The Warwick Postgraduate Colloquium in Computer Science 2012

Programme Committee Chair
Steven Wright

Programme Committee
Shahzad Malik Awan
David Beckingsale
Robert Bird
James Davis
Jonathan Foss
Henry Franks
Andrew Mallinson
John Pennycook
Oliver Perks
Joshua Scotton

29th June 2012

Schedule

Session 1 (09:00 − 10:20)

09:00 − 09:20	Next Generation Digital Libraries	Fault Tolerant based Feedback Control Scheduling for Real Time Embedded Systems
	Saba Khalil Toor	*Oumair Naseer*
09:20 − 09:40	Adaptive Hypermedia in On-line Advertising	On Energy-aware Mobile Workflow Offload with Cloudlets
	Dana al Qudah	*Bo Gao*
09:40 − 10:00	Overview of Mobile Learning: Definitions, Advantages and Barriers and Influence Factors	Validation of Codesign-Enabling Simulation Frameworks
	Maram al Zaidi	*Wilson Tan*
10:00 − 10:20	Exploring Participatory Design for SNS-based AEH System	Stabilizing Information Dissemination in Wireless Sensor Networks
	Lei Shi	*Sain Saginbekov*

10:20 − 10:50   **Coffee (Common Room)**

Session 2 (10:50 − 12:10)

10:50 − 11:10	A Gamma-Gaussian Mixture Model for Detection of Mitotic Cells	Hand Gesture Recognition Based on SURF Tracking and LDCRF
	Adnan Khan	*Yi Yao*
11:10 − 11:30	Fusion of Thermal and 3D Stereo Image	Occluded Face Recognition using Dynamic Class Warping
	Shan-E-Ahmed Raza	*Xingjie Wei*
11:30 − 11:50	Performance Comparison of Advanced Image Coding Schemes	Random Subspace Method for Gait Recognition
	Xin Lu	*Yu Guan*
11:50 − 12:10	Automated Segmentation and Tracking of Focal Adhesion Dynamics in Fluorescene Microscopy Sequential Image	Facial Expression Recognition
	Guannan Li	*Emmanuel Ige*

12:10 − 13:30   **Lunch (Common Room)**

Session 3 (13:30 − 14:50)

13:30 − 13:50	A multilingual Mobile Learning Tool for South African High Schools	Evaluating Co-Array Fortran as an alternative to MPI for Physics Simulation Applications in HPC Environments
	Mmaki Jantjies	*Andrew Mallinson*
13:50 − 14:10	An active learning approach for design and validation of software fault-tolerance	Optimisation of Patch Distribution Strategies for AMR Applications
	Fatimah Adamu-Fika	*David Beckingsale*
14:10 − 14:30	Using Empirical Modelling to Build Parsers	LDPLDS: Improving I/O Performance Without Application Modification
	Hui Zhu	*Steven Wright*
14:30 − 14:50	Checking Equivalence of Quantum Programs and Circuits	Performance Modelling of Magnetohydrodynamics Codes
	Ebrahim Ardeshir Larijani	*Robert Bird*

14:50 − 15:10	**Coffee (Common Room)**

Session 4 (15:10 − 16:30)

15:10 − 15:30	Using Data Mining Algorithms to Gain Insights into Digital Crime	Accelerating Molecular Dynamics with Intel Xeon Processors and the Intel MIC Architecture
	Nentawe Gurumdimma	*Simon John Pennycook*
15:30 − 15:50	The Longest Queue Drop Online Policy	Context-Aware Light-Weight Benchmarking for P2P Computing
	Nicolaos Matsakis	*Shahzad Malik Awan*
15:50 − 16:10	Allocating Resource for Workflows Running Under Authorisation Control	Determining Influence in Complex Network Structures
	Nadeem Chaudhary	*Henry Franks*
16:10 − 16:30	Improved Bounds on Satisfaction Problems	Road Type Classification
	Alexander Wilson	*Philip Taylor*

16:30 − 17:00	**WPCCS Closing Talk and Prize Presentation (CS1.04)**

Session 1 (09:00 – 09:20)

Next Generation Digital Libraries

Saba Khalil Toor

Abstract

Libraries are part and parcel for educational excellence and knowledge. Existence of libraries date back to 2550 BC and its evidence is found in the city of Ebla (Syria) where nearly 2000 books in the form of clay tablets have been discovered. Several historic libraries within Europe and elsewhere have existed including the great library of Alexandria. This tradition of libraries has continued and in this era of cyber space it is flourishing with greater strength and impact.

Although libraries in their physical form exist, the digital form of libraries is becoming exceedingly popular because of its accessibility and ease of use over the internet. Libraries of several universities have their catalogues online. There are services such as 'web of science', 'Web of knowledge', 'CiteSeer' etc. that provide online information about journal articles and books for research work. These sites can be accessed using various modes and protocols.

Our main research work is related to finding the best model for students to do the references and citation task. For this we need information from digital libraries and next generation catalogs. This information is extracted through software that we are developing for the purpose of evaluation of the references and citation task. In this presentation the rational approach and various pitfalls encountered during the development of this software will be discussed, and we will indicate how our research questions are answered through this approach.

Furthermore, this presentation will highlight various protocols that are being currently adopted by various next generation catalog systems and digital libraries, such as Z3950 protocol which was developed by the library of congress, SRU (Search retrieve through URL), SRW (Search retrieve through Web) etc. Bibliographic data returned by searches through these protocols over internet will also be discussed briefly in the presentation.

Fault Tolerant Based Feedback Control Scheduling for Real Time Embedded Systems

Oumair Naseer

Abstract

In order to provide Quality of Service (QoS) in open and unpredictable environment, Feedback based Control Scheduling Algorithm (FCSA) is designed to keep the processor utilization at the scheduling utilization bound. FCSA maintains the desired CPU utilization by avoiding overload, meeting deadlines even if the task execution time is unpredictable and through performance control feedback loop. Current FCSA doesn't ensure Fault Tolerance (FT) of safety critical applications while providing QoS because the design methodology of FCSA is based on the separation of the concerns. These concerns are built on the assumptions that the feedback controllers can be designed by assuming the fixed predefined mapping, fixed time period and hard deadlines. In Real Time Embedded Systems dependability is achieved by adapting one of the classic FT schemes. This research presents a novel approach on integration FT schemes and FCSA for real time embedded systems in the following steps;

1. Defining:
   (a) System Architecture Model.
   (b) Feedback based Control Scheduling Model.
   (c) Processor Scheduling Model.
   (d) Fault Tolerant Scheme Model.

2. Integrating Fault Tolerant Scheme Model with overall System, and

3. Analyzing the Stability of the Integrated System.

This procedure is especially important for control scheduling co-design of real time Embedded Systems.

Session 1 (09:20 – 09:40)

Adaptive Hypermedia in On-line Advertising

Dana al Qudah

Abstract

Adaptive hypermedia is the process of tailoring the content presented based on users' knowledge, capabilities and interests. It was first introduced by Peter Brusilovsky who had deployed his knowledge in the field of e-learning. In e-learning researchers look at creating adaptive content for students based on their personal needs and capabilities.

My research involves another application that is gaining more and more importance over time which is e-business. The Worldwide Web has changed dramatically over the last 20 years giving users more power to perform all types of transactions over it. As a result that's e-business has blossomed and started reaching to consumers over the web using different approaches and models.

In the field of e-business it is crucial that the return of investment is well-formed and the expectations meet the actual profits. In order to achieve that, other models of business have been derived under the umbrella concept of e-business. One of these areas is advertising. On-line advertising is the process of delivering a marketing message using the World Wide Web to attract more consumers .Psychology theories prove that consumers tend to go through at least three different stages when they are exposed to ads, before they actually attempt to complete a sale. It all starts with the exposure to the ad, then cognition or judgment, then the effect of the ad on the consumer which encompasses their attitude to finally make the decision of purchasing or not. My work digs into the creating of personalised ads using adaptive hypermedia techniques to increase the consumers' interest towards completing the sales transaction. The reason to use adaptive hypermedia techniques is that research showed that banner advertising are not attracting consumers as much as it used to be while simple personalised ads have been guaranteed a higher number of clicks.

Also it is essential that personalised ads are differentiated from recommendation systems, as it is expected that they vary on the algorithms and techniques.

This is done by designing a theoretical framework that helps in understanding the different perceptions and approaches to add value and knowledge to online advertisement and then implementing an application that represent this framework and test the proposed theoretical framework.

On Energy-aware Mobile Workflow Offload with Cloudlets

Bo Gao

Abstract

Cloud computing and mobile computing are two of the most influential technologies that look set to change the face of computing in the coming years. Combination of the two provides us with an unprecedented opportunity to provide highly portable and yet content-rich and computation-intensive services to the end user. In this paper we investigate the possibility to use code/task offload techniques between mobile and cloud in order to reduce the energy cost of workflows deployed on mobile devices. In the paper, we first weave a vision in which mobile devices are coordinated over a network, which is equipped with a layer of cloud-like infrastructures called cloudlets whose computational resources can be leveraged by the mobile devices, to host the execution of mission-critical mobile workflows in an energy-aware manner. We then build a model which encompasses various characteristics of the workflow's software and the network's hardware devices. With this model we construct the objective functions that guide the offload decisions. We then present a simple heuristic algorithm that produces offload plans according to these objective functions and their variations both statically and dynamically. We conclude this paper with a series of simulation studies the results of which give an insight into the offload-ability of workflows of different characteristics. The results also illustrate how different hardware specifications can affect offload efficiency. These studies indicate that our offload algorithm can significantly improve the energy efficiency and execution speed of a mobile.

Session 1 (09:40 – 10:00)

Overview of Mobile Learning: Definitions, Advantages and Barriers and Influence Factors

Maram al Zaidi

Abstract
The development of mobile learning (hereafter referred to as m-learning) in many developed countries such as the United Kingdom and the United States has led to remarkable changes in educational technology. This development has become an essential tool not just for communication but also for studying. It now offers great advantages for both learners/students and lecturers. Despite this current educational revolution of m-learning around the world, Saudi m-learning falls far behind in this respect. Thus, there is an urgent need to investigate and promote m-learning in Saudi Arabia as it becomes increasingly necessary to deliver education to students/learners at any time and even beyond the classroom. M-learning can provide the easiest way to access materials or practice activities related to the learning process. However, successful implementation of Saudi m-learning requires first thorough exploration and understanding of the influence factors of adoption/accepting m-learning in higher education in this case specifically in the Saudi Arabian context. Despite this, upon surveying the current literature, there is a noticeable lack of understanding of the driving factors regarding the adoption of m-learning from learners'/ students' perspectives in Saudi universities. Therefore, this paper provides a historical review by summarizing m-learning, its definitions, its advantages and its barriers and influence factors in many regions in the world with a view to better understanding of the Saudi situation. This, it is hoped, will provide a framework which is directly applicable to the specifics of m-learning in Saudi education.

Validation of Codesign-Enabling Simulation Frameworks

Wilson Tan

Abstract
Exascale computing is necessary for us to be able to deal with the rising amount of data that is being generated by businesses and to advance frontiers of scientific knowledge. Unfortunately, the techniques that got us to the level of petascale computing would no longer suffice to take us to exascale; to reach exascale, new paradigms and new approaches must be developed, and developed quick. One promising approach is codesign, which was originally utilized in the domain of embedded computing.

Codesign, in the simplest sense, is the simultaneous design and optimization of the hardware and software component of a system. In contrast, most hardware systems today are designed first, and the software that could run well(or be optimized to run well) for the hardware are then found. Bringing codesign to the domain of high performance computing is not without obstacles however: tools must be developed that could effectively bridge the disparate disciplines of hardware and software design. These tools include simulation frameworks that span both applications and hardware systems. Our work to date has focused on the determining the validity, accuracy and limitations of simulation frameworks(SST-Macro and WARPP) that could be used(and in some cases, had already been used) for HPC hardware-software codesign. We have already identified a number of weak points in the two simulator frameworks that could limit their usefulness and accuracy; in some cases, we have also begun working on the fixes ourselves.

The most significant result of our work so far is our discovery/realization that one of the key assumptions underlying the operations of SST-Macro and WARPP, the existence of a Computation-Communication Divide, may actually have limited validity for at the level of the node. The Computation-Communication Divide states that a message passing program could be segmented into computation and communication sections which could then be simulated independently of one another. While a reasonable assumption for purely Internode setups(setup where the processing elements are separated by a medium such as Ethernet), it is no longer accurate for Intranode setups, since the "medium" being used, which is shared memory, is not exclusively used for communication: it largely figures in computation as well. At the level of the node, the Computation-Communication Divide is blurred. We are currently determining how much inaccuracy this "proven-to-be-invalid" assumption contributes to application simulations, and working on a new Intranode communication model that is more reflective of real world systems.

Session 1 (10:00 – 10:20)

Exploring Participatory Design for SNS-based AEH System

Lei Shi

Abstract

The rapidly emerging and growing Social Networking Sites (SNS) offers an opportunity to improve adaptive e-learning experience by introducing a social dimension, connecting learners within the system. Making connections and providing communication and sharing tools can engage students in creating an effective learning environment and enriching learning experiences with the collaboration and feedback from instructors and their peers. Researchers have been working on introducing social networking features into adaptive educational hypermedia (AEH) systems. The next stage research is centered on how to enhance social networking facilities of SNS-based AEH systems, in order to engage students' participation in collaborative learning and generating and/or enriching learning materials. Students are the core participants in the adaptive e-learning process, so it is essential for the SNS-based AEH system designers to consider students' opinions. This paper aims at exploring how to apply participatory design methodology in the early stage of SNS-based AEH system design.

Stabilizing Information Dissemination in Wireless Sensor Networks

Sain Saginbekov

Abstract

Wireless Sensor Network is a network of small devices called sensors. Sensors are different from other wireless devices mainly with their size, processing capability, power and memory. They are small in size and have small memory. They have limited processing power and small power supply. One important component of network reprogramming is code dissemination, when the new program code is distributed to the relevant nodes. Very few information dissemination protocols, if any, tolerate transient data faults, i.e., faults that corrupt the memory state of nodes. We address this limitation by proposing a novel protocol, called Repair, that transforms any fault-intolerant information dissemination protocol into a stabilizing protocol where, eventually, all nodes obtain the updated code. We conduct simulations to show the performance of Repair, and we integrate Repair with existing reprogramming protocol called Varuna, and our results show that Repair induces low overhead on Varuna, and causes all nodes to receive the new code. Our main contribution is the first corrector protocol that stabilizes information dissemination in the presence of transient faults.

Session 2 (10:50 – 11:10)

A Gamma-Gaussian Mixture Model for Detection of Mitotic Cells

Adnan Khan

Abstract

Detection of Mitotic Cells (MCs) in breast histopathology images is one of the three components, the other two being tubule formation and nuclear pleomorphism, required for developing computer-assisted grading of breast cancer tissue slides. This is very challenging since the biological variability of the MCs makes their detection extremely difficult. Existing methods for mitosis detection either use an additional dye (e.g. PHH3) to stain MCs exclusively and detect exclusively stained MCs in the images or use a video sequence to detect mitotic events over time by incorporating spatial and temporal information. Since the exclusive stain costs additionally and videos are not used in standard histopathological practices, therefore, a gap exists in the literature.

Our proposed MC detection system mimics a pathologists top-down approach to MCs detection under microscope. The system consists of the following components: (1) Isolate tumor regions from non-tumor regions (lymphoid/inflammatory/apoptotic cells) using an unsupervised tumor segmentation algorithm, which we term RanPEC; (2) Search for potential MCs in the tumor regions by modeling the pixel intensities in mitotic and non-mitotic regions using a Gamma-Gaussian mixture model; and (3) Context evaluation of potential MCs: in order to enhance precision of the system. Textural features in a small window around candidate MCs are used to accept or reject the candidates.

The system has been validated on 35 H&E stained Aperio images, belonging to 5 different slides, containing more than 200 MCs. The experimental results show that the proposed system achieves a high recall of 0.82 with precision of 0.29. Employing CAPP on these results produces 100% increase in precision at the cost of less than 15% decrease in recall.

Hand Gesture Recognition Based on SURF Tracking and LDCRF

Yi Yao

Abstract

The process of communication is to transfer information from one to another. Naturally, body language is a important component of human communication, especially upper body gesture, namely, hand gesture. Imagining all machineries and computers can understand our gesturing and body language. There is no longer need for keyboards, buttons, handles, mice and any other human machine interaction hardware. Making the way that people control machinery more intuitive and convenience is one of the primary objectives in computer science community. Through the image and video capturing devices like digital camera, web cameras, stereo cameras, surveillance cameras and other kinds of sensors including laser scanners and infrared sensors, programme can see the world just like we do. Analysis can be performed using image processing, pattern recognition techniques to recognise the meaning of the input human hand gesture. Even though automatic hand gesture recognition technology has been applied to real-world applications with relative success, there are still several challenges remain, including solving skin-like region in the background problem, hand-like region moving in background problem, background with complex textures problem. We propose a novel method using SURF based tracking to solve distractions in the background. In the input video, skin color detection used to locate possible hand like regions in the first frame. Then in the rest of the video sequence, SURF feature is extracted from every frame to detect moving hand-like region. The hand trajectory when extracted as feature for the later classification. Our method uses Conditional Random Field for classification.

Session 2 (11:10 – 11:30)

Fusion of Thermal and 3D Stereo Image

Shan-E-Ahmed Raza

Abstract

Thermography can be used as a non-destructive method for stress and early disease detection. Infrared thermometers have been used since the early 1980's by researchers to determine the temperature differences in plants and different parts of canopy for irrigation scheduling purposes. The development of thermal imagers has extended the opportunities for analysis of thermal properties of plants and canopies. Various researchers have used thermal imaging to determine water stress and for disease detection. Different stress indices can be found in the literature and researchers have conducted experiments to find a relationship between stress indices and temperature values found by thermal imaging. Thermal imaging has been tested in different environmental conditions and the conditions best suited for this kind of imaging have been explored. Plant genotyping and phenotyping have also been studied with the help of thermal images.

However, the temperature profile of a region in a thermal image is largely affected by leaf angles and the distance of the region from the thermal sensor. Stereo vision helps to create a depth map of a region under observation and when combined with thermal imaging or other sensors can help to get a better estimate of variation in the thermal profile. Here we report results of the fusion of thermal and 3D stereo images of an anthurium plant. Our analysis shows the resulting visual stereo image along with overlaid thermal image. The result shows that the regions of the plant which are closer to the thermal imaging sensor show higher temperature than the regions which are at a distance. Also the leaves which are at an angle appear to be cooler in the thermal profile. This shows that angle and height variation can help to correct for temperature variations. Aside from this, we have also obtained thermal images from plants undergoing different water stress treatment. The images were analysed using extreme value theory and the results show that the images from different type of treatments can be distinguished on the basis of their statistical parameters.

Occluded Face Recognition using Dynamic Class Warping

Xingjie Wei

Abstract

Face recognition (FR) in real-world environment often has to confront uncontrolled and uncooperative conditions. The collected face data is easily occluded by sunglasses, carves, hats, shadows or other objects in front of a face, which introduces more intra-class variations and degrades the recognition performance. For FR based on still images, the difficulty caused by occlusion is twofold. First, occluded pixels usually cause a high distance between two images from the same person. Second, the occlusion may be similar with others' facial features, for example, a scarf and a full bread, which can misguide recognition by computers. Holistic methods such as PCA and LDA are not applicable since all the extracted features will be corrupted by occlusion. If the occluded area can be well detected, accurate recognition can be achieved using only the unoccluded area. However, in real environment, the sources of occlusion are varying. The location, size or shape of occlusion is unknown leading it is difficult to simply segment the occlusion parts from a face image.

There are two categories of approaches to handle occlusion related problem. The first one treats occluded FR as a reconstruction problem. A clean image is reconstructed from an occluded image by a linearly combination of gallery images. Then the occluded image is classified as the class with the most accurate reconstruction. Those approaches usually require sufficient samples per person in the gallery set in order to cover the variations might occur in the probe set, which is not always available especially in the real environment. The other category is patch-matching based approaches. An occluded image is first partition into sub-patches where features are extracted. Then matching is performed based on the similarity of each patch-pair. In order to decrease the large matching error due to occlude patches, different weights calculated through training are given to different patches. But the training process is usually data-dependent. The weights needs to be trained on specific data set that they deal with.

We propose a patch-matching method Dynamic Class Warping (DCW) to deal with occlusion. A face image is partition into overlapping sub-patches followed by being concatenated as a sequence. DCW calculates the Image-to-Class distance from a probe sequence to all the gallery sequences of a given class. Compared with other methods, neither occlusion detection nor training process is needed in our method and our method achieves excellent recognition rate on public databases.

Session 2 (11:30 – 11:50)

Performance Comparison of Advanced Image Coding Schemes

Xin Lu

Abstract
Three advanced image and video coding schemes, JPEG2000, Scalable Video Coding (SVC) and Wavelet-based Scalable Video Coding (WSVC) are compared. In order to gain an insight into the application scenarios of the three coding techniques, the coding performance of JPEG2000 and both SVC and WSVC's intra frame coding tools are evaluated on still images with a variety of resolutions and characteristics.

JPEG2000 is the first still image compression standard based on the wavelet transform. Due to blocking artifacts, use of the Discrete Cosine Transform (DCT) was rejected in favour of the Discrete Wavelet Transform (DWT). The core encoding tools of JPEG2000 include the DWT and Embedded Block Coding with Optimized Truncation (EBCOT) for entropy coding. With these tools, JPEG2000 outperforms conventional JPEG in terms of compression performance by nearly 30%. JPEG2000 also provides an intrinsic multi-resolution characteristic which is useful in multimedia applications.

SVC is a video coding extension of H.264/AVC for the provision of temporal, spatial and SNR scalability. For intra frame coding, SVC employs spatial prediction on variously-sized blocks in the current frame, inter layer prediction, integer Discrete Cosine Transform (DCT) transform, and Context-Adaptive Binary Arithmetic Coding (CABAC) for entropy coding. Whereas intra frame coding does not reference information from neighbouring frames, in SVC, coding of the enhancement layers does utilize information from the base layer. Consequently, SVC intra frame coding can be used for still image compression.

Barbell-lifting wavelet-based SVC (WSVC) was proposed by Microsoft Research Asia (MSRA) in competition with H.264-based SVC during the call by the Moving Picture Experts Group (MPEG) for SVC proposals. It also provides a very good platform for continued research on wavelet-based technologies. The core technologies suggested in WSVC include spatial 2D DWT, motion compensated temporal filtering (MCTF) and Embedded Sub-band Coding with Optimized Truncation (ESCOT) based on Arithmetic Coding.

A variety of images of resolution CIF to HDTV were used in the comparative evaluation. The results show that at low bitrates SVC out-performs JPEG2000. The reverse is true at high bitrates. However, when the quality of the fully reconstructed picture is considered, WSVC out-performs both SVC and JPEG2000. In terms of computational complexity, JPEG2000 has a clear advantage over both SVC and WSVC, and is therefore more suited for high definition picture compression. In terms of spatial scalability, performance evaluation of which has not previously been reported in the literature, at high bitrates SVC achieves a better rate-distortion performance than both JPEG2000 and WSVC, regardless of picture resolution and content.

Random Subspace Method for Gait Recognition

Yu Guan

Abstract
Overfitting is a common problem for gait recognition algorithms when gait sequences in gallery for training are acquired under a single walking condition. In this paper, we propose an approach based on the random subspace method (RSM) to address such overlearning problems. Initially, two-dimensional Principle Component Analysis (2DPCA) is adopted to obtain the full hypothesis space (i.e., eigenspace). Multiple inductive biases (i.e., subspaces) are constructed, each with the corresponding basis vectors randomly chosen from the initial eigenspace. This procedure can not only largely avoid overadaptation but also facilitate dimension reduction. The final classification is achieved by the decision committee which follows a majority voting criterion from the labeling results of all the subspaces. Experimental results on the benchmark USF HumanID gait database show that the proposed method is a feasible framework for gait recognition under unknown walking conditions.

Session 2 (11:50 – 12:10)

Automated Segmentation and Tracking of Focal Adhesion Dynamics in Fluorescene Microscopy Sequential Image

Guannan Li

Abstract

Focal adhesions (also called cell-matrix adhesions) are subcellular macromolecules that play essential roles in many important biological events including cell motility, cell proliferation, cell differentiation, regulation of gene expression and cell survival. Fluorescence microscopy imaging of focal adhesions is the primary method used to understand focal adhesion dynamics. Data acquisition of the focal adhesion dynamics is generally done by manual segmentation and tracking of focal adhesion. The performance of manual work can be significantly hurt by the long time required in large data set. Also manual segmentation and tracking is highly time consuming and onerous. In my recent work, a sequential image processing and object tracking system is developed for automated segmenting, tracking and data extraction of focal adhesion. The system coordinates several image processing approaches in segmentation processing and focal adhesions can be successfully recognized in each image frame. The dynamic of each focal adhesion in the sequential image is tracked via a novel object tracking algorithm introduced in my work. The system is developed in Matlab and its GUI offers many useful tools that helps user to better understand and process the focal adhesion dynamic.

Facial Expression Recognition

Emmanuel Ige

Abstract

The human ability to detect and interpret faces and facial expressions in a given scene poses little or no effort. Up till date, the development of reliable machine expression recognition system is still a challenge. In a facial expression recognition system there are several related problems, for example, detection of face as an image segment, extraction of information related to facial expression and expression classification (i.e. categorization of emotions). However, researchers are striving hard to achieve steps towards human-like interaction between man and machine, that can perform these operations accurately and in real-time. The advancement recorded in area of image analysis and pattern recognition has paved way for the automatic detection and classification of emotional and conversational facial signals (Kaur and Vashisht, 2011). The Facial Action Coding System (FACS) link facial changes with muscle actions that produce them, the major problem with this system is that: a) it is a human-observed based system and b) it needs experts to manually perform FACS coding frame by frame. The focus of my talk will be on machine recognition of facial expressions exploring the robustness of temporal dynamics of different expressions between genuine and fake expressions.

Session 3 (13:30 – 13:50)

A Multilingual Mobile Learning Tool for South African High Schools

Mmaki Jantjies

Abstract

Inadequate learning material and resources available to learners in some South African schools has prompted the investigation of mobile phones as potential learning tools. The increasing accessibility of mobile phones has influenced their widespread use in various socio-economic communities, as affordable means of basic communication and technology. This has also influenced their potential as a contributory solution to South African education challenges. Language has also been identified as an obstacle in the education system, as most of the South African population is not first language speakers of the instructional language. Switching between two languages (code-switching) to converse is a popular method of communication in South African schools, and this is also attributable to the country's diverse language community. A learner's inability to adequately communicate in English consequently contributes to their poor performance in school. This paper examines the development of a mobile learning tool that supports bilingual education through bilingual learning material and also creates a pervasive learning environment for the learners. In the research project learners were provided with low cost mobile phones and access to a mobile bilingual learning tool with content that supports their curriculum. Through surveys, learners from four South African public schools were subsequently assessed on their use of a multilingual mobile learning tool that enables to access adequate learning material in their language of choice.

Evaluating Co-Array Fortran as an alternative to MPI for Physics Simulation Applications in HPC Environments

Andrew Mallinson

Abstract

As HPC (High Performance Computing) systems continue to evolve in an era in which Moore's law continues to hold but CPU (central processing unit) clock speeds are at best static and in several cases decreasing. It will only be possible to realise performance improvements in future HPC systems by effectively harnessing the massive increase in concurrency available at the hardware level in these systems. Constructing HPC applications, which are able to take advance of the computational power available through this uplift in hardware concurrency will become increasingly challenging. Existing approaches mainly based on the dominant MPI (Message Passing Library) are already struggling to scale on present day systems and developing and maintaining them is becoming increasingly problematic.

Recognising this the DARPA (Defence Advanced Research Projects Agency) in the USA initiated a programme to address these challenges. The HPCS (High Productivity Computing Systems) programme developed several new languages mainly based on the PGAS (Partitioned Global Address Space) paradigm. These languages promise to improve programmer productivity whilst maintaining and potentially improving application performance and scalability.

Co-array Fortran (CAF) is a language extension developed to incorporate many of the PGAS concepts into the existing Fortran programming language. It was standardised into the main Fortran specification in Fortran2008. To date little work has been conducted into the applicability of CAF to the HPC applications used within the physics simulation community.

This talk will present experiences and experimental results obtained during the porting of an existing physics mini-application (CloverLeaf) to CAF. CloverLeaf will be introduced together with an overview of the CAF language itself, including how it differs from traditional message passing languages/libraries (eg MPI). Several different approaches for porting CloverLeaf to CAF will be evaluated. The success of each of these strategies, in terms of performance and scalability, will be assessed under both strong and weak scaling scenarios, relative to the performance of the existing MPI implementation. Experiment results on up to 240 processor cores of the Raven system at Cray Inc. will be presented. This is one of the few HPC platforms which is generally available and also has hardware support for the one-sided communication constructs required by CAF.

In addition to the experimental results I will also assess the applicability of CAF to improving HPC programmer productivity. The presentation will conclude with a discussion of the future directions of this work.

Session 3 (13:50 – 14:10)

An active learning approach for design and validation of software fault-tolerance

Fatimah Adamu-Fika

Abstract

Software-based injection consist of reproducing at software level the errors that would have been produced upon occurring faults in the hardware. Usually, the most frequently occurring and most difficult to detect are transient faults. Flipping program variable bits is the most prevalent method of emulating the effects of these transient faults. Data mining techniques have been adopted not only to identify key infrastructural factors in determining the behaviour of systems in the presence of faults but also to develop an approach for design of highly efficient error detection predicates for error detection mechanisms for real-time software systems. In addition to high computational cost associated with exhaustive bit flipping, most of the bit-flip stress the system is similar ways. Furthermore, during system execution multiple bits are corrupted as an upshot of high power dissipation. Similarly, executing all possible bit-flip combinations will be computationally infeasible. Existing bit-flip fault injection analysis are limited to performing single bit flips exhaustively.

We propose a notion of using active learning in systematically selecting bits to flip and injecting multiple bits flips into a target system for each test case. Using the dataset generated from the fault-injection data, for each test case, it intuitively selects the most informative instance(s), i. e the best bit or combination of bits to flip, that will stress the system in varied ways. For example, for a 32-bit test case, the method may discern the best option is to flip the 8th and 16th bit to simultaneously by selecting the two instances that summarises the behaviour of the target system when these two bits were flipped; alternatively it can present the best bit to flip as the 20th bit singly by selecting the instance captures the system state when that single-bit was flipped.

Optimisation of Patch Distribution Strategies for AMR Applications

David Beckingsale

Abstract

As core counts increase in the world's most powerful supercomputers, applications are becoming limited not only by computational power, but also by data availability. In the race to exascale, efficient and effective communication policies are key to achieving optimal application performance. Applications using adaptive mesh refinement (AMR) trade off communication for computational load balancing, to enable the focused computation of specific areas of interest. This class of application is particularly susceptible to the communication performance of the underlying architectures, and are inherently difficult to scale efficiently. In this paper we present a study of the effect of patch distribution strategies on the scalability of an AMR code. We demonstrate the significance of patch placement on communication overheads, and by balancing the computation and communication costs of patches, we develop a scheme to optimise performance of a specific, industry-strength, benchmark application.

Session 3 (14:10 – 14:30)

Using Empirical Modelling to Build Parsers

Hui Zhu

Abstract
Most parsing methods fall into one of two classes, called the top-down and bottom-up methods. I will introduce shift-reduce parsers, a general style of bottom-up syntax analysis, and how shift-reduce parsers are modelled in Empirical Modelling.

There are many reasons for making such a model. A finite-state-automaton generated by L-R parser are often quite complex. A finite-state automaton is an abstract machine that can be in one of a finite number of states. The machine is in only one state at a time, called the current state. It can change from one state to another when initiated by an event or condition, this is called a transition. A particular finite-state automaton is defined by a list of the possible transition states from each state, and the condition for each transition. In LR-parsing: first, the grammar is complied into a finite-state machine; second a push-down automata is used to parse strings according to the grammar. Hence if the language is complex, there will be many states in the finite-state machine. Empirical Modelling is powerful of modelling parsers. This is due to the fact that in Empirical Modelling, modeller can link different states in finite-state-automaton,current content on the stack, next symbol to be read by the parser and other features of the parser, using network of dependencies. Once any of the above changes, the others' value will be updated immediately. Hence, the model can switch between different modes(shift or reduce) smoothly if the modeller defines correctly the variable which represents the current mode of the model.

There are three main motivations to use Empirical Modelling to model parsers. The first purpose is educational. We could understand how a parser works and understand how a parser is designed. Conventional programming aims to automate everything, while Empirical Modeling makes the state-changing activity explicit, intelligible and controllable. Unlike traditional programming, in Empirical Modelling, a human agent can be involved in the process of state-changing.

The second is practical. Empirical Modelling could help us to achieve building parsers on-the-fly and with better error recovery.

The third is philosophical. Empirical Modelling is a different approach to computing. In Empirical Modelling, people build models based on network of dependencies and they focus on representing states of models using network of definitions not procedures or actions. Hence we have more flexibility to change parsers.

The process of building parsers in Empirical Modelling includes the graph building, the data storing and designation of network of observables and dependencies. Every observable(variable) has a correct definition and then it will update automatically, no extra procedures are needed.

My talk will include a demonstration of parsing models and the tool. One example will be the eddi parser(eddi is a relational database notation in Empirical Modelling).

LDPLFS: Improving I/O Performance Without Application Modification

Steven Wright

Abstract
Input/Output (I/O) operations can represent a significant proportion of run-time when large scientific applications are run in parallel and at scale. In order to address the growing divergence between processing speeds and I/O performance, the Parallel Log-structured File System (PLFS) has been developed by EMC Corporation and the Los Alamos National Laboratory (LANL) to improve the performance of parallel file activities. Currently, PLFS requires the use of either (i) the FUSE Linux Kernel module; (ii) a modified MPI library with a customised ROMIO MPI-IO library; or (iii) an application rewrite to utilise the PLFS API directly.

In this paper we present an alternative method of utilising PLFS in applications. This method employs a dynamic library to intercept the low-level POSIX operations and retarget them to use the equivalents offered by PLFS. We demonstrate our implementation of this approach, named LDPLFS, on a set of standard UNIX tools, as well on as a set of standard parallel I/O intensive mini-applications. The results demonstrate almost equivalent performance to a modified build of ROMIO and improvements over the FUSE-based approach. Furthermore, through our experiments we demonstrate decreased performance in PLFS when ran at scale on the Lustre file system.

Session 3 (14:30 – 14:50)

Checking Equivalence of Quantum Programs and Circuits

Ebrahim Ardeshir Larijani

Abstract

Quantum Information Processing is a rapidly growing area in Computer Science and Physics which aims to harness quantum world to enhance computation and security of communications. Since its appearance nearly three decades ago, many advances in theoretical and practical aspects of QIP has been made. On the theoretical side, Shor's polynomial time factoring algorithm has shaken the so called "extended Turing-Church" thesis which asserts any physically realizable computation can be efficiently simulated with probabilistic Turing machine. Many unconditional cryptographic protocols like Quantum Key Distribution, Quantum Secret Sharing etc. have been introduced and their security, rigorously examined. On the practical side physicists have been able to cope with de coherence in noisy environments and build apparatus to carry out simple quantum computations and communications. Even more, recently European scientists have been able to realise quantum Teleportation (an important primitive for quantum cryptography) over 143 km in the free space which promises satellite quantum communication in the near future.

There are two important issues regarding the above progresses: First, How information flow occurs in quantum systems? For instance when we measure a quantum system during the execution of a protocol (like Teleportation) we end up with classical information, now, how we combine this and quantum information? To understand the quantum information flow, Abramsky and Coecke pioneered Categorical Quantum Mechanics. This area is inspired by using Category Theory for the analysing the semantic of programming languages.

Secondly, there is a lack of high level understanding in QIP. The current algorithms and protocols are usually described in the level of quantum circuit formalism. This led computer scientists to design quantum programming languages.

In this talk I will focus on the second issue and present our technique for verification of quantum programs by equivalence checking. Then I will give the details of implementation of this technique into our tool "Quantum Equivalence Checker". I will also discuss some ideas for verification of quantum circuits.

Performance Modelling of Magnetohydrodynamics Codes

Robert Bird

Abstract

Performance modelling is an important tool utilised by the High Performance Computing industry to accurately predict the run-time of science applications on a variety of different architectures. Performance models aid in procurement decisions, and help to highlight areas for possible code optimisations. This presentation covers a performance model for a magnetohydrodynamics physics application, Lare.

We demonstrate that this model is capable of accurately predicting the run-time of Lare across multiple platforms with an accuracy of 90%. We then utilise this model to evaluate the performance of future optimisations.

The model is generated using SST/macro, the machine level component of Sandia's simulation toolkit, and is validated on both a commodity cluster located at the University of Warwick and a large scale capability resource located at Lawrence Livermore National Laboratory.

Session 4 (15:10 – 15:30)

Using Data Mining Algorithms to Gain Insights into Digital Crime

Nentawe Gurumdimma

Abstract

Consider an investigation into a criminal act that took place, and the investigators, probably the police, decided to gather several eyewitness reports on the incidence from different people. These multiple documents containing the eyewitness reports will be too large for the police to go through each, analyse and concisely build a structure and gain an understanding of what exactly happened.

We are looking at how we can analyse these documents individually and integrate them (multiple data sources) so that a concise picture of events that took place before, during the crime can be built. We want to determine the order in which these events occur by inferring their temporal relationships and the actors involved and associate them to their corresponding events. Also, we want to build up a structural description of events and actors from each document for identifying conflicting reports among the eyewitness documents and the gap between them.

Using Natural Language Processing (NLP) and text mining techniques, the event and their associated actors across the document corpus can be identified. Information Extraction is a technique used to identify useful information in a document or document corpus, both structured and unstructured. It can be used to analyse textual data and identify relevant pattern or information contained in the text. Recently, most NLP techniques use the Machine Learning approach, as opposed to earlier Rule-based approach. These NLP techniques are used to perform Named Entity Recognition, which identifies entity mentions in a text such as persons, locations, organisations. This can be used to identify our actors involved in crime. Co-referencing, to determine which entity mentions refer to same object or entity. This is used to identify similar actors (same name entity actors) of an event in a document.

Others include: Part of Speech Tagging to determine the part of speech for each word in a sentence, Parsing to obtain parse tree and / or dependency parse tree for each of the sentences.

A dependency parse tree is an ordered, rooted tree that shows the relationships between the words in a sentence. We will briefly look at some of the existing tools that are used to handle some of these techniques.

The POS tags and Dependency Parse trees obtained from parsing the sentences is utilised for event and actor identification to obtain a more structured representation and accurate identification.

Accelerating Molecular Dynamics with Intel Xeon Processors and the Intel MIC Architecture

Simon John Pennycook

Abstract

In recent years, the use of co-processors (e.g. Cell, FPGAs, GPUs) alongside traditional CPU architectures has gained significant traction within the high-performance computing (HPC) industry. Such hybrid system designs are employed by some of the world's fastest supercomputers and have been suggested as being key to reaching exascale performance within an acceptable power budget; these co-processors' simpler control logic and smaller on-die storage enables higher compute density, leading to a larger peak flops-per-Watt. One area of HPC that has seen significant application speedups through the use of co-processors is molecular dynamics, due to the large amount of exploitable parallelism found in many simulations.

We present a new implementation of Sandia's miniMD benchmark (a simplified version of the popular LAMMPS package, intended for use in optimisation studies) that has been tuned for Intel Xeon CPUs and Intel Many-Integrated Core (MIC) co-processors. We detail a number of code-level and algorithmic improvements targeted at multi-threading and SIMD issues, and demonstrate how to achieve high levels of performance on server node configurations using a combination of both Intel Xeon processors and Intel MIC co-processors. In single precision, our implementation is up to $5\times$ faster than the original code running on Intel Xeon processors with AVX support, and adding a single MIC co-processor provides up to an additional $2\times$ increase in performance.

Session 4 (15:30 – 15:50)

The Longest Queue Drop Online Policy

Nicolaos Matsakis

Abstract
Memory management policies constitute a large area of research for online algorithms. In the case of shared memory switches, memory may completely fill up, therefore an online policy should serve packets trying to maximize the total number of served packets towards the total number of served packets of an optimal offline policy, for any possible sequence of packet arrivals σ. Two cases of online policies have been analyzed extensively: the preemptive policies where the rejection of packets already accepted is possible and the nonpreemptive policies where accepted packets cannot be rejected later. We consider packets of unit values and costs of transmission, organized in FIFO queues in case they are admitted.

The idea of preempting packets from the longest queue present in the buffer when the shared memory becomes full, was first proposed by Wei et al. in 1991. According to this policy, every packet is accepted as long as the buffer is not full. If it is full and the packet waiting for admission belongs to the longest queue, this packet is rejected, otherwise a packet from the longest queue is preempted and the packet waiting for admission is accepted. This policy is called Longest Queue Drop (LQD).

Time is assumed to be discrete. For every timestep, the packet positioned in the head of every active queue is sent to an output link and afterwards, but before the next timestep begins, the policy decides which packets to accept between those that have been placed in its input ports. An incoming port may accept many packets at the same timestep and many packets belonging to the same queue may arrive to one or more input ports simultaneously.

It has been shown that LQD is $(2 - o(1))$-competitive. A lower bound of $\sqrt{2}$ has, also, been shown for it. There exists a significant gap between the two bounds, however no $(2-\epsilon)$ upper bound (for some positive constant) has been shown, except for the special case of memory switches with two output ports, where an upper bound of $(4M-4)/(3M-2) < 4/3$ has been proved (M is the buffer size). Our research target, currently underway, is to show that LQD is at most 1.5-competitive, closing significantly the gap between the $\sqrt{2}$ lower bound.

Context-Aware Light-Weight Benchmarking for P2P Computing

Shahzad Malik Awan

Abstract
The peer-to-peer (P2P) computing paradigm is a practicable solution for providing cheap yet significant computational power by involving volunteer participation of the general public, which was otherwise infeasible under the traditional supercomputing paradigm due to the required computational power and its associated operational cost. As the shared resources contributed by the general public have no legal obligation with respect to resource provisioning, project administrators face two challenging tasks of: i) efficiently and effectively utilizing the best available resources from the available resource pool; and ii) motivating and encouraging participants to continue with the project as well as attracting new volunteers by using a reward system. The Berkeley Open Infrastructure for Network Computing (BOINC) is one of the widely used middleware platforms in P2P systems for creating, executing and managing loosely-coupled complex scientific applications in a dynamic and uncontrolled environment. As the volunteered computing resources are not under the direct control of the project administrators, running large benchmarks with long execution time on the participating nodes is not feasible.

A new light-weight synthetic benchmark – MalikStone has been developed for performance characterization of computing systems particularly those participating in a P2P project. MalikStone focuses on characterizing the performance of individual operations, which are widely used in real-world P2P programs. P2P applications commonly use primitive operations belonging to the categories of floating-point, integer, logic, programming constructs and string manipulation operations for performing lengthy complex scientific simulations. MalikStone is composed of five main function categories: floating-point, integer, logic, programming constructs and string manipulation; pertaining to operations which are commonly used in complex scientific P2P applications e.g., SETI, Einstein and Milkyway. Each of the operations in these function categories are mainly evaluated by the benchmark in a similar manner as they have been used in SETI, Einstein and/or Milkyway and provide a clearer insight into these individual function categories. Based on MalikStone, a new credit unit – MalikCredit – for P2P computing has been developed. MalikCredit addresses some of the limitations of the existing credit unit Cobblestone, which is used by BOINC as part of its rewards system.

Session 4 (15:50 − 16:10)

Allocating Resource for Workflows Running Under Authorisation Control

Nadeem Chaudhary

Abstract
Automating the execution of workflows (or business process) on computer resources has been the subject of much research. However, many workflow scenarios still require human involvement, which introduces additional authorization concerns. Role-Based Authorization Control (RBAC), under which the users are assigned to certain roles while the roles are associated with prescribed permissions, is a popular authorization control scheme. When we allocate resources for workloads and plan system capacities, it is often assumed that when a task is allocated to a resource, the resource will accept the task and start the execution once the processor becomes available. However, the authorization policies impose further constraints on task executions, and therefore may incur performance penalty and affect both application- and system-oriented performance. This paper investigates the issue of allocating resources for running workflows under the role-based authorization control. The resource allocation strategies are developed in this paper for both human resources and computing resources. The allocation strategy for human resources takes into account the authorization constraints and establishes the optimization equation subject to the constraint of the budget available to hire human resources. Then the optimization equation is solved to obtain the number of human resources allocated to each authorization role. The allocation strategy for computing resources also takes into account authorization constraints, calculating not only the number of computing resources, but also the proportion of processing capacity in each resource allocated to serve the tasks assuming each role. The simulation experiments have been conducted to verify the effectiveness of the developed allocation strategies. The experimental results show that the allocation strategy developed in this paper outperforms the traditional allocation strategies, which do not consider authorization constraints, in terms of both average response time and resource utilization.

Determining Influence in Complex Network Structures

Henry Franks

Abstract
In many open Multi-Agent Systems (MAS) there is no centralised control and individuals have equal authority, which can mean that ensuring cooperative and coordinated behaviour is a challenge. Social conventions and norms are particularly useful in such settings, since they offer a means of supporting cooperation, and can be established in an emergent decentralised manner. Identifying influential individuals can enable us to target mechanisms to encourage convention or norm emergence. Existing research has been limited with respect to (i) the extent to which metrics of location indicate influence in a network, (ii) how influence, in general, can be determined, and (iii) the extent to which differences in network structure distort the predictive power of particular metrics. In this talk, we propose a general methodology for learning the network value of a node in terms of influence, and demonstrate our methodology by evaluating it on a range of real-world and synthetic networks, using a model of convention emergence with realistic assumptions. We show that (i) eigenvector centrality, HITS (a measure of hubs and authorities), highest edge embeddedness, and node degree all effectively and robustly correlate with influence across heterogeneous network structures, (ii) using these metrics allows an agent to exert many times more influence over a population than a randomly placed agent, and (iii) common synthetic network generators give rise to structural and influence properties that differ significantly from those in real-world networks, limiting the extent to which results from synthetic networks can be generalised to real-world applications.

Session 4 (16:10 – 16:30)

Improved Bounds on Satisfaction Problems

Alexander Wilson

Abstract

Constraint-satisfaction problems are a class of problems that are of fundamental importance in Computer Science, with all NP-Complete problems being reducible to a 3-SAT formula, and a lot of effort has gone into understanding them. Of particular interest are the phase-transitions of a k-SAT problem: in terms of the number of variables and clauses, how "big" can a random instance become before it becomes unsatisfiable (with high probability)? We build on previous results by using subclasses of this problem (specifically NotAllEqual-SAT), and use this to derive improved lower bounds for $k = 4$ and 5.

Road Type Classification

Philip Taylor

Abstract

n this paper we investigate data mining approaches to road classification for roads based on controller area network (CAN) bus data collected from vehicles on UK roads. We consider three kinds of UK road classifications; the road type (A, B, C and Motorway), signage (None, White, Green and Blue) and carriageway type (Single or Double). Knowledge of these classifications enables the engine to be tuned for the environment to provide better efficiency or for the user interface to adapt to the situation. For example, some engine set-ups are more efficient on A roads than on motorways, especially in hybrid vehicles, and some user interface features such as a lane departure warning system (LDWS) improves safety on a multi-lane road, but are inconvenient in a residential setting. Further, the current road type and surrounding area gives an indication to the driver's workload. In a residential area the driver is likely to be overloaded, and on a highway they are likely to be under-stimulated.

Several data mining and temporal analysis techniques are investigated, along with initial attempts to deal with a class imbalance present in the data, aiming to stop classification algorithms over-fitting the majority, single carriageway, class. Such temporal analysis techniques include the representation of the data in frequency and frequency/time space through the discrete Fourier transform (DFT) and the discrete wavelet transform (DWT) along with aggregating as a Gaussian distribution. Data mining algorithms investigated include decision rules, decision trees, multilayer perception, random forest and hidden Markov models.

The area under the ROC curve (AUC) is used to evaluate classification algorithms, with a perfect classifier gaining an AUC of 1 and the uniformly random classifier an AUC of 0.5. It is found that the random forest ensemble algorithm has the best performance, with an AUC of 0.89 when used with the wavelet-Gaussian summary of the previous 2.5 seconds speed and steering wheel angle recordings. It is also shown that this technique is at least as good as a model based solution, which was manually created using domain expertise, demonstrating the efficacy of temporal data mining on CAN-bus data in the automotive domain.