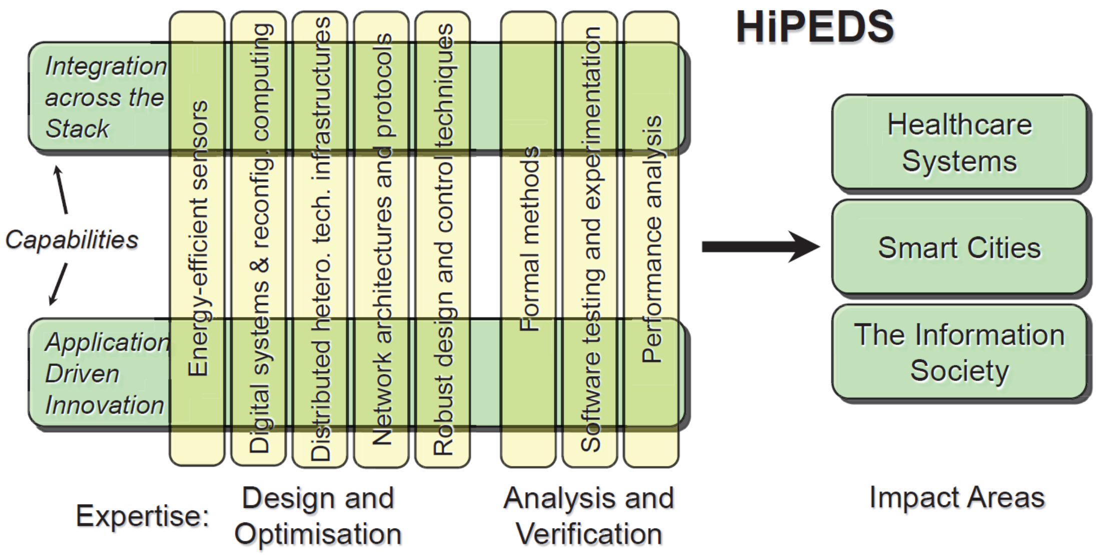
EPSRC Centre for Doctoral Training
in High Performance Embedded and Distributed Systems
(HiPEDS)

Imperial College London
Department of Computing
Department of Electrical and Electronic Engineering

Advisory Board Meeting 30th of November 2015



Student Project Narratives
2015

Contents

1. Jo Schlemper (Daniel Rueckert, DoC)

A computationally efficient framework for the joint segmentation and reconstruction for undersampled magnetic resonance (MR) imaging

Medical image analysis (e.g. segmentation) has traditionally been regarded as a separate process from image acquisition and reconstruction, even though its performance directly depends on the quality and characteristics of these first stages of the imaging pipeline. Adopting an integrated acquisition-reconstruction-segmentation process can provide a more efficient and accurate solution. In this project we propose to develop a computationally efficient framework for the joint segmentation and reconstruction for undersampled magnetic resonance (MR) imaging.

Medical imaging technologies such as Magnetic resonance imaging (MRI) produce highly detailed images with excellent soft tissue contrast. In many cases images are not an end in themselves, but rather a means of access to clinically relevant parameters which are obtained as post-processing steps, such as tissue characterisation/quantification. Acquiring the data necessary to reconstruct MR images is a time consuming process that can impose significant demands on the patient, but at the same time a lot of information is discarded during post-processing. In cases where the clinically relevant parameters sought are known a priori, the design of image acquisition and reconstruction would ideally be application-driven, such that they are tailored to the information necessary to determine them within some reliability standards.

Compressed sensing (CS), in combination with concepts from machine learning, has emerged as an effective way of reducing acquisition time. Incoherently acquiring a fraction of the data normally needed, perfect recovery is possible imposing a sparsity condition on the image, hence allowing scan acceleration without disrupting quantitative analysis. However, it is possible that better measurements could be achieved from fast acquisition data by directly focusing on the final analysis to be performed and treating any reconstruction as an enabling step rather than a distinct endpoint to be achieved first. This will help to establish a new paradigm in image reconstruction and analysis which is *application-driven* MR and can therefore be tailored to clinical applications such as the early detection of Alzheimer's disease.

2. Miten Mistry (Ruth Misener, DoC)

Integrating Mixed Integer Nonlinear Programming and Satisfiability Modulo Theories for Next-Generation Optimisation Algorithms

Mixed Integer Nonlinear Programming (MINLP) and Satisfiability Modulo Theories (SMT) have complementary weaknesses and strengths for solving difficult optimisation problems. This project will develop new foundations, algorithms, and tool prototypes exploiting the complementary benefits of MINLP and SMT to advance optimisation scalability and precision. This project is interesting because it opens up the possibility of incrementally solving difficult design problems; this avenue corresponds directly to the HiPEDS goal of establishing "connections between optimisation and verification" because the current state-of-the-art in computer-aided verification typically uses SMT technology whereas optimisation typically uses algorithms such as MINLP. The primary application area of the project is expected to be the Information Society, with possible applications also in Healthcare Systems.

3. Daniel Castro (Ben Glocker, EEE)

Machine Learning for Medical Image Analysis in Large-Scale, Multi-modal Image Databases

The aim of this project is to develop algorithms for automated analysis of medical images in the context of large-scale, multi-modal population databases (e.g., UK Biobank). In this context, novel machine learning algorithms are developed that scale to large databases and allow, for example, propagation of annotations between images. Annotations can correspond to any local or global semantic information, such as per-voxel structural labels, phenotypes, or non-image information such as diagnostic categories. The main challenge is to solve the problem of subject-specific atlas selection, i.e. for a new unlabelled dataset efficiently finding a labelled dataset (an atlas) in a multi-modal, heterogeneous database from which information can be successfully propagated. The atlas and the new image should be similar in a sense that they belong to the same population sub-group. Identifying sub-group membership for a new unlabelled dataset represents an unsupervised clustering problem, which needs to be tackled using high-performance and distributed algorithms. It is also important to estimate a level of confidence for the accuracy of the propagation in the absence of ground truth, such that the newly labelled datasets can serve as atlases for subsequently acquired data, and eventually serve as clinically useful information in tasks such as diagnosis, therapy and intervention.

4. Andrei Lascu (Andrew Donaldson, DoC)

Verifying compilers for multi-core programming languages

Andrei Lascu's PhD project fits the "analysis and verification" theme of the CDT. The plan is for his work to investigate novel methods for improving confidence in compilers and runtime systems for many-core programming languages, using testing and formal verification. A particular challenge which Andrei will focus on during this research is that of testing implementations where there is uncertainty or a permitted envelope of possibilities in the results that a program might compute. This is especially relevant in high-performance graphics applications, where there is significant scope for deviation from the real-valued result that a function approximates, and no formal semantics with which to build a high-confidence analyser. Nevertheless, the notion of a compiler or driver bug that produces the wrong image is clear, and there is a need for automated tools to help diagnose such problems. This aspect of the compiler correctness problem is what differentiates the approach Andrei is taking from existing approaches to compiler analysis.

5. Sanket Kamthe (Marc Desienroth, DoC)

Extracting intentions from high velocity data streams

Sanket Kamthe is working at the intersection of statistical machine learning and neurotechnology. Sanket's PhD project centers around the extraction of intentions from high-velocity human data streams (e.g., EEG signals), which are naturally very noisy. The inferred intentions are then used for real-time control of virtual and real robots in the context of personalized healthcare. The overall system will reduce the training time of disabled patients (e.g., amputees or paralyzed people) compared to the current state of the art and allows for an unprecedented level of personalization. This project fits nicely into the realm of the HIPEDS CDT because it combines robust inference and control techniques and large-scale computations for online data processing and analysis. Due to its relevance to personalized healthcare and the information society, this project directly addresses two of the three main impact areas of the CDT.

6. Giannis Evagorou (Thomas Heinis, EEE)

Efficient algorithms for large scale spatial data

The project Giannis Evagorou will carry out addresses the need for new algorithmic support in the analysis of large amounts of spatial data. Today's methods do not scale to the enormous amounts of data we are witnessing today. This is particularly true for the spatial data Giannis is working with which originates from collaboration with Transport for London (TfL). To assess impact of weather on the road network, particularly travel times, large-scale spatial join operations and spatial analyses need to be performed that as of today cannot be executed efficiently. Despite decades of or related work, the state of the art cannot efficiently deal with data at this scale and of this complexity. Giannis will consequently design novel algorithms to execute large-scale spatial joins between complex road networks and weather data. Working on optimising the delivery of road space as Giannis' does consequently directly contribute to Smart Cities, one of the CDT's impact areas.

7. Davide Cavezza (Dalal Alrajeh, DoC)

Repairing Software Specification

The detection of errors, their causes and suitable repairs in distributed systems is a complex problem, owing to their non-sequential nature. Though much effort has been spent on developing formal, automated methods to support error detection, rigorous, general mechanics for explaining their causes and suggesting alternative repairs is yet an open challenge. Recently, there has been a surge in the application of machine learning to provide some form of support for repairing software specification. However, their deployment has been limited. Their use typically depends on the ability to provide suitable translations from software specifications languages and semantics to languages and semantics understood by the machine learning environment. Though recent applications have shown promising results for some instances, their progress has been hindered by the constraints of operating within languages and semantics of the learner that are not necessarily best suited for capturing the full scope of the software models. This PhD is focused on (1) developing a formal underpinning of the classes of explanation and repair problems for specifications (described Linear Temporal Logic) that may be resolved through learning, and (2) designing suitable learning-based algorithms and tools for repairing temporal specifications.

8. Andrea Paudice (Emil Lupu, DoC)

Anomaly detection and diagnosis in large scale distributed environments

Detecting anomalies and in particular signs of compromise in large scale distributed systems and cloud environments is particularly challenging yet necessary in order to protect the security of the system from malicious threats whether by outsiders or insiders. The size, heterogeneity and architecture of the system requires techniques that can correlate information across the system layers and that can adapt to requirements and the level of threats. The information streams to analyse are nevertheless very large so any algorithms design will need to take into account performance and overhead induced aggregating information where necessary. This fits in well with the objectives of the *design and optimisation* research theme and in particular aspects relating to the design and implementation of *distributed heterogenous technical infrastructures* as well as with the *analysis and verification theme* and in particular aspects linked to *distributed systems experimentation.* A related problem that we also aim to investigate is the robustness of anomaly detection and diagnosis in the face of malicious behaviour, that is where the attack seeks to undermine and perhaps use for malicious purposes the anomaly detection algorithm itself by modifying their decision boundary. Building and evaluating techniques for improving the robustness of anomaly detection in large scale environment to malicious compromise also requires us to engineer systematic tests at scale that can reveal new vulnerabilities. Again this fits in well with the *distributed systems experimentation* theme. Finally, at large scale anomaly detection and diagnosis need to be adaptive to balance detection accuracy against performance overhead. This also integrates well with some of the *adaptation aspects of* the HiPEDs. CDT.

9. Anastasios Andronides (Cristian Cadar, EEE)

Testing and verification in multicore platforms

Anastasios' PhD project fits within the Analysis and Verification theme of the HiPEDS centre for doctoral training. In particular, his project is motivated by the emergence of parallel platforms such as multicore CPUs, which provide an abundance of often idle resources that could be successfully exploited for testing and verification. To take advantage of such idle resources, one could run the native version of an application in parallel with several other versions instrumented to find certain classes of bugs, e.g., memory or concurrency errors. This would enable one to apply testing and verification techniques in production, making it possible to catch bugs that are not revealed during the offline testing stage. In order to achieve this ambitious goal, we would need to address several difficult challenges, such as synchronising the execution of the native and instrumented version while minimising the overhead incurred by the native version, and effectively dividing the testing effort across the instrumented versions. Besides deployment-time testing and analysis, Andronidis's PhD project will explore other mechanisms that exploit idle resources to improve software reliability and security.

10. Ilya Kisil (Danilo MAndic/ Wayne Luk, EEE/DoC)

Efficient representations for big data analytics

This project will look into tensor decompositions for big data analytics, with a particular emphasis on bringing the solutions to the stage which is ready fo hardware implementation. This will, on the one hand, make possible efficient compressed representations of Petabites of data on a standard computer, while on the other hand using efficient parallel representations to split the data and computational effort into smaller, "easier to swallow" chunks.

11. Effrosyni Simou (Danilo Mandic, EEE)

Reverse engineering of human auditory cortex

We will use the in-ear EEG for 24/7 monitoring of the brain responses to the auditory scene. Particular attention will be paid to the perception of competing sounds in a cluttered environment. A successful solution will be important for hearing aids wearers as it will enable hearing aids to automatically adjust their parameters according to the complexity of the surround sound field.

12. Tianjao Sun (Paul Kelly, DoC)

Domain-specific optimisations for solving partial differential equations

Numerical solution of partial differential equations (PDEs) is fundamental to a huge range of applications in computational science, engineering and medicine. This PhD project concerns the development of tools that automate the generation of numerical PDE solvers from a high-level specification - essentially weaving together the PDE itself with a specification of how the continuous fields are discretised. Automation creates opportunities beyond just a productivity advantage: firstly we can automate sophisticated performance optimisations, and secondly we can automate the search for the best solution technique and implementation choices for a given problem on given hardware. The goal is to achieve higher performance than can reasonably be achieved by hand, on HPC and low-power hardware, while enabling flexible selection of alternative models, numerical methods, discretisations and implementation techniques. This work is in close collaboration with researchers modelling aeroplanes, engines, racing cars, arteries, seismography, the atmosphere, oceans, and tidal turbines.

13. Domagoj Margan (Natasa Przulj, DoC)

Extracting knowledge from large-scale biological networks

Large and complex "Big Data" networks have been used to model and analyse various biological and medical phenomena, such as protein-protein interactions, metabolic reactions, functional associations of genes, drug-protein bonds, and structural or side-effect similarities of drugs. Each type of a network provides a specific perspective on a different slice of biomedical information and, as such, each network plays an important role in describing the system's cellular functions and complex relations that occur between various biological entities. The main aim of the PhD is the development of novel methods for the extraction of knowledge from large-scale biological networks, with a particular emphasis on the machine learning aspects of network data integration and fusion. The PhD will encompass the exploration of machine learning methods for collective and integrated mining of different biological networks and the construction of a unified model. The problem of fast and efficient large-scale network analysis opens up questions in high-performance computing, which necessitate the creation and improvement of computational techniques designed for parallel processing on distributed multi-core architectures. The PhD will develop and apply these to open biomedical problems.

14. Matt Douthwaite (Pantelis Georgiou, EEE)

Fault tolerant and robust embedded wearable devices

Matthew's area of research will be focused on developing CMOS ASIC systems for monitoring and controlling chronic health conditions, overcoming some of the challenges in integrating wearable systems on patients, by making them more robust, discreet, power efficient and connected. To do so we will need to develop a new methodology using embedded systems to guarantee autonomy of the device and fault tolerance and robustness when being used in free-living environments.