space utilization for indexing, through cautious data packing.
- ***PKT-ACK*** – An improvement of the new algorithm for initial cluster centers in k -means algorithm
- ***P-Data*** – A NOVEL efficient model for imputing missing data in a dataset