**Computer Vision, Scene Analysis and Adaptive Pattern Recognition**

J R Barker, S. Sens, F. Smoes and F Young

Scene analysis, for both single frame and sequences of frames, is an important area of artificial intelligence which is making computer vision a reality. It is important for applications which range from surveillance, computer-assisted tomography, microscopy to industrial product monitoring and robotics. Our interest is specifically in the analysis of images obtained from scanning tunnelling microscopy and related techniques. It is particularly important to recognise a particular molecule against a particular background and to discriminate it from “dirt” or artefacts of the imaging process. It is also important to measure the detected target and to find its distortion from the perfect form. We have therefore developed algorithms which can detect one or more molecular targets within a scene of adsorbed molecules plus substrate. At this level we use a self-organising or synergetic algorithm which represents an unknown scene as a potential surface in an abstract feature space. the input scene is filtered against known images using a non-linear overdamped technique which locates the most dominant recognisable sub-image in the scene and identifies it. By shutting off the attention parameters for the recognised component(s) the program them seeks the next identifiable objecxt and proceeding recursively finally obtains a hierarchical description of the scene in terms of known objects. The unknown features are filtered out and may be “learned” if they represent a new object or further analysed. A second algorithm is used to determine the orientation, position and scaling of any of the recognised objects with repect to a training image (the affine transformation parameters). The novel elements of this project involve the ability to perform scene analysis, invariant pattern recognition up to an affine transformation and for adaptive learning. The techniques are proving particularly useful in molecular electronics applications where the “scenes” involve “semi-transparent” distorted molecules covering layers of other molecular and inorganic materials. Our long term interest invoves the applications of the above ideas to general problems in computer vision. Funded by the EU TOPFIT project (see grants)

Algorithms for  scanning tunnelling microscopy

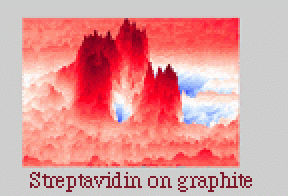
**IMAGE PROCESSING AND ADAPTIVE PATTERN RECOGNITION FOR SCANNING TUNNELLING MICROSCOPY**

J R Barker and F Young

As part of the ESPRIT programme TOPFIT we are developing tools for the visualisation and analysis of molecular images obtained locally and with collaborators in Mainz from scanning tunelling microscopy. At the simplest level we are developing software for producing fully rendered “landscape” images in both direct and stereo pair forms. Simulation tools are being developed which allow the experimenter to effectively fly over and around the surface terrain of a target molecule viwed by the STM and to then produce fully rendered computer movies of the scene. Morphing and compression techniques are being used to make this procedure cmputationally efficient.

It has been found that electrochemical STM images of arrays of molecules covering a suitable substrate such as graphite (HOPG) have a number of surprising features which makes interpretation of the images very difficult or ambiguous. For example in recent data from our partners in Mainz it has been found that a monolayer of didodecylbenzene on graphite presents a multiple image rather than a single surface image and the result is very sensitive to the tip-substrate bias potential. For example at low bias the image reveals high contrast cones associated with the benzene group but the alkyl chains are virtually invisible but there expected location is filled with a direct image of the graphite substrate which appears to shine through the monolayer. At higher bias levels the graphite image fades and is replaced with the characteristic images of the alky chains. To assist the interpretation of such complex “scenes” we are developing adaptive pattern recognition programs which are trained to recognise features in an STM image and to classify a scene into a superposition of known features and a residual “novelty” element. The technique is based on a synergetic non-linear dynamical systems algorithm which uses known images to construct attracting basins in a generalised potential function to filter out known components of a target scheme. The method is complemented by a new algorithm for determining the position, scale and rotation of recognised objects based on affine transformations. By using test images which include noisy substrates, STM artefacts and defects as well as known or predicted molecular images it is possible to give “expert” assistance to the task of image interpretation.

**Molecular Electronics**

[](http://johnreginaldbarker.co.uk/wp-content/uploads/2011/06/avadin.gif)

Molecular Electronics has two principal strands; firstly, the use of the novel combinations of macroscopic properties offered by organic materials in electronic and optoelectronics devices; and secondly, the study of systems at the molecular scale in order to determine their potential as components of molecular scale devices. The use of liquid crystals in displays and semiconducting polymers in the xerographic process are the prime examples of the impact that molecular materials can make on technology. Conjugated, conducting polymers offer good prospects for mid-term application but still require basic studies to underpin the development of materials suitable for use in practical devices. An example of how rapidly this field has developed is illustrated by the fact that Philips have now developed an electrolytic capacitor using a conducting polymer counter electrode. True molecular scale electronics (< 10 nm) is a much longer term target which, at present, lacks a strong fundamental science base. However, atomic scale resolution microscopies, and other emerging techniques offer the tools with which to develop the science essential for the exploration of this topic.

The EPSRC in the UK has in recent years supported a UK-wide initiative in Molecular Electronics to extend the use of molecular materials in electronics. In generation research has focused on the investigation of electronic properties at the molecular scale that might lead to new devices. Many different types of molecules have been proposed for synthesis or further investigation. A practical approach to the problems of molecular electronics has been taken in Glasgow by concentrating current efforts on developing reliable techniques for substrate attachment and patterning of molecules for new devices or materials. The techniques of photolithography and electron beam lithography have been successfully applied to patterning techniques at nanoelectronic scales and there is a strong interest in exploiting the electronic properties of biological molecules imparted by either their structure and/or their functionality (see Bioelectronics).

COOPER,J.M. , SHEN, J.,YOUNG,F M.,CONNOLLY ,P, BARKER, J.R., AND MOORES, G.  
The Imaging of Streptavidin and Avidin using Scanning Tunnelling Microscopy  
J. Materials Science: Materials in Electronics 5 (1994)

**How important is machine learning for bioinformatics?**

B Machine learning plays an important role in a lot of bioinformatics problems. To list a few -

* **Gene Finding Algorithms**: Hidden Markov Models (HMM)
* **Gene Expression**: Clustering Algorithms  like k-means
* **Genome Alignment**: HMM
* **Population Stratification :** PCA, MDS, manifold learning
* **GWAS :** Linear and logistic regression (Mixed linear models)
* **Genomic Selection :** Classification Algorithms like Random Forest, Decision Trees, Naive Bayes,  SVM, Logistic Regression

In last one and half year that I have worked in a bioinformatics lab, I have worked on a bunch of projects involving machine learning. I have built classification models for predicting structural variants, used Bayesian Networks for incorporating prior information in a genomic selection model, used Bayesian optimization to estimate parameters for a model etc.   
Machine learning is immensely helpful in doing bioinformatics, in fact I see a new idea along the interface of ML in bioinformatics almost every week.ioinformatics is really data intensive. As bionformaticians we viably  survive as data parasites and essential to our survival is the ability  to grow the right set of "teeth" in order to munch and munge through our  data rich diet. Gene annotation, motif finding, pathway analysis are  all but routine applications for mining, classification, functional  prediction and further downstream analysis.  
  
Few  years ago we thought that next generation sequencing was the game  changer, it made data generation cheaper and revolutionized  bioinformatics in many ways, today we sit at the top of untapped  insights yet to be released from all the data we've accumulated and it's  clear that better ways to intelligently integrate and query this data  are needed so we make sense of what is hidden underneath it. Machine  learning allows just that through an arsenal of algorithms you'd be able  to classify, aggregate,  project insights and have fun while at it.   
  
So the short answer is machine learning is very important to the big biology data that drives bioinformatics.  
  
This paper here provides examples for machine learning applications in biology  
[Page on nature.com](http://www.nature.com/nrg/journal/v16/n6/full/nrg3920.html)  
  
Another relevant read that puts things in perspective is the recent Nature article on outlooks of  data titled "[Big data: The power of petabytes](http://www.nature.com/nature/journal/v527/n7576_supp/full/527S2a.html)" by Michale Eisenstein.

[Written Nov 11](https://www.quora.com/How-important-is-machine-learning-for-bioinformatics/answer/Hisham-Eldai) • [View Upvotes](https://www.quora.com/api/mobile_expanded_voter_list?type=answer&key=dyPZFrxspfi) • Answer requested by [David Tran](https://www.quora.com/profile/David-Tran-109)

You can only go so far by analyzing experimental data; whether your own or publicly available. The next step in becoming a good computational biologist is to be able to develop models and classifiers learning differences between different conditions and being able to predict what class a new data point will be assigned to. Machine learning has been a transformational subfield in computer science and traditionally support vector machines (SVM's) in particular lead the charge in classifying objects. This is a problem routinely encountered in biology. Imagine a scenario in which you need to learn from existing experimental data features that distinguish DNA sequence bound by different proteins and next classify novel DNA sequence based on which protein it will interact with. Machine learning techniques are very good at solving such problems. Especially now with the advent of deep learning machine learning in biology is more important and hotter than ever!

Ability to get things done with no structure, no clear objectives, and very little if any supervision.

earning how to learn (not just skills); techniques/technologies/skills become outdated soon- how do you pick up new ones when you need?