SAT-based techniques for Approximate Circuit Design

Professor: Laura Pozzi

Abstract:

As energy efficiency becomes a crucial concern in every kind of digital application, a new design paradigm called Approximate Computing (AC) gains popularity as a potential answer to this ever-growing energy quest. AC provides a different view to the design of digital circuits, by adding *accuracy* to the set of design metrics.

So, while traditionally one could sacrifice area for delay, for example, or energy for area, etc, now the idea is to play with accuracy also, and pay a small loss in accuracy for a large improvement in energy consumption. This is particularly suited for error-resilient applications, where such small losses in accuracy do not represent a significant reduction in the quality of the result. While Approximate Computing can be applied at different levels – from software to hardware – in our group we are particularly interested in the design of approximate boolean circuits. In particular, we are research Approximate Logic Synthesis, which is the process of automatically generating, given an exact circuit and a tolerated error threshold, an approximate circuit counterpart where the error is guaranteed to be below the given threshold. The resulting circuit will be a functional modification of the original one, where parts will be substituted, or even completely removed.

While various algorithms have been proposed – in and out of our group – for the design of approximate circuits, we are currently exploring new SAT-based solutions. The SAT (or boolean satisfiablity) problem states the following: given a formula containing binary variables connected by logical relations, such as OR and AND, SAT aims to establish whether there is a way to set these variables so that the formula evaluates to true. If there is, the formula is SAT; if there isn't, the formula is UNSAT.

An astonishing number of problems in computer science can be reduced to the SAT problem – including our approximate circuit design question – and, in addition to this, astonishingly fast SAT solvers exist. Hence, in this project we aim at designing (and improving our existent) SAT-based formulations and algorithms for circuit design, in order to generate ever more efficient approximate circuits.

You will need a very basic knowledge of gate-level design, and your programming skills! Drop me an email if you are interested and you want to have more information.

Supervision by Laura Pozzi, Ilaria Scarabottolo, Morteza Rezaalipour

Useful links:
An introduction to SAT:
https://www.borealisai.com/en/blog/tutorial-9-sat-solvers-i-introduction-and-applications/
An example of approximate circuit design technique:
https://ieeexplore.ieee.org/document/8342067
A survey of approximate circuit design techniques:
https://www.inf.usi.ch/phd/scarabottolo/papers/ALS_survey.pdf

Analyzing Runtime Variations of JVM Workloads across Multiple Executions

Running the same workload multiple times in separate Java Virtual Machine (JVM) processes often results in different runtimes. This behavior can generally be explained by nondeterminism. Even if the workload itself is deterministic, nondeterminism can affect many components involved in the application execution. For example, in different runs of the same workload:

- the operating system (OS) may decide a different thread-scheduling;
- nondeterministic lock acquisitions may alter the order in which threads execute.
- nondeterministic executions of the garbage collector (GC) may affect object allocations, which in turn may cause different access patterns to the thread-local allocation buffer (TLAB).
- the ergonomic phase performed by the JVM during warm-up may tune heap- and GC-specific parameters in different ways;
- the JVM may trigger dynamic compilation of different loops or methods;
- the just-in-time (JIT) compiler may apply different optimization decisions;
- the processor may rearrange instructions in a different way;
- cache-access patterns may significantly differ, leading to different frequencies of cache misses.

Such nondeterminism may significantly impact application performance, resulting in different runtimes across multiple executions of the same workload. In some cases, the runtime variability can be very high and complicate performance evaluation. For example, accurately assessing the execution-time overhead of a dynamic profiler can be challenging due to the high variability of the runtimes obtained in different executions, which may give unrealistic measurements (e.g., they may indicate a speedup instead of a slowdown). Therefore, an in-depth analysis on such performance phenomena is fundamental.

This project makes a first step in this direction, aiming at analyzing runtime variations of applications executed on the JVM. In particular, for a given workload (that exhibits significant runtime variations in different JVM processes), we aim at identifying the root causes of the observed runtime variability. To this end, we aim at correlating the observed runtimes with different dynamic metrics. Such metrics span the whole system stack, describing attributes of the application, the JVM, the JIT compiler, the OS, and the hardware. We will design and develop profilers to collect metrics of interest with high accuracy and low overhead.

To achieve the goals of the project, the student will be involved in several activities:

1. Identification of workloads where the runtime varies significantly in different runs.
2. Identification of metrics from the whole system stack that can be correlated with the runtime of applications running on the JVM.
3. Development of efficient and accurate profilers to collect metrics of interest.
4. Analysis of the correlation between runtimes and metrics of interests in the considered applications.
5. Detection of the root causes of significantly different runtimes and identification of possible ways to reduce runtime variability.

The project is a unique opportunity for the student to deepen her/his knowledge in the domains of dynamic program analysis, the JVM architecture, the system stack, and empirical evaluation, all being important skills for a software engineer. The student will work side-by-side with the members of the Dynamic Analysis Group at USI, and will receive support in learning advanced topics that will strengthen her/his abilities as a computer scientist. Applicants interested in this project should be enrolled in the BSc program, have a good knowledge of the JVM architecture, Java, C/C++, and UNIX-based operating systems, excellent programming skills, and deep interest in the field of dynamic analysis.

Advisors: Prof. Walter Binder and Dr. Andrea Rosà
Assistant: Matteo Basso

EEG-RECALL: Analysis of the influence of distractions on human memory
EEG Signals and distraction

Memory Augmentation Systems could be capable of selecting specific to-be-remembered events during an experience, e.g., by detecting the individual as distracted or unengaged. Such information then could be used to generate memory cues for the specific periods during which the user was distracted. One way used to detect the cognitive state of users, e.g. if the user is being interrupted, is through physiological signals, like electroencephalography (EEG). The goal of this thesis is to evaluate the potential of EEG Signals captured by an OpenBCI device to detect distractions of individuals presented with a memory task. The final purpose is to use this work as an additional input in future Memory Augmentation Systems. The specific tasks of the project are: (i) Overview methods for EEG signal processing and review memory augmentation literature; (ii) Develop the necessary code for processing the EEG data; (iii) Design and run an experiment to test the setup and pipeline.

Required skills and knowledge

- Good knowledge of Python programming (i.e., using Jupyter notebook)
- Knowledge of data analysis and machine learning
- Ability to run a small user-study to collect data (5-10 participants)

Expected outcomes

- Related work summary
- Collected dataset
- Implemented a machine learning pipeline to automatically recognize distractions from EEG data.

Supervisors:

- Prof. Marc Langheinrich
- Matías Laporte
- Martin Gjoreski

More information about the research group available at: https://pc.inf.usi.ch/

Physio-RECALL: Analysis of human memory and physiological signals

Memory Augmentation Systems could be capable of selecting specific to-be-remembered events during an experience, e.g., by detecting the individual as distracted or unengaged. Such information then could be used to generate memory cues for the specific periods during which the user was distracted. One way used to detect the cognitive state of users, e.g. if the user is focused on a task, is through physiological signals, like electrodermal activity (EDA) and interbeat interval (IBI). The goal of this thesis is to evaluate the potential of physiological signals captured by a wrist-worn device to detect the cognitive load of individuals presented with a memory task. The final purpose is to use this work as an additional input in future Memory Augmentation Systems. The specific tasks of the project are: (i) Overview methods for physiological signal processing and review memory augmentation literature; (ii) Develop the necessary code for processing the EDA and IBI data; (iii) Design and run an experiment to test the setup and pipeline.

Required skills and knowledge

- Good knowledge of Python programming (i.e., using Jupyter notebook)
- Knowledge of data analysis and machine learning
- Ability to run a small user-study to collect data (5-10 participants)

Expected outcomes

- Related work summary
- Collected dataset
- Implemented a machine learning pipeline to automatically recognize distractions from physiological signals.

Supervisors:

- Prof. Marc Langheinrich
- Matías Laporte
- Martin Gjoreski

More information about the research group available at: https://pc.inf.usi.ch/

PrivAffect: Privacy-aware personal-video sensing for affect recognition

Affective computing is an interdisciplinary field that aims at the development of computer science techniques that enable machines to recognize, understand and simulate human affective states. Video-based sensing is one promising approach for affect recognition, however, it is also privacy intrusive. Thus, this project will develop a method for privacy-aware personal-video sensing for affect recognition. The method will utilize personal camera (e.g., smartphone or laptop camera) and will include privacy-aware features such as: to record only when the users grant permission, to record only when a specific user is in front of the camera (user identification). Once the video is collected in a privacy-aware manner, existing video-based affect recognition software will be sued to extract informative. The specific tasks of the project are: (i) check existing software (e.g., on GitHub) for user identification, software for counting faces in a vide, and software for Facial Action Coding System (FACS) [1]; (ii) implement privacy-aware user identification; (iii) implement privacy-aware method for extracting Facial Action Units (based on FACS); (iv) test the overall processing pipeline is a small user-study (e.g., 5 to 10 participants).

Required skills and knowledge

- Good knowledge of Python programming (i.e., using Jupyter notebook)
- Knowledge of data analysis and machine learning. (Knowledge in video analysis is a plus)
- Ability to run a small user-study to collect data (5-10 participants)

Expected outcomes

- Overview of existing software for the task
- Collected a small dataset
- Implemented a simple method for privacy-aware personal-video sensing for affect recognition

Supervisors:

- Prof. Marc Langheinrich
- Martin Gjoreski
- Matías Laporte

More information about the research group available at: https://pc.inf.usi.ch/

Fed-CogLoad: Federated Cognitive Load Estimation

Federated learning (FL) is a state-of-the-art machine-learning technique developed by Google, where the users' privacy is guaranteed by implementing one simple rule: "No personal data leaves the user-device". This project will investigate FL techniques for cognitive load estimate. Cognitive load can be estimated through the analysis of from pupillometry data, brain activation data (EEG), breathing rate, heart rate, heart rate variability and other related physiological responses. In the project you will: (i) overview existing datasets for cognitive load estimation; (ii) develop a centralized machine learning pipeline for cognitive load estimation; (iii) develop a FL pipeline for cognitive load estimation, (iv) compare the centralized and the FL pipeline and write a short report.

Required skills and knowledge

- Good knowledge of Python programming (i.e., using Jupyter notebook)
- Knowledge of data analysis and machine learning

Expected outcomes

- Related work summary
- Find a test dataset
- Implemented a simple method for federated cognitive load estimation

Supervisors:

- Prof. Marc Langheinrich
- Martin Gjoreski
- Matías Laporte

More information about the research group available at: https://pc.inf.usi.ch/

Python Expressions Analyzer

Motivation

In most programming languages, a sizable fraction of source code consists of expressions. Expressions are not limited to arithmetic and are of paramount importance when learning to program. While some expressions seem easy to understand, in our work we have identified a significant number of misconceptions students hold about expressions.

Goal

In this project you will thoroughly analyze the Python language to determine which AST (Abstract Syntax Tree) nodes represent parts of expressions. You will leverage the API of a compiler frontend to parse source code, programmatically detect expressions, and analyze their structure with the ultimate goal of producing an augmented tree-like graphical representation for each expression.

Prerequisites

Excellent analytical and algorithmic thinking skills and ability to quickly familiarize yourself with new formalisms (such as formal syntax and semantics of programming languages).

Platform / Languages / Frameworks

In this project you will complement prior work carried out focusing on Java. Prior experience with the following technologies is beneficial: Python, compilers.

More Information

If you are interested in this project, please contact luca.chiodini@usi.ch and matthias.hauswirth@usi.ch to discuss the details.

This is me: Enhancing self-representation with Multisensory Experiences

UROP project proposal at the Università della Svizzera italiana (USI), Lugano, Switzerland

Most of our everyday experiences are multisensory in nature. What we can see, hear, feel, taste, smell, and much more are part of our daily lives. Almost any experience you can imagine, such as walking at the park or watching content on your smartphone, involves a magnificent sensory world.

Multisensory experiences, where the senses meet technology, are available in multiple scenarios and can be used to enhance self-representation. Self-representation is an instance of representing one person and can be important for people with cognitive impairments as the materialized version of their feelings and preferences.

Using multisensory technology to enhance objects can add fun to the process while supporting creativity in freely expressing their identity and preferences. An example of self-representation objects can be seen in the image below.

In this project, you will help design and develop multisensory self-representation objects for people with intellectual disabilities from an association we have a long-standing collaboration.

You will be under the supervision of Dr. Monica Landoni and Leandro Guedes, a Doctoral assistant at USI. Collaborations can involve international researchers from different research areas.

Are you interested in this project? Contact us if you have any questions, we will be happy to answer them.

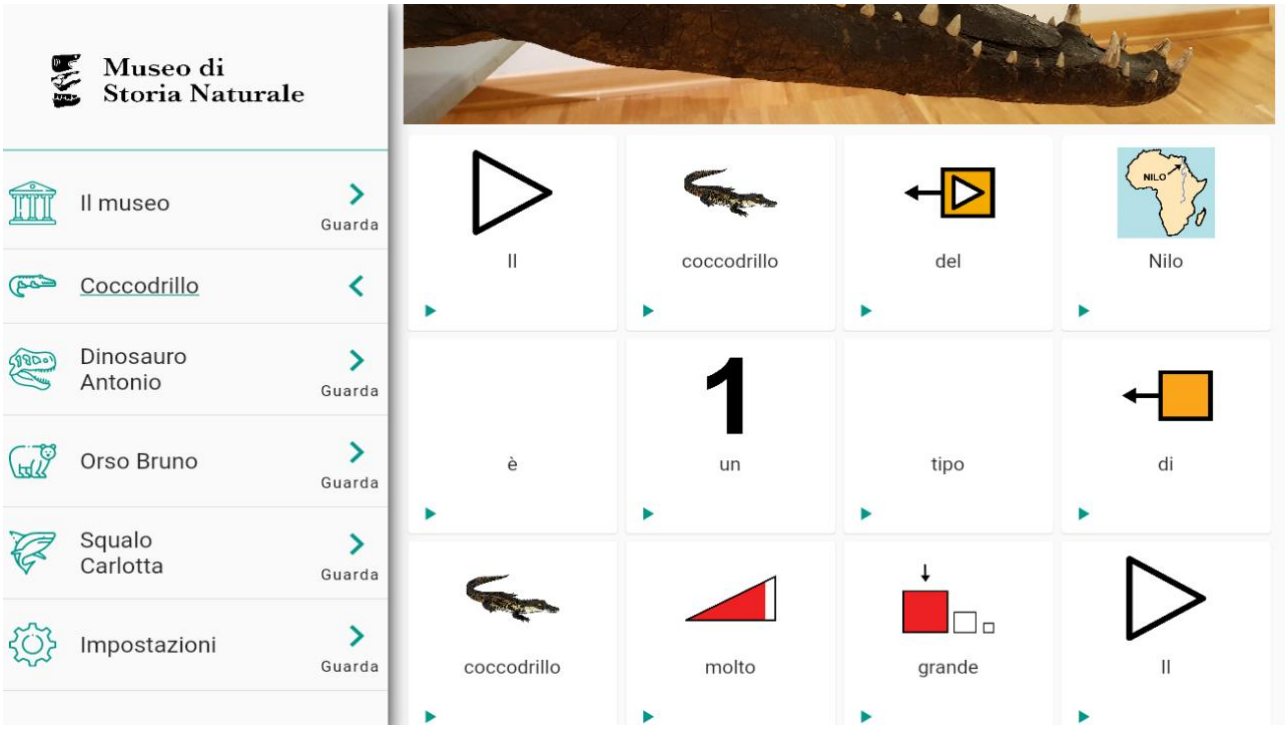
Redesigning and improving an accessible application for people with intellectual disabilities

UROP project proposal at the Università della Svizzera italiana (USI), Lugano, Switzerland

Digital accessibility and assistive tools have been used to enhance the experience of people with different abilities. Developing accessible solutions with and for people with disabilities makes a huge contribution to the quality of their life.

In this project, you will follow user experience design methods and fulfill essential accessibility guidelines. The work includes developing prototypes and a final redesigned version of an application to be used before, during, and after a museum visit.

The application already allows our users to read texts in the written form and with augmentative and alternative communication (AAC), listen to the content (with text-to-speech), see pictures, and give feedback input, such as comments and ratings. An image of the current application can be seen in the image below.



You will redesign the interface and menu items, add videos, possibility to record audio, integrate an Augmented Reality functionality, and a section to write annotations and create drawings. In the future, your work will be assessed by participants from an association of people with disabilities focusing on users with mild intellectual disabilities: real users will benefit from your contribution.

You will be under the supervision of Dr. Monica Landoni and Leandro Guedes, a Doctoral assistant at USI. Collaborations can also involve international researchers from different research areas.

Are you interested in this project? Contact us if you have any questions, we will be happy to answer them.

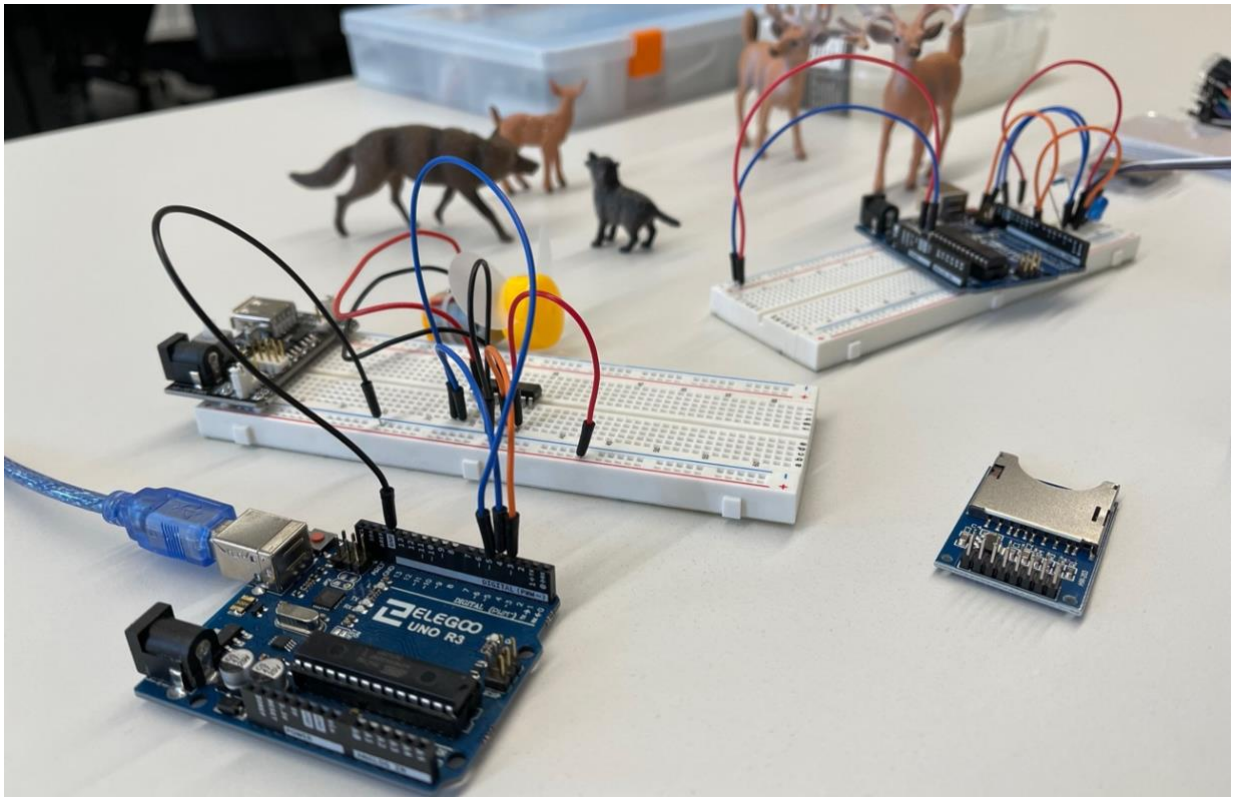
Multixp: Designing multisensory experiences for learning museum contents

UROP project proposal at the Università della Svizzera italiana (USI), Lugano, Switzerland

Most of our everyday experiences are multisensory in nature. What we can see, hear, feel, taste, smell, and much more are part of our daily lives. Almost any experience you can imagine, such as walking at the park or watching content on your smartphone, involves a magnificent sensory world.

Multisensory experiences, where the senses meet technology, are available in multiple scenarios, including museums and exhibitions. We can support them and enhance the user experience, adding gamification and learning to the process.

This summer project aims to use tangible user interfaces with visual and audio feedback to build a multimodal experience for users visiting an exhibition at a museum. Currently, we have a project in progress developing multisensory experiences for a Museum of Natural Sciences. We are using Arduino and a lot of creativity to build a multisensory environment where our users can play and feel different feedbacks. An image of some sensors and objects used in this project can be seen in the image below.



You will contribute by supporting the current project, designing, and developing brand-new solutions. This will help users to understand museum content and have fun with technology.

You will be under the supervision of Dr. Monica Landoni and Leandro Guedes, a Doctoral assistant at USI. The work will be assessed with Anffas Association and the Museum of Natural History of Trieste. Collaborations can involve international researchers from different research areas.

Are you interested in this project? Contact us if you have any questions, we will be happy to answer them.

The medial axis of a simple polygon in linear time.

The medial axis transform (MAT) is a well-known shape descriptor, which is used in numerous and diverse scientific areas. Given a simple polygon P, its medial axis is a skeleton contained in the Voronoi diagram of the sides of P as restricted in the interior of P. If the polygon P is convex, its medial axis can be computed in linear time by a very simple randomized incremental algorithm, which is known since the late 80s. If the polygon P is simple, its medial axis can still be computed in linear time, however, the available algorithms, which were developed in the late 90s, are too complicated to be of any practical use. Recently, we have generalized the original extremely simple randomized technique to general simple polygons. The randomized incremental algorithm remains surprisingly simple, almost as simple as the original one for convex shapes, and it can compute the medial axis of any simple polygon in expected linear time. We would like to bring this elegant technique to life through student projects in computational geometry. The technique is applicable to all kinds of Voronoi diagrams with a tree structure, thus, several extensions are possible for the interested student. The ideal candidate for UROP project should combine good algorithmic, programming, and analytical skills while having an interest in geometric or graph problems.

The UROP project would first focus on building the required simple data structure to keep a Delaunay graph (such data structures are available through the internet) which are essentially the same whether the polygon is convex or simple. Given the Delaunay graph, it would then visualize the Voronoi diagram. Once this infrastructure is in place we will focus on the so called "in-circle"test. To make it simple (and approximate) we will sample the polygon edges by points computing the Delaunay triangulation of fours such edges. Packages to compute the Delaunay triangulation of points are also available through the internet. After this additional infrastructure is also in place, we will then focus on the simple randomized incremental construction to compute the medial axis of a simple polygon in expected linear time (the time to answer the in circle test is counted separately).

Prof. Evanthia Papadopoulou

Learning shapes as signed distance functions for numerical simulations

What's a shape? Computationally, it could be many things: a collection of points and triangles, a patchwork of simpler shapes, a parametric equation, and so on. Signed distance functions encode a shape in their zero-level set. This representation is therefore very compact, since the signed distance function can smoothly map points in space to a single scalar number, the distance. For instance, the function $f(x,y,z) = x^2 + y^2 + z^2 - 1$ encodes the unit sphere.

This project aims at learning signed distance functions for general shapes. It is based on DeepSDF [1], a deep neural network that can encode signed distance functions. The network has two inputs: a code and a point coordinate and returns the distance of the point from the shape. Thus, the shape is easily recovered, for instance through the marching cubes algorithm. The training of the network consists in learning from a large set of shapes, all encoded as signed distance functions, so that the loss function is simply penalizing distances. DeepSDF is so powerful that can recover a full shape even from a sparse set of points partially covering it.

This project has very interesting implications in numerical simulations. Partial differential equations, a key tool in mathematical modeling, always comes with a computational domain (a bar, a mechanical piece, a human heart), that is hard to parametrize efficiently. DeepSDF provides a convenient way to efficiently encode computational domains. The whole numerical simulation could be even performed on the signed distance function itself, with a mesh-free method. Random sampling of shapes, e.g., for performing uncertainty quantification, consists in just sampling the code in the input layer.

The successful candidate can grow scientifically and to extend the project outcomes to a master thesis. Furthermore, the student will be able to benefit from existing numerical codes and computational resources at the Euler Institute and at the CSCS.

Prerequisites
Basic understanding of neural networks and Python programming language. Previous experience with PyTorch, TensorFlow or similar is good but not strictly necessary. Basic understanding of Calculus (function, derivatives) is welcome.

References
[1] Park et al., "DeepSDF: Learning Continuous Signed Distance Functions for Shape Representation", 2019, arXiv: 1901.05103v1

Contact
Simone Pezzuto simone.pezzuto@usi.ch and Rolf Krause rolf.krause@usi.ch, Euler Institute.

COMPUTING THE EVOLUTION OF SPECIES

Prof. Ernst C. Wit (wite@usi.ch) and Dr. Francisco Richter (richtf@usi.ch)

DESCIPTION

Dynamical networks and trees are commonly used to describe real evolutionary processes. Those networks are typically incomplete and performing statistical inference on available trees is computationally challenging. Well-designed data augmentation algorithms play a key role when dealing with missing-data scenarios, as advanced statistical methodologies rely heavily on them.

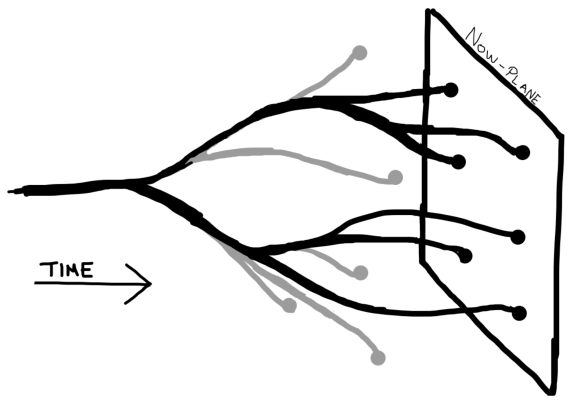


Figure 1: Evolutionary process are usually described by phylogenetic trees. A phylogenetic tree has two components, the topology (or shape) of the tree, and the timings of its nodes.

Statistical inference in species evolutionary processes is crucial for testing biological hypotheses in quantitative and system biology research. Some of the computational methodologies in evolutionary biology inference include EM-type of algorithms, Stochastic Gradient Descent and Bayesian approaches. At the moment, little is known about the relative computational efficiency of these approaches.

In this project, we will quantify the computational efficiency of different tree inference algorithms. For this you will develop your own code. The results from the study will be written up in a manuscript, evaluating various data science methodologies to perform inference on point processes involving trees and networks.

At the completion of the project the students will (1) be familiar with general data science methodologies, (2) understand real applications in evolutionary and theoretical biology, (3) have explored the field of statistical network science and (4) have developed expertise in algorithmic design and data augmentation algorithms for stochastic trees and networks.

DATA & TECHNICAL REQUIREMENTS

We will work with real phylogenies, which are characterized by surviving species, as extinct species are "missing". Therefore, as part of this project, millions of alternative phylogenetic trees need to be simulated efficiently. C(++) programming skills are required to implement these simulations efficiently. Some understanding of optimization techniques and numerical algorithms is required. Affinity with probability is essential to enjoy the project. A background in stochastic processes is an advantage, but can be learned on the job.

REFERENCES

Francisco Richter, Bart Haegeman, Rampal S Etienne, and Ernst C Wit. Introducing a general class of species diversification models for phylogenetic trees. *Statistica Neerlandica*, 74(3):261–274, 2020.

UROP 2022
PROJECT PROPOSAL
Title: Experimental evaluation of failure predictors for containerized cloud applications.
Mauro Pezzè

Summary

This UROP project is part of an ambitious research project, PREMISE, that aims to improve the quality of complex cloud systems by automatically predicting failures for dynamically-scaled, containerized cloud systems. The overall project involves a team of four researchers: Prof. Mauro Pezzè, Prof. Giovanni Denaro, and the PhD students Rahim Heydarov and Noura El Moussa. The UROP project will be carried on in collaboration with the research team, within the infrastructure currently available in the group.

The research goal of the overall PREMISE project is to study the application of leading-edge machine learning technology to the problem of automatically predicting failures of cloud-based applications. In the last years we defined approaches that learn the runtime profile of a target application based on dynamically monitored key-performance-indicators (KPIs), and exploit the learned profiles to discriminate correct executions from executions that can lead to failures. We also explored how to localize the specific components that are responsible for the upcoming failures, aiming to fix the issues before major disasters. We experimented with machine-learning technology based on deep-autoencoders, restricted boltzmann machines, time series analysis, and SVN classifiers, with excellent results. We developed a family of prediction tools, along with tools for runtime monitoring and fault seeding, and we validated our research hypotheses with experiments on large distributed applications.

We are now studying the applicability of the approach to dynamic-scaling, containerized cloud applications, such as applications based on the Kubernetes technology, the most common platform for delivering applications with high availability, performance, and reliability requirements. We are investigating the applicability of our approaches for predicting failures in the context of dynamic-scaling applications, which can both scale up by dynamically deploying additional containers to handle increasing workload, and scale down by dynamically dismissing some containers to free unused resources.

The UROP student will join the team in a set of experiments with an industrial-scale application running on Google cloud. The work consists of

- running experiments a large Kubernetes application that we already installed on Google cloud
- collecting data during both correct and failing executions by using our monitoring and fault-seeding tools
- using the collected data to experiment with our tools for predicting failures
- elaborating the data produced with the experiments.

The student will benefit from a full immersion experience in a leading research team, and will learn:

(i) cutting edge cloud technology, by experimenting with industrial-scale Kubernetes applications,
(ii) applications of cutting edge machine learning technology by experimenting with tools for predicting failures that exploit different machine learning approaches, and
(iii) future software engineering approaches that augment classic software engineering tools with machine and deep learning engines.

Essential references

[1] L. Gazzola, L. Mariani, F. Pastore, and M. Pezzè. An exploratory study of field failures. In Proceedings of the International Symposium on Software Reliability Engineering, 2017
[2] C. Monni, M. Pezzè, and G. Prisco. An RBM anomaly detector for the cloud. IEEE Conference on Software Testing, Validation and Verification, ICST 2019
[3] L. Mariani, M. Pezzè, O. Riganelli, and R. Xin. Predicting failures in multi-tier distributed systems. Journal of Systems and Software, 2020

Towards an Intelligent Post-training Mutation Tool for Deep Learning Systems

Paolo Tonella, Nargiz Humbatova
Software Institute@USI
paolo.tonella@usi.ch,nargiz.humbatova@usi.ch

Deep Learning (DL) has become an integral part of many groundbreaking projects and products we use everyday. As quality and safety remains the main concern for the developers and users of modern products based on Artificial Intelligence, different techniques aimed at assessing their quality are of increasing interest for research community.

Mutation Testing

Mutation testing is a technique that deliberately seeds faults in form of small syntactic changes into the program under test to create a set of faulty programs called mutants. The general principle underlying this approach is the assumption that faults used by mutation testing represent the mistakes that programmers usually make. Mutation testing aims to assess the quality of a given test suite in terms of its capability to detect faults. For this, the test suite is executed on each of the generated mutants. If the result for a given mutant is different from the result of running the original program then the mutation is considered killed. The ratio of killed mutants to the overall number of generated mutants is called mutation score. The higher the mutation score, the better is the quality of the test suite.

The example in Figure 1 shows a method `subtract` that subtracts two integer values and returns the result. It has two mutations: in `Mutant 1` the subtraction is replaced with multiplication and in `Mutant 2` it is replaced with addition. If our test suite has only `test0`, none of the two mutations would be killed (as they all return the expected value `0`) and the mutation score is 0%. If we add test case `test1()` to our test suite, then `Mutant 1` gets killed and the mutation score becomes 50%. Once we add test case `test2()`, both mutations get killed and the mutation score achieves its maximum value of 100%.

Mutation Testing for DL Systems

In traditional software systems the decision logic is often implemented by software developers in the form of source code. In contrast, the behaviour of a DL system is mostly determined by the training data set and the training program, i.e. these are the two major sources of defects for DL systems. Thus, there should be a specific approach to mutation testing of DL systems. There currently exist two tools that are designed specifically for performing mutation testing for DL systems. However, one of the tools is a pre-training one, which means it injects the faults into system prior to the training and thus is computationally expensive, while the second one, a post-training mutation tool, injects faults that are random and not very likely to happen in real world. Such faults usually introduce slight noise or modifications to a randomly selected subset of weights or change a structure of an already trained DL model by adding/deleting its layers or replacing the activation function.

```java
// Original Program
public int subtract(int a, int b) {
   return a - b;
}
```

```java
// Mutant 1
public int subtract(int a, int b) {
   return a * b;
}
```

```java
// Mutant 2
public int subtract(int a, int b) {
   return a + b;
}
```

```java

public void test0() {
   assertEquals(0, subtract(0, 0));
}

public void test1() {
   assertEquals(-4, subtract(-2, 2));
}

public void test2() {
   assertEquals(1, subtract(2,1));
}
```

Figure 1: Mutation Testing Example

Currently, mutation testing is being applied to various tasks for DL systems such as program repair, generation of optimal oracles for self-driving cars, detection of adversarial inputs, prioritisation of test inputs for the labelling, etc. Availability of a mutation tool that can inject changes that resemble the effect inflicted by real faults would be extremely useful also for these approaches as well as the advance of DL testing in general.

Project Proposal

The goal of the project is to develop a new post-training mutation tool that would solve the limitations set by the previous approaches, i.e. to introduce smarter and fast DL-specific mutation operators that produce stable and reliable results and would facilitate the increased interest for mutation testing in DL community.

In the frame of this project, the student will learn about state-of-the-art techniques in the domain of mutation testing for DL systems, their limitations and advantages. The student will practice with most popular DL frameworks and widely-used models and datasets.

Additional Information

The project will be carried out within the TAU research group at the Software Institute (https://www.si.usi.ch) and contribute to the PRECRIME ERC research project (https://www.pre-crime.eu). Students are supervised by researchers of the TAU group who follow them constantly and provide them with timely feedback, advice and directions. The code developed for the projects is typically released as an open source project and the results are often included in scientific publications. Both code and publication would contribute to a stronger CV of the participating student.

Explain the Feature space of DL Systems using Decision Trees

Summer Project Proposal

Paolo Tonella, Vincenzo Riccio, Tahereh Zohdinasab
Software Institute@USI
{paolo.tonella|vincenzo.riccio|tahereh.zohdinasab}@usi.ch

Project Proposal

The Software Engineering research community is working hard at adequately testing the functionality of Deep Learning (DL) systems by proposing a steadily growing number of approaches. Since part of the program logic of DL systems is determined by the training data, traditional code coverage metrics are not effective in determining whether DL systems have been adequately exercised. Therefore, recent testing solutions aim at maximising ad hoc white-box adequacy metrics, such as neuron or surprise coverage, or at exposing misbehaviours. A limitation of these approaches is that their output cannot be directly used to explain the behaviour of the DL system under test, e.g. coverage reports do not provide enough information to understand what input features might have caused misbehaviours.

Feature Maps (i.e., maps where the inputs are positioned based on their characteristics) allow to explain the root cause of misbehaviours. Another approach to explain DL systems are Decision Trees that learn the characteristics of the critical test scenarios and identify critical regions in an input space (i.e., the regions of a test input space that are likely to contain most critical test scenarios).

The goal of this project is to combine Feature Maps and Decision Trees. We will generate decision trees starting from the feature maps and use them as a model to highlight the critical regions of the map and guide the testing process.

Project Environment

The DL framework that will be used for this project is TensorFlow, while the main programming language is python. The DL-based software that will be tested is a lane keeping assist system for self-driving cars. To this aim, the students will evaluate the behaviour of the software by using a driving simulator. The driving simulator that has been chosen for the project is BeamNG.tech, a freely available research-oriented version of the driving game BeamNG.drive which features an accurate driving physics engine.

Prerequisites

We are looking for a student which is curious and motivated to contribute to the proposed project. We do not require prior knowledge on Deep Learning or Software Testing, as we expect the student to learn the basics during the project.

However, we expect the candidates to be curious, since they will apply what they have learned to design new solutions. Moreover, we expect them to be very precise, as they will perform an accurate comparison with the existing solutions.

During the project, the candidate will start from an existing software infrastructure and we will provide assistance when needed.

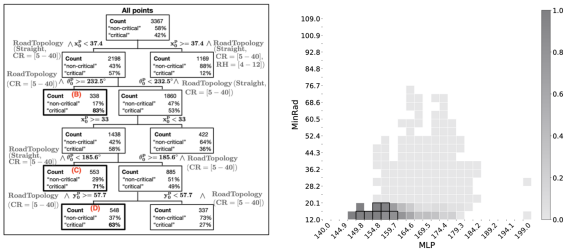


Figure 1: Sample decision tree (left) and feature map (right) for self driving software.

Tasks

For the successful completion of this project, the candidate is expected to perform a series of tasks, among which:

- study the relevant literature and become familiar with Decision Trees from Feature Maps;
- familiarize with the BeamNG driving simulator and the self-driving car DL software;
- implement a technique that derives Decision Trees from Feature Maps and uses them for testing DL systems;
- design an experiment to compare the implemented approach with other existing solutions.

Why You Should Choose This Project

This project is a great chance to gain some money during Summer, acquire multi-disciplinary skills, and hopefully have fun. You will learn and use machine learning concepts to train a self-driving car. You will use and adapt existing software engineering solutions to a relevant problem. Finally, you will acquire the ability to set up an experimental setting to rigorously compare different methods.

This experience will substantially enrich your CV with relevant skills and knowledge, which can be useful if you want to build your career either in academia or in a software company. Upon successful completion of the project, you will contribute to the open-source community and add this software project to your portfolio.

Further Information

Are you interested in this project but you want more details? We are happy to discuss our project with you - send us an email!

The proposed work is part of the Precrime (Self-assessment Oracles for Anticipatory Testing) ERC project. The interested student can find more information about Precrime on the project's website: https://www.pre-crime.eu/.

Neural Style Transfer-based Testing of Autonomous Driving Systems

Summer Project Proposal

Paolo Tonella, Andrea Stocco
Software Institute@USI
{paolo.tonella|matteo.biagiola|andrea.stocco}@usi.ch

Context

Self-driving cars are nowadays a reality. Most major manufacturers including Tesla, GM, Ford, Mercedes, BMW, and Waymo/Google are building and actively testing different types of autonomous vehicles. Recent results show that autonomous cars have become very efficient in practice and have already driven millions of miles without any human intervention.

The key component of an autonomous vehicle is the perception module controlled by Deep Neural Networks (DNNs). DNNs take input from different sensors like camera, light detection, ranging sensor, and infrared sensor, and outputs the steering angle, braking and other commands necessary to operate the car safely.

Preliminaries

DNN-based autonomous driving systems are tested using either (1) physics-based simulators that model the interactions with the environment; or, (2) data-driven simulators that use GANs to use or reproduce a real-world stream of driving data.

The former approach uses a simulation engine to render a virtualized road track in which the DNN-based autonomous driving systems is executed and tested for its capability to stay in lane (see Figure 1). While virtual testing with simulators has emerged as the de-facto practice, researchers are debating on whether the results obtained on a simulation environment are representative enough of real-world situations, as most simulators lacks the photo-realism and are often decoupled from the complexities of the real world.

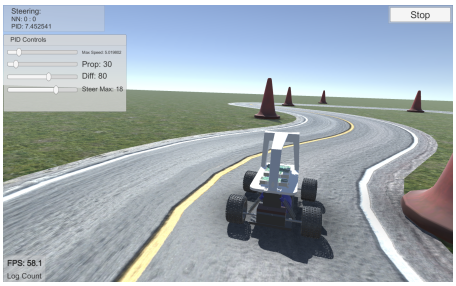


Figure 1: Physics-based virtual testing of self-driving cars.

The latter approach consists in simulators that focus entirely on the usage of real-world data to produce a continuous stream of images. Typically, the rendering part is made by generative adversarial networks (GANs) that are able to customize specific scenarios through latent-space transformations of the initial inputs and produce newly generated inputs that are consumed by the DNN-based autonomous driving systems. However, despite data-driven simulators offer as an interesting alternative, they only address the problem of photo-realism while leaving uncovered observing and exploiting the values of the physical interaction with the environment.

Figure 2: Data-driven virtual testing of self-driving cars.

Goal of the Project

The goal of the project is to tackle the gap between the two approaches described in the introduction and investigate the potential for hybridization, by implementing the capability of using real-world data in real-time within a physics-based simulator. The final outcome of the project will be the first proposition of neural-based testing framework for system-level testing of autonomous driving systems.

Approaches such as CycleGAN will be adopted to process simulated driving data and translate them into real-world driving data (from a given distribution), which will be sent back to the simulator. Effectiveness will be measured against in-field testing with a real small scale self-driving car based on the Donkey Car framework. The goal is to understand whether we can test a real-world autopilot *without the physical car* by means on an "enhanced simulated testing" platform.

In summary, the questions the thesis will aim to answer are: Can we enhance a driving simulator to use real-world data? Do the testing results obtained using such enhanced simulator correlate with those of in-field testing?

The results of the project are expected to increase the degree of support available to engineers in self-driving cars development and testing, and give the candidate experience on the following topics: supervised learning, neural translation, sim2real testing, empirical analysis.

Tasks

For the realization of this project, the candidate is expected to perform a series of tasks, among which:

- familiarize with the Donkey Car simulator, and the self-driving car models;
- familiarize with GAN-related concepts;
- train the CycleGAN model and implement the communication with the Donkey Car physics-based simulators based on Unity.

Prerequisites

We are looking for a student who is passionate and motivated to contribute to a project in the domain of self-driving cars. No prior knowledge of the simulation platform or the Donkey Car is required, nor on supervised learning or generative adversarial networks, but willingness to learn these technologies and adapt them depending on the project's needs. During the project, the candidate will build the project upon an existing software infrastructure that works with such algorithms, and we will work closely with you and provide assistance when needed.

Why You Should Choose This Project

This project is unique in its multi-disciplinary nature: you will learn and use concepts in deep learning (DL), software engineering and the ability to set up an experimental setting to rigorously compare different methods.

Upon successful completion of the project, you will have contributed to an open-source project that will allow developers to compare different self-driving car models based both on neural-based testing. As a result, you will be able to show on your CV practical experience with training and evaluating self-driving car models.

Further Information

The proposed work is part of the Precrime (Self-assessment Oracles for Anticipatory Testing) ERC project. The interested student can find more information about Precrime on the project's website: https://www.pre-crime.eu/. Interested in this project? Any still-unanswered questions? Drop us an email now!

Simulation-Based Testing of Social Networks by means of Autonomously Interacting Bots

Paolo Tonella
Software Institute, USI, Lugano, Switzerland
paolo.tonella@usi.ch

1 INTRODUCTION

Social networks are affecting vital aspects of our society that span from the personal sphere to the economy and health-care. Correspondingly, testing such platforms to avoid the emergence of undesirable or threatening behaviours has become a well recognized problem, for which no effective solution is available yet. Social network providers, such as Facebook, are well aware of the problem and are investigating approaches to test their social networks based on web-enabled simulations [1].

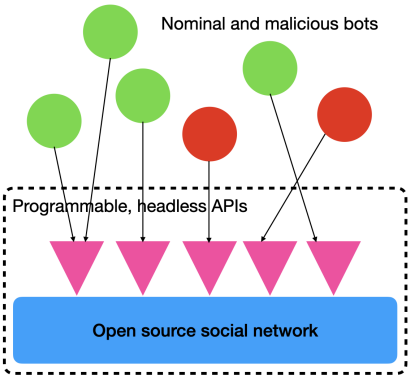


Figure 1: Architecture of the simulation platform

2 GOAL AND TASKS

The **goal** of this Summer project is to develop an infrastructure where a social network can be simulated programmatically and where autonomous bots with predefined or learned behaviours can be deployed (see Figure 1). Oracles on the health state of the social network should be also deployable into the simulated social network, so as to support the automated detection of emergent anomalous behaviours. The project will be organized into the following tasks:

Task 1 Study the relevant literature on simulation based testing and agent based testing.

Task 2 Assess the suitability for this project of existing open source social networks.

Task 3 Develop the social network infrastructure

Task 4 Develop autonomous bots that interact with each other through the social network infrastructure

2.1 Relevant literature

The seminal papers [1–3] by Facebook on web-enabled simulation are a cornerstone for this project. Other simulation based testing works can be found in the area of testing of self-driving cars [4, 7] and robotics testing [5].

The research area of agent oriented software development and agent testing is extremely rich. Among the relevant works, those considering not only agent development, but also agent testing [6] are particularly relevant for this project.

2.2 Existing platforms

A suitable open source social network is one that can be easily extended to operate in a completely headless mode (i.e., with no GUI) and that can be controlled completely by software bots (that replace humans). Other important suitability criteria are the quality and readability of the code that implements it, the maturity of the project, the number of project contributors and their level of activity and responsiveness. Reliability, stability and degree of testing of the platform are also important selection criteria.

As a starting point, the following platforms can be considered as candidate open source social networks:

- Diaspora (https://github.com/diaspora/diaspora)
- Elgg (https://github.com/Elgg/Elgg)
- Humhub (https://github.com/humhub/humhub)
- Bootcamp (https://github.com/vitorfs/bootcamp)

2.3 Infrastructure development

The social network selected at Task 2 should be extended so that it operates in headless mode under the control of bots. A proper API should be defined to support the interaction between bots and social network. The API should be abstract enough to be reusable across different platforms, so that in the future it would be possible to easily replace the selected social network platform with another one.

2.4 Bot development

The behaviour of these bots can be programmed via rules or can be learned either from examples of via reinforcement learning. In particular, two categories of bots will be developed: nominal bots and malicious bots. The former should behave by the rules of the social network and should mimic the behaviour of real human users, possibly observed in existing social networks. The latter should challenge the rules of the social network by behaving at or beyond the limits. An oracle on the healthy state of the network is also developed within Task 4 to recognize the occurrence of anomalous, undesirable behaviours. Such an oracle might take the form of a set of metrics for which the range of nominal values is known: whenever a value exceeding the nominal range is observed, a potential anomaly in the network is detected.

REFERENCES

[1] John Ahlgren, Maria Eugenia Berezin, Kinga Bojarczuk, Elena Dulskyte, Inna Dvortsova, Johann George, Natalija Gucevska, Mark Harman, Ralf Lämmel, Erik Meijer, Silvia Sapora, and Justin Spahr-Summers. 2020. WES: Agent-based User Interaction Simulation on Real Infrastructure. In *ICSE '20: 42nd International Conference on Software Engineering, Workshops, Seoul, Republic of Korea, 27 June - 19 July, 2020*. 276–284.

[2] John Ahlgren, Maria Eugenia Berezin, Kinga Bojarczuk, Elena Dulskyte, Inna Dvortsova, Johann George, Natalija Gucevska, Mark Harman, Maria Lomeli, Erik Meijer, Silvia Sapora, and Justin Spahr-Summers. 2021. Testing Web Enabled Simulation at Scale Using Metamorphic Testing. In *43rd IEEE/ACM International Conference on Software Engineering: Software Engineering in Practice, ICSE (SEIP) 2021, Madrid, Spain, May 25-28, 2021*. 140–149.

[3] Kinga Bojarczuk, Natalija Gucevska, Simon M. M. Lucas, Inna Dvortsova, Mark Harman, Erik Meijer, Silvia Sapora, Johann George, Maria Lomeli, and Rubmary Rojas. 2021. Measurement Challenges for Cyber Cyber Digital Twins: Experiences from the Deployment of Facebook's WW Simulation System. In *ESEM '21: ACM / IEEE International Symposium on Empirical Software Engineering and Measurement, Bari, Italy, October 11-15, 2021*. 2:1–2:10.

[4] Alessio Gambi, Marc Müller, and Gordon Fraser. 2019. Automatically testing self-driving cars with search-based procedural content generation. In *Proceedings of the 28th ACM SIGSOFT International Symposium on Software Testing and Analysis, ISSTA 2019, Beijing, China, July 15-19, 2019*. 318–328.

[5] Carl Hildebrandt, Sebastian G. Elbaum, Nicola Bezzo, and Matthew B. Dwyer. 2020. Feasible and stressful trajectory generation for mobile robots. In *ISSTA '20: 29th ACM SIGSOFT International Symposium on Software Testing and Analysis, Virtual Event, USA, July 18-22, 2020*. 349–362.

[6] Cu D. Nguyen, Anna Perini, Paolo Tonella, Simon Miles, Mark Harman, and Michael Luck. 2009. Evolutionary testing of autonomous software agents. In *8th International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS 2009), Budapest, Hungary, May 10-15, 2009, Volume 1*. 521–528.

[7] Vincenzo Riccio and Paolo Tonella. 2020. Model-based exploration of the frontier of behaviours for deep learning system testing. In *ESEC/FSE '20: 28th ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering, Virtual Event, USA, November 8-13, 2020*. 876–888.

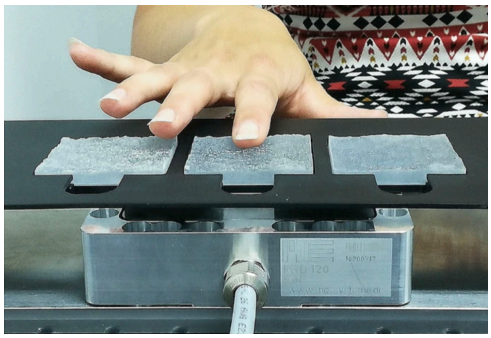
Haptics fabrication

For more details contact: Piotr Didyk (piotr.didyk@usi.ch)
Group website: https://www.pdf.inf.usi.ch/

Motivation

Novel fabrication devices, such as 3D printers enable precise geometry control. Recently, a significant amount of research has been dedicated to reproducing objects' appearance. Unfortunately, little is known regarding using these new fabrication tools to produce specific haptic properties. For example, imagine you want to 3D print a prototype of the newest model of a phone, which has a back-cover made of leather. While there is a body of work on how we can address the geometry and appearance, little is known on how to optimize the object such that it gives an impression of touching specific material.

Goal



The first goal of the project is to investigate the abilities of an observer in distinguishing various geometrical features of 3D surfaces, such as sinusoidal corrugations or Gabor patches. Next, we would like to build a computational model which predicts the perceived differences for general and more complex shapes than those covered by the initial investigations. Finally, we would like to employ the model to optimize 3D shapes such that their haptic (touch) properties are similar to those that we want to reproduce. In the beginning, we would like to consider fabrication of the samples using a high-precision (1-micron resolution) milling machine and develop the prediction model as well as the optimization for such tools. Later, we would like to extend it to 3D printers.

Additional information

This is a research-oriented project. If successful, we envision it to become a part of a submission to one of the top-tear computer graphics conferences such as ACM SIGGRAPH. The research in this project will be carried out as a part of our SNF project.

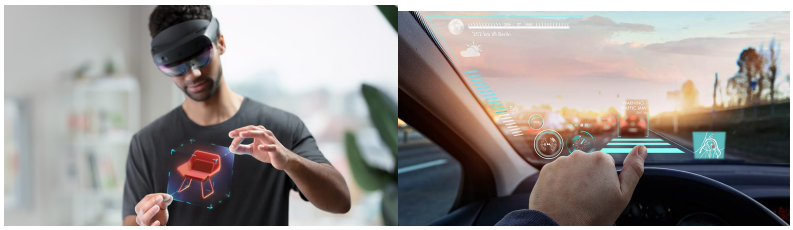
Prerequisites:

- Good programming skills
- Experience in processing geometrical information
- Interest in prototyping using devices such as 3D printers and milling machines

Image-processing for augmented reality

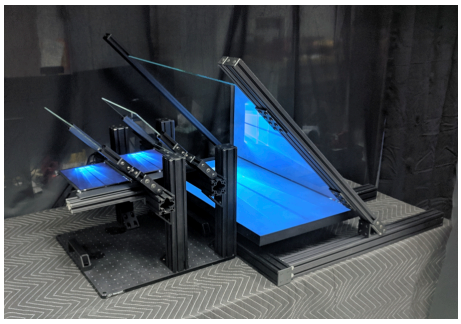
For more details contact: Piotr Didyk (piotr.didyk@usi.ch)
Group website: https://www.pdf.inf.usi.ch/

Motivation



Additive augmented reality displays overlay digital content atop the real-world image. From the viewer's point of view, the perceived image is a sum of the image provided by the augmented reality system and the real world, therefore the term additive. While this technology is popular and used by many devices, e.g., Microsoft HoloLens or head-up displays in cars, it has many problems. One of the main ones is the occlusion problem. The digital image never represents an opaque object appearance, i.e., the virtual objects are mixed with the real world behind them.

Goal



The goal of this project is to explore ways of enhancing the virtual image to improve its visibility compromising the quality of the real-world image. We will build a simple augmented reality system based on a beamsplitter with an integrated eye tracker and a separate camera observing the real world. Then, we will investigate image processing operations, such as contrast enhancement or attenuation, and apply them according to the content of the natural and virtual images as well as where the observer is looking. The final results of the projects should be a complete system (image-processing pipeline + display) which improves the quality of additive augmented reality displays. As an extension, we envision demonstrating performance of such a system on a high-end augmented reality headset, such as, Microsoft HoloLens 2.

Additional information

This is a research-oriented project. If successful, we envision it to become a part of a submission to one of the top-tear computer graphics conferences such as ACM SIGGRAPH. The research in this project will be carried out as a part of our ERC project. Additionally, it will be generously supported by Facebook Research.

Prerequisites:

- Good programming skills
- Basic knowledge on image processing
- The will of building and using experimental display system

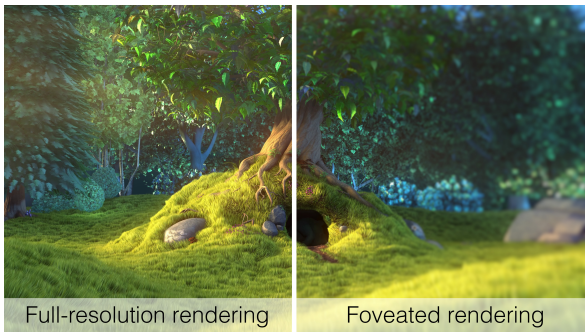
Efficient rendering for VR and AR devices

For more details contact: Piotr Didyk (piotr.didyk@usi.ch)
Group website: https://www.pdf.inf.usi.ch/

Motivation

Generating images for novel virtual and augmented reality headsets is computationally expensive due to the spatial resolution, frame rate, and image quality requirements imposed by the hardware and the human visual system. At the same time, when rendering for these wide-field-of-view devices, not everything can be appreciated by a viewer. In particular, we can achieve significant computational gains without significant quality loss by rendering lower quality for peripheral vision, the so-called foveated rendering approach. However, it is still an open research question what information has to be rendered to guarantee image quality matching the requirements imposed by the hardware and viewer.

Goal



The main goal of this project is to push the boundaries of foveated rendering, which is considered a key enabler of future VR and AR displays. This project aims to develop new image-based enhancement techniques for foveated rendering. More precisely, we postulate that we can generate images of significantly lower quality (low spatial and temporal resolution, inaccurate depth, etc.) and then enhance them using simple image-processing or image-based methods to make them indistinguishable from full-quality rendering. The main focus is on real-time techniques which can replace high-quality rendering. Therefore, also the techniques have to be simple. However, we can take advantage of any byproducts of the rendering procedure, e.g., texture information, depth map, optical flow, or any pre-computed information. We envision the final results of the project to be demonstrated in the end-to-end system consisting of low-cost rendering, enhancement, and display on one of the latest VR and AR devices, such as Varjo VR-3 (https://varjo.com/products/vr-3/).

Additional information

This is a research-oriented project. If successful, we envision it to become a part of a submission to one of the top-tear computer graphics conferences such as ACM SIGGRAPH. The research in this project will be carried out as a part of our ERC project.

Prerequisites:

- Good programming skills
- Experience with graphics pipeline, and rendering API such as OpenGL
- Basic knowledge on image processing

Robin - ROBot for Interactive Narratives to engage children

Storytelling and narrative skills are essential foundations for children to develop essential reading and writing competences; even before they learn how to read, children love listening to stories and being read aloud by their parents and caregiver; however, not all adults have the time or skills to engage children in storytelling, as they might not know the right questions to ask.

While technology could prove a valid support in this task, parents are often wary of letting their children use too much technology, and they want to limit their screen time, by giving more space to screen-free options.

Recently, we ran a user study that allowed us to better understand children needs and preferences regarding storytelling: we found that children are better engaged with interactive activities, such as gamebooks or toy that allow for a back-and-forth conversation, and that at the same time children are strongly attracted by tangible interfaces such as robots, and specifically that the children's favourite was a robot named Tellie, a humanoid, genderless, soft robot with changing lights, able to tell stories, ask questions and play songs.

We propose to design and develop a robot that will bring together the best of these worlds: the interactivity of gamebooks read by skilled adult readers, the aesthetic and functionality of a robot designed according to children's needs and preferences, and a screen-free experience to reassure parents.

If you are interested in:

- Learning how to elicit and understand children's preferences regarding robots.
- Working on the design of a brand-new robot for children, from the aesthetic to the actual functionality.
- Learning how to work with Arduino modules, pixel matrices and other fun technology to create a true interactive experience for children.

This is the perfect project for you! You will work together with a team of researchers, developing your coding and designing skills, and you will participate in a real-world research project with real users and stakeholders!