

Projectvoorstellen

Scriptie BSc. Informatica 2019-2020

Computationele Wetenschap

A 3D Point selection tool for CloudCompare

Supervisor: Rob Belleman (r.g.belleman@uva.nl)

This project will be executed with the 4D Research Lab of the UvA Department of Archaeology.

CloudCompare (<https://www.danielgm.net/cc/>) is a popular open source 3D point cloud processing software package. It is commonly used for aligning, cleaning, georeferencing, and performing various statistical analyses on point clouds. In archaeology and related fields, 3D scan data becomes increasingly important and thus the demand to work with this data efficiently for purpose of visualization and analysis.

For the end-user the program has however limited selection capabilities. For instance, when editing a cloud, the selection tool draws 2D polygons that selects all points in the 3D space behind it. This offers problems when just small groups of points on the inside of a scan need to be selected, a very common occasion when working with scans of interiors. In such cases, the user has to cut up the point cloud in pieces to isolate the target points before he or she can select them. The solution is a 3D selection tool.

The student will explore the possibilities for such a selection tool in CloudCompare and evaluate alternative approaches. Following, a script is written for CloudCompare that enables 3D selection of points.

CloudCompare coding rules:

<https://github.com/CloudCompare/CloudCompare/blob/master/CONTRIBUTING.md>

CloudCompare forum topic with relevant basic information:

<http://www.cloudcompare.org/forum/viewtopic.php?t=2915>

Accelerating computational chemistry: a TURTLE case-study

Supervisor: Ana-Lucia Varbanescu, Remco Havenith (a.l.varbanescu@uva.nl)

GAMESS-UK is a large high-performance computing software package for computational chemistry ([https://en.wikipedia.org/wiki/GAMESS_\(UK\)](https://en.wikipedia.org/wiki/GAMESS_(UK))). Within GAMESS-UK, the calculation of valence bond wave functions is made possible by the TURTLE code. This code, much like the rest of the package, is computationally-intensive and requires large computing resources to reach results in a timely-manner.

In a collaboration with Prof. Remco Havenith, from University of Groningen, we aim to accelerate TURTLE by using GPUs. Specifically, in this project, we plan to use CUDA and/or OpenACC to accelerate a computationally intensive part of the code.

The challenges are: (1) devising a correct (set of) parallel algorithm(s) for the GPU, (2) empirically analysing and iteratively optimizing this algorithm (within the allocated time), and (3) embedding the accelerated kernels in the large GAMESS/TURTLE code, which is written in Fortran.

Requirements: the project requires basic knowledge in parallel processing in general and GPU processing in particular (e.g., at the level of the CPP course). Despite the code being very large, it is expected that the code to be written will not exceed a couple of hundreds of lines. No prior Fortran knowledge is required, but we expect the student working on this project not to be afraid of new programming languages.

Automatic generation of a volumetric mesh from multiple polygonal meshes

Supervisor: Rob Belleman (r.g.belleman@uva.nl)

This project will be executed with the 4D Research Lab of the UvA Department of Archaeology.

Archaeological excavations are complex structures consisting of irregular layers and patches of soil, structural elements and artefacts. Increasingly, archaeologists use 3D digital methods such as photogrammetry to document these. One issue is that the used scanning methods result in many separate polygonal meshes that have no formally defined relation to each other. They are isolated surfaces floating in 3D space. Next to this, no volumetric data is attached to them. This is not a realistic model of an archaeological excavation. A model of an excavation closer to reality would implement a voxel or tetrahedra approach. This allows both for preservation of volume and of the spatial relations between different archaeological features through their interfaces. In other words, the discrete definition of elements allows for no overlap and clear definition of boundaries.

This project focusses on an effective way to automate the generation of a volumetric mesh from multiple surface meshes. What is the most effective method to create a single continuous volumetric mesh from a series of surfaces, representing the boundaries of archaeological layers? Moreover, how to effectively integrate smaller features as part of the volumetric mesh? Ideally, from an end-user perspective, multiple surface meshes are selected and transformed into one volumetric mesh at once. The student is free to choose the package, or packages that is used as a platform for development, as long as it has an open source base. It is possible to approach the problem from either a GIS (e.g. GRASS GIS) angle, a 3D modelling angle (e.g. Blender), or a scientific simulation or visualization angle (paraview, VTK, openFOAM).

Lieberwirth 2008, 3D GIS Voxel-Based Model Building in Archaeology:

https://archiv.ub.uni-heidelberg.de/propylaeumdok/484/1/01_04_lieberwirth_voxel.pdf

Example of an unstructured grid approach implemented in Blender:

https://github.com/tkeskita/unstructured_grids

Resilience and repertoire: computational modeling for novel personalized markers

Supervisor: Rick Quax (r.quax@uva.nl)

Resilience, or the ability to resist and respond to perturbations, is expected to be greater for people with a broader 'repertoire' of available mechanisms to cope with the effects of internal and external disturbances. For example, the body maintains stable cerebral blood flow (CBF) despite highly fluctuating systemic blood flow demands. With age and disease this response repertoire narrows, and disturbances might push CBF out of the optimal range causing serious problems (e.g., syncope and falls). Assessing the breadth of this repertoire is therefore vital for screening for those at higher risk (prevention), tailoring support to restore balance after large disturbances (recovery) and broadening the repertoire to better buffer future disturbances (intervention). However, in geriatric medicine it is currently impossible to estimate this repertoire and quantify the corresponding level of resilience. We propose to develop a novel set of generic metrics that are indicative of the breadth of the repertoire by extracting key features from non-invasive, physiologic time-series data.

This project takes a first step towards extracting these metrics and validating them with empirical data. Compared to perturbations encountered within the repertoire, fluctuations that fall outside this range are expected to require wider, yet slower secondary and tertiary response mechanisms to buffer and result in larger deviations and longer recovery times of regulated variables. Thus, the skewness of the sampled response distribution is hypothesized to be inversely associated with the repertoire. The student will be tasked with developing a series of stylized computational models with increasing complexity. These models will generate simulated (within- and between-person level) data from which we can extract hallmark features from responses to simulated perturbations as candidate indicators of a narrow response repertoire. The main objective of the internship will be to test the hypothesis that the distributional properties (e.g., skewness, kurtosis) of time-series data in response to perturbations hold information about the breadth of a person's repertoire, and thus, their resilience. This work will be carried out under the joint supervision of Dr. Rick Quax (UvA), Dr. Jerrald Rector (daily supervisor) and Dr. René Melis (both Radboudumc) over the course of 3 months.

Spatial selection queries in Blender

Supervisor: Rob Belleman (r.g.belleman@uva.nl)

This project will be executed with the 4D Research Lab of the UvA Department of Archaeology.

Blender (<https://www.blender.org/>) is a popular open source 3D modelling program that is also increasingly being used by archaeologists working with 3D data. It is used for constructing and rendering 3D reconstructions, visualizing excavation data (trenches and objects), and processing and rendering of 3D captured data (scans).

Archaeologists often need to know the specific spatial relation between objects. A common question is: which objects are found inside a particular layer or within a particular feature? Although Blender has many selection options, there is no functionality to select objects based on their spatial relations. So such basic archaeological questions are not easily answered. In this project the student will work on extending the selection functionality in Blender so it can select objects based on their spatial relation. The primary goal is to add a selection function that allows the user to select mesh objects located on the inside or outside of another 3D mesh object. A special feature would be to allow the user to define the amount of overlap objects should have on the boundaries of the selection (eg. 80%, 50% or 10% overlap of the 'selected object' with the 'selection object').

Thermal image manipulation with R Thermal

Supervisor: Rob Belleman (r.g.belleman@uva.nl)

This project will be executed with the 4D Research Lab of the UvA Department of Archaeology.

Drone thermography is increasingly being experimented with as an archaeological remote sensing technique. The processing of thermal data from drones needs a lot of work; calibration, normalisation and optimization for photogrammetrical applications. Much existing software suites for thermal data is however optimized for thermal inspections and reporting, with few options for more advanced processing.

In this project, students will research the Thermal plugin for R to explore the possibilities of more advanced processing of thermal data, such as calibration and normalization. Additionally, a common photogrammetric suite such as AgiSoft Metashape can be explored as well for its potential to incorporate a python plugin for the same procedures. The research can result in a workflow for calibration and normalisation, and an evaluation which of these software environments are best suited for these functionalities.

Translating parts of ISR3D model from Java to C++

Supervisor: Pavel Zun (p.zun@uva.nl)

"Translating parts of ISR3D model from Java to C++

ISR3D is a multiscale model of tissue growth and remodelling in arteries after a surgical intervention. ISR3D couples a lattice-based model of flow to an agent-based model of arterial wall. The coupling and mapping values from agents to lattice cells and vice versa is done via special modules called mappers.

To enable easy uncertainty quantification (UQ) and sensitivity analysis (SA) of the model, ISR3D has to be ported to use MUSCLE 3 (Multiscale Coupling Library and Environment 3) for coupling.

However, the mappers are written in Java, and MUSCLE 3 does not have Java support. Hence, the goal of this project is to translate the existing mappers to C++, which is supported in MUSCLE 3. This would enable large-scale UQ and SA simulations.

Contact person: Pavel Zun, p.zun@uva.nl

Computerarchitectuur

Adaptable RTOS scheduling

Supervisor: Taco Walstra (t.r.walstra@uva.nl)

In realtime embedded systems it is necessary to react on events within a strict time frame where it is not possible that the timing requirement is allowed to fail. Often such systems are constructed with hardware and software components which fulfill such strict requirements using a RTOS (realtime operating system) for the application code. Such RTOS system normally use one of the commonly used scheduling algorithms like Earliest Deadline First. Although many realtime tasks have a periodic character, the system needs also to respond to sporadic events which can overload the system to handle events and normal periodic tasks. Ant colony optimization (ACO) can be used in such situations.

In this thesis you will implement such a system on a simple microprocessor board with a RTOS. You will need to adapt the OS to allow an alternative scheduler and test the performance.

literature pointer:

Kotecha (2008), Adaptive Scheduling Algorithm for Real-Time Operating System

Thomas S. (2018), An Intelligent Adaptive Scheduler for Operating Systems Experimented Using FreeRTOS

Extension of an RTOS with ant colony optimization

Supervisor: Taco Walstra (t.r.walstra@uva.nl)

In realtime embedded systems it is necessary to react on events within a strict time frame where it is not possible that the timing requirement is allowed to fail. Often such systems are constructed with hardware and software components which fulfill such strict requirements using a RTOS (realtime operating system) for the application code. Such RTOS system normally use one of the commonly used scheduling algorithms like Earliest Deadline First. Although many realtime tasks have a periodic character, the system needs also to respond to sporadic events which can overload the system to handle events and normal periodic tasks. Ant colony optimization (ACO) can be used in such situations.

In this thesis you will implement such a system on a simple microprocessor board with a RTOS. You will need to adapt the OS to allow an alternative scheduler and test the performance.

literature pointer:

Kotecha (2008), Adaptive Scheduling Algorithm for Real-Time Operating System

Harshl Garodi et al (2015), Performance Enhancing in Real Time Operating System by Using HYBRID Algorithm

Generic assembly code simulation

Supervisor: Taco Walstra, Toto van Inge, Edwin Steffens (t.r.walstra@uva.nl)

This project is related to the SystemC computer architecture simulator.

This project consists of implementation of a generic and user extensible ISA (instruction set architecture) for a processor design which allows the running of relative small assembly programs like we do in the computer architecture course with SIM-PL. In this case not only MIPS structures but also other processor systems should be able to be simulated depending on a generic ISA description. The main idea is to have a separation between graphical description of a design and the actual code simulation. The simulator should step through the code and report the number of used clock cycles and contents of memory and register structures. You should show that an assembly file containing a (limited) MIPS instruction set can be simulated, but also other different processor types like for example a VLIW or Intel processor.

Literature pointer:

Kohl (2016), A systemc based framework for cycle accurate processor simulation and parameter analysis

Node definition for distance measurement in MANET based swarming, from simulation to real-world implementation.

Supervisor: Toto (a.vaninge@uva.nl)

Distance estimations between nodes by the nodes themselves is essential for true swarming. Based on respeaker.io (<https://respeaker.io/>) and already implemented algorithms (in simulation and in hardware) a real-world implementation will be explored. The research can focus, among other topics, on: net-centric computing, load balancing, distributed image processing.

The nodes will be based on a FPGA/ARM Linux-based processing board. All processing, also for the drone its autopilot- and sensor-data, is carried out exclusively by the processing board forming a so-called MANET with the other boards.

Remark: The complexity of the platform is such that it is possible to aim for a group consisting of several graduates who work on different aspects of the project. The previous projects have also been successfully implemented in this composite form. Further work can be based on the work and graduation theses of several predecessors.

Performance analysis of a run time reconfigurable processor using partial configuration in a FPGA

Supervisor: Taco Walstra (t.r.walstra@uva.nl)

In embedded systems there is often a constraint in delivering of efficient performance and the energy consumption of the system. Depending on the application of the system the system can sometimes benefit from adapting the hardware of the system at runtime to accelerate the performance. FPGA's are very useful to provide such flexible behaviour using partial configuration. In this project you investigate how performance can benefit from such partial configuration. Possible routes are extension of the register file of a very basic processor (like the MIPS processor you have seen during the Computer Architecture course in the first year of your study) or Instruction Level Parallelism (ILP) changes at runtime. You will be using the Xilinx toolchain Vivado and simulator environment to analyse the performance and energy consumption for a certain FPGA chip.

Some literature pointers:

Souza J.D. (2018), Dim-Vex: Exploiting design time configurability and runtime reconfigurability, International symposium on applied reconfigurable computing

Wong S. Et al (2010), Dynamically reconfigurable register file for a softcore VLIW processor, 2010 Design, Automation and Test in Europe Conference

Albuquerque G.(2017), LP-P2IP A low power version of P2IP architecture using partial reconfiguration, 13th international symposium ARC2017 Delft Vivado user guide, partial reconfiguration, UG909 / UG947,

https://www.xilinx.com/support/documentation/sw_manuals/xilinx2018_1/ug909-vivado-partial-reconfiguration.pdf And :

https://www.xilinx.com/support/documentation/sw_manuals/xilinx2018_1/ug947-vivado-partial-reconfiguration-tutorial.pdf

SystemC computer architecture simulator

Supervisor: Taco Walstra, Toto van Inge, Edwin Steffens (t.r.walstra@uva.nl)

For Computer Architecture course the UvA have been using SIM-PL as platform for education of basic processor structures. Although SIM-PL fulfills many educational requirements, the platform is not easy extended with different features. One example is the export of a design into a real hardware system like a FPGA. A possible solution can be provided if the design is done using the descriptive language SystemC which allows translation into a low level RTL language like VHDL or Verilog.

In this project you will design a system which is able to describe describes basic processor and sequential or combinatoric logic in SystemC. We like to simulate such a design in a user friendly way showing specific logic signals or values of this design when providing one or more clock signals or - in case of non-synchronous systems - changes of (some) input signals provided. Testing of the final design can be done using off-the-shelf software. The system should be easily extendible, controllable by a script and the emphasis should be on the SystemC part. A strict project-based approach should be followed.

A second project for the simulation of assembly code is related this project

Literature pointer:

Kohl (2016), A systemc based framework for cycle accurate processor simulation and parameter analysis

Netwerken

An attack analysis on container networking architectures

Supervisor: Sara Shakeri, Paola Grosso (s.shakeri@uva.nl, p.grosso@uva.nl)

In recent years, containers are becoming a leading approach for setting up the applications quickly, easily, and reliably. In fact, they are introduced as an alternative to virtual machines. However, the method of bringing the connectivity among containers in a secure and efficient way is a big challenge that substantially depends on network setup architecture. Although there are multiple network architectures for bringing the connectivity among containers, every method is different in providing security and performance. In this project, you compare three possible network architectures of container networking from security perspective. You perform attack analysis on each method and see how each architecture is vulnerable to different possible attacks.

During this project, you will be familiarized with different concepts of containers and also start working with containerization tools like Docker, Swarm, and Kubernetes. Your results will be used in the SecConNet project.

The project link: <https://www.esciencecenter.nl/project/secconnet>

Detection and Correction of Incompatible Service Interfaces

Supervisor: Benny Akesson (k.b.akesson@uva.nl)

Context

Cyber-physical systems with long life times need to continuously evolve after deployment in response to changing technology and business needs. Lacking this ability not only prevents systems from quickly reacting to these changes, but also increases risk, as many small updates are collected into big infrequent upgrades. Service-oriented software architectures support continuous evolution by decoupling the application from a particular product, technology, and implementation using service interfaces. These interfaces hide the details of the component implementing the service, allowing them to be replaced with new implementations, possibly using other technology, when necessary. Although this arrangement has many benefits, there are challenges when the service interfaces themselves need to be updated. Changing the interface may be quick and easy, but it often means that all server and client components implementing them stop working. Manually addressing this problem by updating components is expensive and time-consuming. Automation is hence required to detect and correct incompatibilities, such as different message structures and protocol mismatches, when a service interface is updated.

*Assignment *

Initial research towards a methodology to address this problem have been taken by ESI (TNO) in collaboration with Thales. The methodology uses a domain-specific language to specify the structure and behavior of service interfaces. Formal models based on Petri Nets, representing the possible interactions with the service interface, are automatically generated from this specification and are used as a base for identifying incompatibilities and to synthesize adapters to correct them when needed.

Additional work is still needed to: 1) evaluate and improve the scalability of the methodology, and 2) deal with complex interface behaviors, such as data-dependencies. In the former case, software for generating synthetic interfaces with varying complexity w.r.t the number of incoming and outgoing signals and states in the protocol needs to be developed. In the latter case, we need to find and implement suitable transformations on a Petri Net to account for data dependencies. Each of these two parts is suitable work for one student.

The project may be organized as a paid internship at ESI (TNO), assuming the candidate is willing to travel to Eindhoven one or two days per week, and is an opportunity to learn about conducting applied research in an industrial context.

Your Profile

You are a Bachelor student in Computer Science or a similar program, looking for a graduation

assignment of 3 months. You have some experience with software development in (e.g. Java, C++, or Python) and have some familiarity with, or at least interest in the use of formal methods for development of software-intensive cyber-physical systems. Lastly, you would like your graduation assignment to show the benefits of scientific system development methods in a high-tech organization.

EPI infrastructure for secure data sharing in healthcare applications

Supervisor: Jamila Alsayed Kassem, Paola Grosso (j.alsayedkassem@uva.nl)

Data interoperability across different healthcare providers is getting increasingly vital. The EPI* project extends on that to support the advancement of healthcare services and applications such as real-time monitoring, personalised diagnosis, and digital health twin. One of the problems that arises when discussing health data sharing is the stringent data protection and legalisation. Different requirements, governance, and resources are asked to be considered per sharing request. EPI works on providing an adaptive infrastructure that considers the mentioned variables, and avoids the “one-fits-all” approach. The question to be answered is, When can health data move between different providers?

The idea here is to assign a number of nodes per health domain/provider, each one is characterized with a set of infrastructural attributes. The mentioned attributes describe what a node can provide and support in terms of computing/networking. In the EPI project, we want to find ways to group nodes with similar characteristics in 'EPI areas'. You will evaluate and improve the current algorithm for area creation. We aim to evaluate the algorithm not just by simulation, but also with real life testbed with the collaboration of SURFsara. If your results were reliable and good, it will be used in the EPI project.

*EPI - Enabling Personalized Interventions

Green traceroute

Supervisor: Paola Grosso, Emile Aben (RIPE NCC) and Marijke Kaat (SURFnet)
(p.grosso@uva.nl)

This is a project in joint collaboration with RIPE NCC (www.ripe.net) and SURFnet (www.surf.nl). We will schedule an intake to assess suitability of the students interested.

The "traceroute" tool is the bread and butter for network operators to figure out how packets traverse the Internet. A traceroute exposes the IP addresses of intermediate hops in a path between a source and a destination. We want to investigate if and how we can enrich traceroute data with energy data for paths that are visible with traceroute. Energy data would be both the type of energy (eg. green, grey) and amount of energy used. We will research how IP addresses map to energy data in the physical locations of that IP address, and if the information we can get on this is useful or not for understanding sustainability characteristics of Internet paths that are exposed in traceroutes.

Challenges here are understanding to what type of unit we need to map IP addresses (ie. individual router, data centre, city, country, network operator, network operator+data centre), and how to get meaningful energy usage data for these particular units. A stretch goal is to see if we can develop this to a point where we can order paths by their sustainability characteristics (ie. qualitative comparison with regards to minimising energy cost, and type of energy that was used)

Protocol containers for data transfers between DTNs

Supervisor: Reginald Cushing (r.s.cushing@uva.nl)

Data Transfer Nodes (DTNs) are dedicated servers with high bandwidth and storage meant to facilitate large data transfers on the Internet. How such distributed infrastructure is setup and programmed is still an area of research. In this project the student will look into the programmability of such an infrastructure using containers to deploy custom application protocol stacks (e.g. gridFTP, FDT, etc) on these nodes for the duration of the the transfer. This would, potentially, allow the application to customize the infrastructure to best suit its needs. That said, the problem becomes more complex when considering that such an infrastructure is a shared one and scheduling needs to be in place to maintain fairness.

Time synchronization simulation for wireless sensor networks

Supervisor: Taco Walstra (t.r.walstra@uva.nl)

In wireless sensor networks accurate timestamping of acquired data is often necessary. For simple internet connected systems the Network Time Protocol is sometimes sufficient, but for more accurate timing we need something else. Especially in wireless systems this can be difficult to achieve. The Precision Time Protocol is a standard which achieves more accurate timing and extensions of this protocol with other methods can even reach nanosecond accuracy between nodes on a wired version (White Rabbit protocol). Large wireless sensor networks are often investigated with simulators and in this project you will use/extend a simulator for time synchronisation in a WSN. Part of the project to give an overview of the several simulation options which have been recently researched for accurate synchronisation of wireless sensor networks. You will implement a simulation of PTP for several scales of such networks and investigate the accuracy for different configurations. Possible extensions of the simulation are ofcourse also possible.

Literature pointer:

Huang Y. Et al (2015) TS2: a realistic IEEE1588 time-synchronization simulator for mobile wireless sensor networks

Utilizing p4 switches in container networking

Supervisor: Sara Shakeri, Paola Grosso (s.shakeri@uva.nl, p.grosso@uva.nl)

Container technology is a virtualization technology at the operating system level. In recent years, they are becoming a popular way of deploying applications quickly, easily and reliably. There are still a lot of challenges about the method of connecting multiple containers together in a secure and efficient way. On the other hand, p4 is a well-known language for defining how packets are processed by the data plane of a hardware or software switch. In this project, you investigate how a p4 switch can be deployed in a network to connect multiple containers together.

During this project, you will be familiarized with different concepts of containers and also start working with containerization tools like Docker and Swarm. You will also learn to work with p4 switches. Your results will be used in the SecConNet project.

The project link: <https://www.esciencecenter.nl/project/secconnet>

P4 Language link: <https://p4.org/p4-spec/p4-14/v1.0.4/tex/p4.pdf>

Software

(Micro)benchmarking AKKA

Supervisor: Ana-Lucia Varbanescu (A.L.Varbanescu@uva.nl)

Together with Thales, we are using Akka to develop a framework for implementing decision making applications - like, for example, intelligent surveillance in national parks (aiming to catch poachers).

Our decision for using Akka was based on the features of the language. However, no in-depth analysis of the actual performance of the language has been performed. In this project, we aim to setup a microbenchmarking for the analysis of the framework.

The project will allow the student to get familiar with the framework, design a microbenchmark, and test it on a real system. The project will involve collaboration with Thales (one of their employees will co-supervise the work).

Automated containerization of legacy scientific code

Supervisor: Spiros Kolouzis, Zhiming Zhao (z.zhao@uva.nl)

In research communities, scientific programs are often developed by domain experts, using the script languages like python, and R. Those scientific programs are often results of many iterations of algorithm improvement, and only service specific experimental purposes. Those scripts often tightly couple the steps of experiments, and result in very limited reusability and portability to different platforms. Docker and singularity are OS-level virtualization technologies that allow for easy deployment, migration and scaling of applications.

In this project we aim to automate the containerization of legacy scientific code so we may transparently integrate, migrate and scale such code for the needs of large scale scientific workflows.

CoCoNut Code Generation Reloaded (Compiler Construction)

Supervisor: Clemens Grelck (c.grelck@uva.nl)

Background:

The Compiler Construction course makes use of compiler construction framework that has been developed over many years in the course of the SAC (Single Assignment C) research project. While it does accomplish its job in supporting the rapid construction of compilers in the programming language C, it is a hardly documented, fairly heterogeneous collection of tools developed in isolation over a long period of time.

In 2017 we started the development of the CoCoNut ("Compiler Construction in a Nutshell") framework that is supposed to replace the currently used framework. CoCoNut is based on a collection of DSLs (domain specific languages) and meta-compilation to generate C code that is then used for building the actual compiler.

Your Challenge:

As of now the CoCoNut meta-compiler generates C code for compiler construction that, very much unlike the old framework, is geared towards static type safety. For example, each node type leads to a dedicated C type. One downside of static type safety is that replacing a node by a node of different type requires awkward auxiliary functions. In this project we aim at exploring a more dynamic approach where the CoCoNut meta compiler would generate different C code, closer to what the old framework does, that would replace static type checking by comprehensive automatically generated dynamic consistency checks.

For evaluation we expect an experimental comparison of the two approaches (static vs dynamic) in terms of code elegance and programming productivity.

Prerequisites:

You must successfully have completed the Compiler Construction course.

Code smell analysis of concurrent Go programs

Supervisor: Damian Frolich, Ana-Lucia Varbanescu (a.l.varbanescu@uva.nl)

The Go programming language provides a lightweight alternative to threads in the form of goroutines. Besides being lightweight, they are also first-class citizens in the programming language. Therefore, it is rather easy to create concurrent programs in Go. Nonetheless, goroutines do not make writing concurrent programs less complicated.

In this project, you will be analysing Go programs and finding code smells in the context of goroutines. For example, Go already has the following slogan "Do not communicate by sharing memory; instead, share memory by communicating" - this guideline can form the basis of your thesis. Additionally, Go provides a document with common mistakes, which includes mistakes involving/using goroutines. The code smell analysis can be extended to include possible blocking and non-blocking bugs that are currently not reported by the two main Go tools [2, 1].

[1] Nicholas Ng and Nobuko Yoshida. "Static deadlock detection for concurrent go by global session graph synthesis"

[2] Tengfei Tu et al. "Understanding real-world concurrency bugs in go".

Determining incorrectness in concurrent Go programs

Supervisor: Damian Frolich, Ana Oprescu (a.m.oprescu@uva.nl)

The Go programming language provides a lightweight alternative to threads in the form of goroutines. Besides being lightweight, they are also first-class citizens in the programming language. Therefore, it is rather easy to create concurrent programs in Go. However, it is still not easy to reason about the order of execution in these programs.

In this project, you will be looking at determining incorrectness in concurrent Go programs by analysing the source code and analysing stack and traces generated by the Go runtime. This will be combined with a testing library [test-lib], giving the programmers options to define invariants and pre- and post-conditions about their concurrent Go programs.

The goal of this project is to create a tool that simplifies finding incorrectness in concurrent Go programs, which might result in simplifying programming concurrent Go programs.

Do programming languages have distinct energy footprints?

Supervisor: Ana Oprescu (ana.oprescu@gmail.com)

With EU Climate Law targeting climate neutrality by 2050, while ever growing the Knowledge Economy, the energy footprint of software services remains an important topic of research. In this project, we set out to understand whether some programming languages are inherently more “green” and whether established software smells (anti-patterns) are also energy-related smells. Possible directions include investigating the impact of the difference at code level on the energy consumption. The work would extend research conducted in two master projects[1, 2], and would use the DAS-based measurement framework developed in those projects.

References

[1] Lukas Koedijk. “Finding Significant Differences in the Energy Consumption when Comparing Programming Languages and Programs.”

[2] Stephan Kok. “The Impact of Refactoring Code Smells on the Energy Consumption of Java-based Open-source Software.”

Energy efficiency benchmarking

Supervisor: Ana-Lucia Varbanescu (A.L.Varbanescu@uva.nl)

Energy efficiency is becoming an important metric for comparing applications and their implementations. In this project, we aim to test different methodologies for energy measurement and analysis, and deploy them on different applications. The result of the project should be a list of guidelines and potential models to be used for energy efficiency estimation for a specific type of hardware.

The applications we have in mind are either from existing benchmarking suites like SPEC or from graph processing. The platforms we have in mind are CPUs, GPUs, or FPGAs.

Energy efficiency for heterogeneous computing

Supervisor: Ana Lucia Varbanescu (a.l.varbanescu@uva.nl)

Heterogeneous computing is a generic name for multiple types of processing units working cooperatively on the same workload. A good example of a heterogeneous platform is a CPU + GPU machine, where both units are used to execute the same application. While it is often clear what is the benefit of heterogeneous computing for *_performance_*, it is not entirely clear what the benefits are (if any) for energy efficiency.

Therefore, in this project, we aim to investigate and systematically improve the energy efficiency of heterogeneous CPU+GPU computing. To this end, we will focus on a set of applications and (1) build heterogeneous versions of these applications, (2) measure their energy efficiency, and (3) systematically determine what is the best workload mapping in terms of energy efficiency. Our ultimate goal is to determine whether there is tension between performance and energy efficiency, i.e., whether there is a loss in performance when energy efficiency comes first.

From concurrent to sequential Go programs

Supervisor: Damian Frolich, Ana-Lucia Varbanescu (a.l.varbanescu@uva.nl)

The Go programming language provides a lightweight alternative to threads in the form of goroutines. Besides being lightweight, they are also first-class citizens in the programming language. Therefore, it is rather easy to create concurrent programs in Go. However, concurrent programs are difficult to test due to the possible different orders of non-deterministic execution.

In this project, you will be transforming concurrent Go programs into sequential ones. This results in multiple configurations of a concurrent program, which then can be tested in a deterministic manner. You need to make sure that the program remains correct (compared to the concurrent version) and you will be providing a framework that can test the multiple configurations.

To achieve this "sequentialization", we will use the Go SSA package. The constructed framework will be evaluated on open-source Go programs, freely available on Github. Your framework will make it simpler for a programmer to test a concurrent Go program by writing tests for the several sequential configurations, instead of writing tests for the concurrent Go program.

Microbenchmarking for OpenACC

Supervisor: Ana-Lucia Varbanescu (a.l.varbanescu@uva.nl)

"OpenACC is the leading programming model in terms of performance portability. Endorsed by NVIDIA, it provides a pragma-based approach for GPU acceleration, where productivity and performance are kept reasonable. Moreover, OpenACC is supported by regular CPUs, where the same code can be executed.

However, preliminary results demonstrate that although the code is portable, its actual behavior is different on different types of hardware. As the compilers are closed source, the only way we can further investigate these anomalies are by microbenchmarking.

In the context of this project, microbenchmarking refers to the design and implementation of very small (as small as possible) test kernels for OpenACC primitives, which can then be further measured on different types of hardware (e.g., CPUs and GPUs), something like unit-testing for performance. Thus, specifically, the goal of this work is to generate a set of such benchmarks (3-5 of them), and analyze them empirically on 3 different platforms (CPUs and GPUs).

This project requires basic understanding of parallel processing, and basic experience with pragma-based languages (like OpenMP). The implementation effort is minimal. Most of the effort will be spent in empirical data collection and analysis. "

Performance patterns identification and analysis

Supervisor: Ana-Lucia Varbanescu (a.l.varbanescu@uva.nl)

For many applications, discovering the performance bottlenecks - i.e., the resource/code region that limits the overall performance is often a problem solved on an application-by-application basis. This approach is not scalable. Instead, we would like to provide a set of “performance patterns”, based on collected performance data, which are indicators of different performance problems. Preliminary work has demonstrated this approach is feasible, but the automation is very poor. In this project, we aim to investigate whether we can provide a set of performance diagnostics and curing tools, using performance patterns. To reach this goal, we must identify performance patterns from analyzing collected performance data.

Programming FPGAs accelerators in OpenCL

Supervisor: Ana-Lucia Varbanescu (a.l.varbanescu@uva.nl)

FPGAs (Field Programmable Gate Arrays) are circuits that can be reconfigured/programmed at runtime, thus enabling the execution of certain applications/kernels in hardware. FPGA accelerators connect FPGAs, via PCI/e, to a host processor. As such, the host processors manages the execution of the main application, and employs the FPGA to accelerate parts of the application.

In the past years, FPGA accelerators have seen an increase in popularity. This increase is due to the improvement of HLS (High-Level Synthesis), a process that enables FPGA programs to be written in high-level languages (instead of hardware-design languages). Example of such languages are C, Java, or OpenCL.

In this project, we aim to investigate the use of OpenCL for FPGA programming through the means of a case-study. Specifically, we aim to understand how well OpenCL performs in comparison with lower-level models like C or VHDL. In this process, we will (1) provide a first OpenCL implementation of the target application, (2) tune its performance, and (3) compare it against a native version (provided). All the empirical analysis will be performed on the FPGA itself.

Requirements: while OpenCL knowledge is not required, basic accelerator programming knowledge could be useful (GPU programming, in CUDA or similar models - at the level of CPP level).

Representing Access Control Policies using Semantic Web

Supervisor: Milen G. Kebede (m.g.kebede@uva.nl)

Several access control methods exist (role-based, attribute-based, capability-based, etc.) and are currently used in data-sharing infrastructures, but it is generally difficult to understand how they interrelate with each other. A first step to improve interoperability would be to attempt to describe access control policies in a neutral format, as for instance defining a policy ontology using semantic web technologies, as a basis of comparison. As a second step, one could imagine to increase the depth of representation/normative reasoning involved in the access-control management, and generate on-the-fly policies for the current access control infrastructure.

Following this idea, the student will:

- design an OWL ontology for access control policies (e.g. RBAC, ABAC, etc.), identifying potential limitations of the target representation,
- study potential mappings between the different types of policies, possibly identifying incompatibilities,
- develop a converter tool to the format of an actual system for access control, to be selected together with the supervisor.

Test-driving Intel oneAPI

Supervisor: Ana-Lucia Varbanescu (a.l.varbanescu@uva.nl)

In 2019, Intel has launched a new unified programming model, called oneAPI. To quote Intel, oneAPI is "a unified, standards-based programming model to simplify cross-architecture development and improve efficiency and innovation".

In this project, we aim to test OneAPI in terms of productivity and performance, using a set of benchmarking applications as case-studies. Our goal is to get familiar with this new technology, discover its limitations and flaws, and see how applicable it is in practice.

Onderwijs

Computer aided real-time feedback system for non-verbal personal presentation skills

Supervisor: Toto (a.vaninge@uva.nl)

This project will be based on former bachelor projects. The objective of the project is to research the possibility of real-time and offline feedback for performance improvement of novice speakers on their nonverbal skills without the need of audience feedback. Because there is no objective measure the tool compares speaker profiles. By calculating the profile discrepancy between 'professional' speakers and the profile of an arbitrary speaker it possible to make a quality statement.

The former tool was based on input from Kinect like devices. However, with the tools <http://densepose.org/> or <http://pose.mpi-inf.mpg.de/> (see pose estimation literature*) new and previous recorded presentations can be analysed and used as a source for generating representative profiles. In case online available presentations contain audience their scoring the generated profile can be linked to the scoring. A new profile can now get a scoring through the linked score.

Performance grades are given related to a predefined rubric and metrics (comparable with text editor generated readability statistics). Moreover, to have more robust reference profiles, as many as possible recordings will have to be analysed.

Follow link for more information (also for *):

<https://drive.google.com/drive/folders/1ZthrQmgqv3fd7faAwAHKh0VOTHrkU7Xn?usp=sharing>

Computer aided real-time feedback system for verbal personal presentation skills.

Supervisor: Toto (a.vaninge@uva.nl)

This project will be based on former bachelor projects. The objective of the project is to research the possibility of real-time and offline feedback for performance improvement of novice speakers on their verbal skills without the need of audience feedback. Because there is no objective measure the tool compares speaker profiles. By calculating the profile discrepancy between 'professional' speakers and the profile of an arbitrary speaker it possible to make a quality statement.

New and previous recorded presentations can be analysed and used as a source for generating representative profiles. In case online available presentations contain audience their scoring the generated profile can be linked to the scoring. A new profile can now get a scoring through the linked score.

Performance grades are given related to a predefined rubric and metrics (comparable with text editor generated readability statistics). Moreover, to have more robust reference profiles, as many as possible recordings will have to be analysed.

Improving collaborative writing through real-time feedback and monitoring of writing characteristics.

Supervisor: Toto (a.vaninge@uva.nl)

Researchers, but also students, often work together in a project setting that will result in, next to other deliverables, a paper through collaborative writing.

Text statistics could be as simple as word count produced over a day. However, this kind of quantitative data is no guaranty for the quality of the produced texts. Still, it can give for example insight in text consistency among the writers as shown in [J. Bottelier].

Because it is not possible to give an absolute or objective value for text quality, we propose to analyse existing texts of known writers and generate 'reference' profiles. The claim is that it is possible to generate a text discrepancy rating between reference profiles and the new text profile.

The research spin-off is a proof of concept tool for real-time feedback and monitoring. Essential here is that each of the members, who collectively produce texts, will do so in some prearranged digital environment. For instance Overleaf is a product to investigate as a possible writing tool candidate together with some repository environment. Free qualitative text analysis tools are available, but their quality and working is indeterminate.

Moreover, statistics across the groups can unlock information (monitoring dashboard) that is helpful not only for the group, but also for an individual. Additionally, user level permissions will prohibited specific data accesses. Furthermore, former bachelor thesis work (e.g. software) could be a starting point.

- Neomy Storch, Collaborative writing: Product, process, and students' reflections, *Journal of Second Language Writing* 14 (2005) 153–173
- Erik Duval, Attention Please! Learning Analytics for Visualization and Recommendation, LAK'11 February 27-March 1, 2011
- Dyckhoff, A. L., Zielke, D., Bültmann, M., Chatti, M. A., & Schroeder, U. (2012). Design and Implementation of a Learning, Analytics Toolkit for Teachers. *Educational Technology & Society*, 15 (3), 58–76.
- Derick Leony, Abelardo Pardo, Luis de la Fuente Valentín, David Sánchez de Castro, Carlos Delgado Kloos, GLASS: A Learning Analytics Visualization Tool, LAK '12, 29 April – 2 May 2012
- Brian J. McNely, Paul Gestwicki, J. Holden Hill, Philip Parli-Horne, Erika Johnson, Learning Analytics for Collaborative Writing: A Prototype and Case Study, LAK'12, 29 April – 2 May 2012
- M.S. Szyguła, Facility for collaborative consistent writing, bachelor thesis, 17 August 2016
- J. Bottelier, Comparing and Analyzing Individual Writing Statistics in a Collaboratively Written Document in Real-Time, bachelor thesis 2017.
- Automated Evaluation of Scientific Writing Shared Task 2016;
<http://textmining.lt/aesw/index.html>

Theorie

Developing a modular Python package for Judgment Aggregation

Supervisor: Ronald de Haan (me@ronalddehaan.eu)

The setting of Judgment Aggregation (JA), that involves voting on combinatorial domains, has been studied in the literature on computational social choice (see, e.g., http://www.cambridge.org/download_file/951600). In general, many computational problems related to JA are computationally intractable (NP-hard or worse), and consequently, there does not seem to be an optimal choice for algorithms to solve these problems---for different inputs, different heuristic algorithms work better. Various types of algorithmic approaches have been developed for JA, including approaches using Boolean circuits (<https://arxiv.org/pdf/1808.03043.pdf>; <https://github.com/blyxxyz/python-nnf>), and logic programming methods (<https://github.com/rdehaan/ja-asp>), for example. For a structured investigation of these different algorithmic approaches, and in which settings which of these approaches works better, it would be useful to have a single unified computation framework where (1) JA scenarios can be stored/loaded, (2) different algorithms for different problems in the setting of JA can be implemented/used, and (3) inputs and outputs of the problems can be presented in a user-friendly way (e.g., using visualization). As new algorithmic approaches for JA problems are likely to be developed in the future, a modular framework would be useful, where new algorithms/modules can easily be added. The aim of this project is to develop such a package in Python. This involves, among others, (i) designing the general set-up of the package so that new functionality can be added/used modularly, (ii) designing a storage format for JA scenarios that can handle different representation languages, that is as simple as possible (yet efficient) and that can also be interpreted by human readers, and (iii) implementing various proposed algorithms for JA problems within this modular package.

Investigating how people form beliefs about generics

Supervisor: Patricia Mirabile (p.l.mirabile@uva.nl)

Generics are statement of the form "Birds can fly" and "Lions have manes." We are studying the semantics of those statements (i.e. when those statements are believed to be true) using psychological experiments. One idea we would like to test is that learning processes play an important role in how we come to believe certain generics. In this project, you would help us develop an online experiment (sort of like a game) where participants will be able to interact with an environment, discover information about the objects inside it, and evaluate or verify generic statements about those objects. The experiment's UI should have a graphical component, which could be very simple (such as turning over cards with images) but could also be more creative. Finally, the experiment will need to be hosted online and be able to collect the data provided by participants.

Visualisatie en VR

A web-based framework for the composition of interactive 360 video experiences

Supervisor: Rob Belleman (r.g.belleman@uva.nl)

360 graden video is een populair mediaformaat om mensen situaties te laten ondergaan die als doel hebben een diepere indruk te maken dan met conventionele media het geval zou zijn. De toepassingsgebieden zijn talrijk en variëren van instructie en training tot zelfs patiënttherapie. Een recente ontwikkeling zijn 360-video ervaringen waarin interactie wordt geïntroduceerd zodat afgeweken kan worden van een lineaire verhaallijn. Een uitdaging is echter dat het creëren van dergelijke 360-video ervaringen niet voor iedereen is weggelegd.

Binnen dit project werk je aan een web-gebaseerde omgeving waarin niet-technische eindgebruikers op een eenvoudige manier interactieve 360-video ervaringen kunnen maken. Je gaat hier o.a. samenwerken met onderzoekers van De Bascule in Amsterdam bij een onderzoeksproject om een 360-video variant te maken voor Parent-Child Interaction Therapie (PCIT-VR).

Lab42 Digital Twin

Supervisor: Rob Belleman (r.g.belleman@uva.nl)

Lab42 is de naam van het nieuwe huis voor het Instituut voor Informatica (IVI) en Logica (ILLC) dat zal worden gebouwd naast Science Park 904 en in 2022 af moet zijn. We hebben van de architect een 3D model ontvangen dat interessante mogelijkheden biedt om simulaties uit te voeren over de toekomstige inrichting en gebruik van het gebouw, nog voordat het er staat. In dit project ga je werken aan een simulatiemodel dat moet helpen bij het beantwoorden van vragen die in de nabije toekomst gaan komen, zoals: wat is de meest efficiënte manier om de medewerkers in werkkamers te plaatsen (het zogenaamde "vlekkenplan")? Hoe kan de stroom van mensen binnen het gebouw onder verschillende omstandigheden (eind van colleges, noodsituaties, open dagen) geoptimaliseerd worden?

Learning analytics voor educatieve and training VR/AR applicaties

Supervisor: Rob Belleman (r.g.belleman@uva.nl)

In toenemende mate wordt VR/AR gebruikt om uitleg en/of training te geven aan studenten op onderwerpen die in de echte wereld te moeilijk, te duur, te gevaarlijk of domweg onmogelijk zijn.

Een onmisbaar onderdeel bij het gebruik van dergelijke applicaties is een mogelijkheid om de verrichtingen van een student bij te kunnen houden. Binnen dit project bouw je een “learning analytics” systeem om de verrichtingen van de gebruikers van VR/AR apps te monitoren, analyseren en presenteren. Dat systeem bestaat uit drie delen die zijn op te splitsen in twee afstudeerprojecten:

1. een "monitoring" deel dat relevante data over de prestaties van de student verzamelt terwijl VR/AR applicaties gebruikt worden (bv: aantal juiste/foute antwoorden, tijd die een eindgebruiker nodig heeft om een probleem op te lossen, bewegingen die de gebruiker maakt, etc.);
2. een “server” die deze data verzamelt in een database;
3. een "front-end" die de data in de database analyseert en presenteert aan student en docent om de voortgang van de leerprestaties te kunnen monitoren.

Onderdeel 1 en 3 zijn losse projecten. Op onderdeel 2 zal samengewerkt moeten worden.

Binnen dit project zal samengewerkt worden met Tata Steel die VR applicaties gebruikt om hun personeel op te leiden.

Scientific Visualization in Virtual Reality

Supervisor: Rob Belleman (r.g.belleman@uva.nl)

This project is for students with personal access to a high-end VR device (such as Oculus Rift, HTC Vive).

The purpose of visualization is to obtain insight in scientific data. Many scientific research projects generate large amounts of data that are difficult to analyze numerically. In those cases it can help to visualize the data and visually detect patterns that would have been difficult or even impossible to detect otherwise.

In previous years, students have worked on a framework for the Visualisation Toolkit (VTK) library that allows the output of visualization pipelines to be rendered in VR. However, the applications that can be built with this framework are largely non-interactive. We want to build upon this work and achieve the ambitious goal to create a fully immersive interactive VR application that allows a researcher to construct and control a visualization pipeline from within a virtual environment. There are several challenges in this project. First; the candidate needs to acquaint him/herself with the theory and practice of creating scientific visualizations as well as VR applications. Second; the candidate will design and implement a collection of interaction methods that enable the construction of a visualization pipeline from within a VR application. Last; the implementation will be tested on a real data set so that a critical assessment can be made on the effectiveness of the proposed solution.

Web-based visualisation of ultrasound images

Supervisor: Rob Belleman (r.g.belleman@uva.nl)

This project will be done in collaboration with the Embryology Department of UMC/AMC.

Congenital heart disease (CHD) is the leading cause of non-infectious neonatal mortality. Prenatal detection of CHD provides the opportunity to deliver potentially life-saving care immediately after birth, improving the prognosis of several types of severe CHD. The accuracy of ultrasound screening for CHD heavily depends on the training and experience level of the sonographer. In expert hands, nearly all severe CHD cases are detected, whereas the yield of population-based screening is about 50%. Given that only 10% of CHD occurs in pregnancies with known risk factors, improving the performance of population-based screening is the only way to increase prenatal detection of CHD, which requires investments in improving education and training. Through the proposed project, we aim to provide a clinical and educational resource that presents fetal echocardiography in a more approachable and intuitive way. By this we aim to support (trainee) sonographers in developing and maintaining the knowledge and skills required for fetal cardiac evaluation and to facilitate introduction of more advanced cardiac views in primary care settings. Three-dimensional (3D) modelling is an effective tool to provide a more comprehensive understanding of cardiac anatomy and can aid putting 2D ultrasound views of cardiac structures in context. Up until now, 3D modelling and reference atlases for fetal echocardiography have been separate entities. We propose to integrate 3D modelling techniques with ultrasound imaging by collecting ultrasound volumes of normal pregnancies and pregnancies complicated by CHD and reconstructing fetal cardiac structures and major vessels in these ultrasound volumes to create interactive 3D digital models of the heart, bundled in an open access online atlas of fetal echocardiography. This novel approach to presenting fetal echocardiography for educational purposes provides an exciting opportunity to increase and spread knowledge of fetal echocardiography, an important step towards improving population-based prenatal screening for CHD.

Our goal is to present the ultrasound data and 3D cardiac models in such a way that the user can manipulate the ultrasound volume to practice obtaining the imaging views used in clinical practice, as well the 3D model to review the cardiac anatomy with the benefits of depth perception and being able to appreciate the complex interrelationships of cardiac structures. In addition to the ultrasound volumes and 3D models, we will share ultrasound clips taken during the examination, still images, and (color flow) Doppler data, along with a case description and explanation of the basic and more advanced cardiac views used to evaluate the case.