2016

December

Matthieu DERU
TSwoozy: Intelligente und benutzerzentrierte Interaktionen für das semantische Fernsehen
(Advisor: Prof. Wolfgang Wahlster)
Wed, 21.12.2016, 14:15h, building D3 2, Reuse seminar room

Diese Arbeit befasst sich mit der Thematik der parallelen Benutzung der Medien Fernsehen und Internet bzw. des semantischen Web, mit dem Fokus auf einer Suche und benutzerzentrierten Interaktionen, die intuitiv vorm Fernsehgerät durchgeführt werden können. Ziel ist es, Zuschauern eine neue Möglichkeit zu geben mit Videoinhalten zu inter-agieren und effizient nach Hintergrundwissen zu suchen. Hierbei wird der Fokus auf die Extraktion und Gewinnung von Semantik aus einem Video bzw. Live-TV-Programm durch automatische Annotationsverfahren und kombinierte Extraktionsverfahren aus den Bereichen Text- und Videoanalyse gesetzt. Daraus entsteht das Konzept des neuen interaktiven semantischen Fernsehens (Semantic TV). In der vorliegenden Arbeit liegen die Schwerpunkte auf der Konzeption dieses Systems, der technischen Realisierung einer Verbindung zwischen (nicht- bzw. linearen) Videos, den Cloud-basierten Diensten und der semantischen Verarbeitung.

Christopher HACCIUS
Towards a Synthetic World
(Advisor: Prof. Thorsten Herfet)
Wed, 21.12.2016, 10:15h, building C6 3, room 10.10

Visual information needs to fulfill various requirements. On the one hand such information should be realistic, while on the other hand it should be easily modifiable. Both of these requirements are met by different kinds of visual data: Computer generated imagery can be modified, but is generally less realistic than captured content, whereas captured data is photo-realistic, but merely modifiable. This thesis approaches the interplay of generated and captured data in different ways, with the goal of joining the advantages of both kinds. Towards this goal we introduce a representation format which allows the merge of content coming from different data sources. We have advanced algorithmic approaches that enhance captured data by information comparable to generated data. On the basis of our novel representation various processing steps can be implemented, of which some are presented here. An important contribution is our analysis of subjectively perceived quality of augmented and virtual reality scenarios, and an approach to determine this quality algorithmically.

Hassan HATEFI ARDAKANI
Analysis of Markov Automata on Finite Horizon
(Advisor: Prof. Holger Hermanns)
Tues, 20.12.2016, 14:00h, building E1 7, room 001

Markov automata constitute an expressive continuous-time compositional modelling for-malism, featuring stochastic timing and nondeterministic as well as probabilistic branching, all supported in one model. We study mathematical foundations of Markov automata since these are essential for the analysis addressed in this thesis. This includes, in particular, understanding their measurability and establishing their probability measure. Furthermore, we address the analysis of Markov automata in the presence of both reward acquisition and resource consumption within a finite budget of resources. More specifically, we put the problem of computing the optimal expected resource-bounded reward in our focus. We develop a sound theory together with a stable approximation scheme with a strict error bound to solve the problem in an efficient way. We report on an implementation of our approach in a supporting tool and also demonstrate its effectiveness and usability over an extensive collection of industrial and academic case studies.

Srinath SRIDHAR
Tracking Hands in Action for Gesture-based Computer Input
(Advisor: Prof. Christian Theobalt)
Fri, 16.12.2016, 16:00h, building E1 4, room 019

This thesis introduces new methods for markerless tracking of the full articulated motion of hands and for informing the design of gesture-based computer input. Emerging devices such as smartwatches or virtual/augmented reality glasses are in need of new input devices for interaction on the move. The highly dexterous human hands could provide an always-on input capability without the actual need to carry a physical device. First, we present novel methods to address the hard computer vision-based hand tracking problem under varying number of cameras, viewpoints, and run-time requirements. Second, we contribute to the design of gesture-based interaction techniques by presenting heuristic and computational approaches. The contributions of this thesis allow users to effectively interact with computers through markerless tracking of hands and objects in desktop, mobile, and egocentric

scenarios.

Helge RHODIN
From motion capture to interactive virtual worlds
(Advisor: Prof. Christian Theobalt)
Thu, 15.12.2016, 18:00h, building E1 4, room 019

Performance-driven character animation is chosen as a representative application and motion capture algorithms and animation methods are advanced to meet its high demands. Limitations of existing approaches are relaxed, such as coarse resolution and restricted capture volume, the dependence on expensive and complex multi-camera systems, and the use of intrusive suits and controllers. For motion capture, set-up time is reduced using fewer cameras, accuracy is increased despite occlusions and general environments, initialization is automated, and free roaming is enabled by egocentric cameras. For animation, increased robustness enables the use of low-cost sensors input, custom control gesture definition is guided to support novice users, and animation expressiveness is increased. The important contributions are: 1) an analytic and differentiable visibility model for pose optimization under strong occlusions, 2) a volumetric contour model for automatic actor initialization in general scenes, 3) a method to annotate and augment image-pose databases automatically, 4) the utilization of unlabeled examples for character control, and 5) the generalization and disambiguation of cyclical gestures for faithful character animation. In summary, the whole process of human motion capture, processing, and application to animation is advanced. These advances on the state of the art have the potential to improve many interactive applications, within and outside of virtual reality.

Endre PALATINUS
Physical Design in Databases
(Advisor: Prof. Jens Dittrich)
Thu, 15.12.2016, 12:00h, building E1 1, room 407

We live in an age where data has become one of the most important assets of companies. Data in itself is valuable, yet it has to be turned into information to become useful. This is were database management systems come into the picture. They allow for efficient processing of even terabytes of data, and thus provide the basis of knowledge extraction and information retrieval. A high-performance database system is an essential requirement for making big data analysis possible. The performance of database systems can be improved at multiple levels of the system, and using various approaches. In this work we focus on data layouts, and also investigate the performance implications of compiling hand-written queries and whole database systems as well. We present an exhaustive experimental study on vertical partitioning algorithms. Vertical partitioning itself is a physical design technique to partition a given logical relation into a set of physical tables, called vertical partitions. It is a crucial step in physical database design in legacy row-oriented databases. We show a survey of query processing on top of flat files, which are text files containing data encoded in some standard text format. Flat files are commonly used in various fields of science to store experimental results in a human-readable format. We explore the performance implications of compiling both hand-written queries, and whole database systems as well. We present two techniques for improving query performance that build upon changing compiler setups, and apply them in a main-memory database system.

Marek KöSTA
New Concepts for Real Quantifier Elimination by Virtual Substitution
(Advisor: Dr. Thomas Sturm)
Tue, 13.12.2016, 11:00h, building E1 5, room 002

This thesis studies quantifier elimination for the reals. We propose a novel, and practically applicable framework for real quantifier elimination using virtual substitution without any limitations on the possible degrees of quantified variables. That framework corresponds to a generic algorithm parameterized with substitution tables up to a chosen degree bound. We explicitly give such tables for degrees up to 3, which are optimized using our newly developed technique of clustering. On the algorithmic side our algorithm features structural elimination sets, which take into consideration not only the contained atoms but the Boolean structure of the input formula. Within this general framework we have furthermore developed improvements for various existing concepts, including degree shifts for heuristically pushing the degree bounds and extended quantifier elimination for computing sample values with, possibly parametric, satisfiability problems. An implementation of our algorithm up to degree 3 is available in the Redlog system, which is distributed with the open-source computer algebra system Reduce.

Fabienne EIGNER
A Theory of Types for Security and Privacy
(Advisor: Prof. Matteo Maffei)
Mon, 12.12.2016, 14:00h, building E9 1, room 001

Cryptographic protocols are ubiquitous in the modern web. However, they are notoriously difficult to design and their manual security analysis is both tedious and error-prone. Due to the lack of rigorous security proofs, many protocols have been discovered to be flawed. To improve the security and privacy guarantees of cryptographic protocols and their implementations and to facilitate their verification, a lot of research has been directed towards the formal analysis of such protocols. This has led to the development of several automated tools based on symbolic abstractions of cryptography. Unfortunately, there are still various cryptographic protocols and properties that are out of the scope of current systems.
This thesis introduces three novel frameworks for the verification of security protocols and their implementations based on powerful types for security and privacy, overcoming the limitations of current state-of-the-art approaches. With AF7 we present the first type system that statically enforces the safety of cryptographic protocol implementations with respect to authorization policies expressed in affine logic. Furthermore, our novel approach for the automated analysis of e-voting systems based on refinement type systems can be used to enforce both privacy and verifiability. Finally, with DF7, we present the first affine, distance-aware type system to statically and automatically enforce distributed differential privacy in cryptographic

protocol implementations.

Felix SCHUHKNECHT
Closing the Circle of Algorithmic and System-centric Database Optimization: A Comprehensive Survey on Adaptive Indexing, Data Partitioning, and the Rewiring of Virtual Memory
therapies
(Advisor: Prof. Jens Dittrich)
Thu, 8.12.2016, 16:30h, building E1 1, room 407

Over the decades, with the increase of computing resources, the amount of data to manage also increased tremendously. Besides of the sheer quantity of information, the quality of it highly varies today. Indexing all this data with equal effort is cumbersome and wasteful. Thus, adaptive indexing algorithms refine parts of interest more carefully. Unfortunately, the adaptivity also introduces a set of new problems. High variance in response times and low robustness against certain workloads are just two issues to mention. A vast amount of methods have been proposed to deal with these problems. Thus, in the first part of this thesis, we will reinvestigate, analyze, and enhance the class of adaptive indexing methods in a comprehensive evaluation on the algorithmic level. In total, we discuss 18 cracking methods, 6 sorting algorithms, and 3 full index structures, including our own proposed methods. Consequently, we identify data partitioning as the common component. Thus, in the second part, we analyze the surprising amount of optimizations possible to enhance partitioning. Interestingly, they mostly originate from a more sophisticated mapping of the method to the system properties, thus shifting our perspective to a system-centric view. Subsequently, in the third part, we dig down to the ground level by exploiting a core feature of any modern operating system, the virtual memory system. We investigate how virtual and physical memory can be separated in user space and how the mappings between the two memory types can be rewired freely at run-time. Using rewiring, we are able to significantly enhance core applications of data management systems. Finally, we apply the techniques identified in this thesis to the initial adaptive indexing algorithm to significantly improve it — and close the circle.

Alejandro PIRONTI
Improving and validating data-driven genotypic interpretation systems for the selection of antiretroviral
therapies
(Advisor: Prof. Thomas Lengauer)
Wed, 7.12.2016, 16:00h, building E2 1, room 001

Infection with Human immunodeficiency virus type 1 (HIV-1) requires treatment with combinations of antiretroviral drugs. These combinations must be selected under consideration of the drug resistance of the virus and of the prospects that the drug combination has for attaining sustained therapeutic success. Genotypic drug-resistance interpretation systems are frequently used for selecting combinations of antiretroviral drug compounds.
In this work, I present and validate novel, interpretable methods for deriving genotype interpretation systems that are trained on HIV-1 data from routine clinical practice. These genotype interpretation systems can be used for predicting drug exposure, drug resistance, and the time a certain combination of antiretroviral drugs will remain effective. Validation of the methods shows that their performance is comparable or, most frequently, superior to that of previously available methods. Their data-driven methodology allows for automatic retraining without the need for expert intervention. Their interpretability helps them gain the confidence of the users and delivers a rationale for predictions that could be considered surprising. Last but not least, the ability of the therapy-success interpretation system to consider cumulative, long-term therapeutic success allows it to produce predictions that are in line with the results of clinical studies.

Bernhard REINERT
Interactive, Example-driven Synthesis and Manipulation of Visual Media
(Advisor: Prof. Hans-Peter Seidel)
Fri, 2.12.2016, 15:15h, building E1 4, room 019

This thesis proposes several novel techniques for interactive, example-driven synthesis and manipulation of visual media. The numerous display devices in our everyday lives make visual media, such as images, videos, or three-dimensional models, easily accessible to a large group of people. Consequently, there is a rising demand for efficient generation of synthetic visual content and its manipulation, especially by casual users operating on low-end, mobile devices. Off-the-shelf software supporting such tasks typically requires extensive training and in-depth understanding of the underlying concepts of content acquisition, on the one hand, and runs only on powerful desktop machines, on the other hand, limiting the possibility of artistic media generation to a small group of trained experts with appropriate hardware. Our proposed techniques aim to alleviate these requirements by allowing casual users to synthesize complex, high-quality content in real-time as well as to manipulate it by means of simple, example-driven interactions.

November

Daniel STÖCKEL
Bioinformatics methods for the genetic and molecular characterisation of cancer
(Advisor: Prof. Hans-Peter Lenhof)
Fri, 25.11.2016, 15:15h, building E2 1, room 007

Cancer is a class of complex, heterogeneous diseases of which many types have proven to be difficult to treat due to the high genetic variability between and within tumours. To improve therapy, some cases require a thorough genetic and molecular characterisation that allows to identify mutations and pathogenic processes playing a central role for the development of the disease. Data obtained from modern, biological high-throughput experiments can offer valuable insights in this regard. Therefore, we developed a range of interoperable approaches that support the analysis of high-throughput datasets on multiple levels of detail. Mutations are a

main driving force behind the development of cancer. To assess their impact on an affected protein, we designed BALL-SNP which allows to visualise and analyse single nucleotide variants in a structure context. For modelling the effect of mutations on biological processes we created CausalTrail which is based on causal Bayesian networks and the do-calculus. Using NetworkTrail, our web service for the detection of deregulated subgraphs, candidate processes for this investigation can be identified. Moreover, we implemented GeneTrail2 for uncovering deregulated processes in the form of biological categories using enrichment approaches. With support for more than 46,000 categories and 13 set-level statistics, GeneTrail2 is the currently most comprehensive web service for this purpose. Based on the analyses provided by NetworkTrail and GeneTrail2 as well as knowledge from third-party databases, we built DrugTargetInspector, a tool for the detection and analysis of mutated and deregulated drug targets. We validated our tools using a Wilm's tumour expression dataset and were able to identify pathogenic mechanisms that may be responsible for the malignancy of the blastemal tumor subtype and might offer opportunities for the development of novel therapeutic approaches.

Tomáš DAVIDOVIč
Light Transport Simulation on Special Hardware
(Advisor: Prof. Philipp Slusallek)
Fri, 25.11.2016, 13:00h, building D3 2, DFKI VisCenter

In this thesis we examine the hardware-algorithm synergies in the context of ray tracing and Monte-Carlo algorithms. First, we focus on the very basic element of all such algorithms - the casting of rays through a scene, and propose a dedicated hardware unit to accelerate this common operation. Then, we examine existing and novel implementations of many Monte-Carlo rendering algorithms on massively parallel hardware, as full hardware utilization is essential for peak performance. Lastly, we present an algorithm for tackling complex interreflections of glossy materials, which is designed to utilize both powerful processing units present in almost all current computers: the Central Processing Unit (CPU) and the Graphics Processing Unit (GPU). These three pieces combined show that it is always important to look at hardware-algorithm mapping on all levels of abstraction: instruction, processor, and machine.

Noram AZMY
A Machine–Checked Proof of Correctness of Pastry
(Advisor: Prof. Christoph Weidenbach / Prof. Stephan Merz, co-tutelle Univ. de Lorraine)
Thu, 24.11.2016, 16:00h, building E1 4, room 001

Protocols implemented on overlay networks in a peer-to-peer (P2P) setting promise flexibility, performance, and scalability due to the possibility for nodes to join and leave the network while the protocol is running. These protocols must ensure that all nodes maintain a consistent view of the network, in the absence of a centralized control, so that requests can be routed to the intended destination. This aspect represents an interesting target for formal verification. In previous work, Lu studied the Pastry algorithm for implementing a distributed hash table DHT) over a P2P network. He suggested a variant of the algorithm, together with a machine-checked proof in the TLA+ Proof System (TLAPS), assuming the absence of node failures. Building on this work, this thesis makes three contributions. First, I identify and correct problems in Lu's roof, due to unchecked assumptions concerning modulus arithmetic and underlying data structures. Second, I introduce abstractions that help improve the degree of automation. Third, I formally prove that a simplified version of Lu's algorithm, in which the final phase of the join protocol is omitted, is still correct, again assuming that nodes do not fail.

Sebastian MEISER
Quantitative Anonymity Guarantees for Tor
(Advisor: Prof. Michael Backes)
Wed, 23.11.2016, 16:00h, building E9 1, room 001

We present a methodology to compute sound anonymity guarantees for anonymous communication protocols, which we apply to the Tor protocol. To this end, we present AnoA, a formal framework for quantifying the anonymity of anonymous communication protocols against modularly describable adversaries. We show how the Tor protocol can be analyzed in an abstract way by utilizing an existing formalization of Tor in the universal composability (UC) framework and prove that this analysis is sound. Moreover, we derive efficiently computable, sound anonymity guarantees for Tor against passively eavesdropping adversaries that may control Tor servers or wiretap communications between parties. For known adversaries that perform perfect traffic correlations we show that our bounds are tight. Finally, we perform an extensive analysis of the Tor network against such adversaries in which we calculate the anonymity of Tor when using Tor's current path selection, thereby considering many real-world properties of the path selection algorithm, as well as when using one of several variants, against a variety of eavesdropping adversaries that can control Tor nodes (depending on various properties) or wiretap the communication between them.

Sergio ROA-OVALLE
Autonomous learning of object behavior concepts and models through robotic interaction
(Advisor: Prof. Hans Uszkoreit)
Wed, 16.11.2016, 16:00h, building D3 2, Reuse seminar room

We need robots that learn from the environment by interacting with it, and deduce models of causal relations and associations from these interactions. In this dissertation, we address the particular problem of predicting object trajectories when manipulating objects, using robot arm pushes. To solve this problem we derive models which can describe the behavior of objects put in motion. For a learning robot it is essential to learn in an incremental and active way when new information is coming in, and to do so without losing generalization and without overfitting. First, we tackle this problem by estimating the density of a sensorimotor space after a robot performs a new action by using a modification of the incremental Growing Neural Gas (RobustGNG) algorithm. RobustGNG performs a space quantization which is robust to noise and overfitting issues. Subsequently, we infer models useful

for prediction of object trajectories in terms of object poses. The same machinery is useful for obtaining more coarse-grained predictions, for instance categorizations of object behaviors. Last but not least, these prediction models provide a qualitative and temporal description of the state space, so that they can eventually be used in planning tasks. We infer cause-effect models by using RobustGNG's results in a new version of the CrySSMEx algorithm to generate substochastic finite-state machines.

Fritz MÜLLER
On confluence and semantic full abstraction of lambda calculus languages
(Advisor: Prof. Reinhard Wilhelm)
Mon, 14.11.2016, 16:00h, building E1 3, room 528

My thesis comprises four papers whose common background is the semantic full abstraction problem.
First, I repeat the definition of full abstraction in the context of the "toy" programming language PCF, a typed lambda calculus with integers and recursion on all types. Plotkin showed that Scott's semantic model of PCF becomes fully abstract, if we extend PCF by the "parallel if".
In my paper "Full abstraction for a recursively typed lambda calculus with parallel conditional", I transfer this procedure to a language with a more elaborated type system. To prove the confluence of this calculus, we need a new theorem, in my paper "Confluence of the lambda calculus with left-linear algebraic rewriting".
The other two papers will be only briefly summarized:"On Berry's conjectures about the stable order in PCF" and "From Sazonov's non-dcpo natural domains to closed directed-lub partial orders".

Petr KELLNHOFR
Perceptual modelling for stereoscopic 3D
(Advisor: Dr. habil. Karol Myskowski)
Fri, 04.11.2016, 15:00h, building E1 4, room 019

Virtual and Augmented Reality applications typically rely on both stereoscopic presentation and involve intensive object and observer motion. A combination of high dynamic range and stereoscopic capabilities become popular for consumer displays, and is a desirable functionality of head mounted displays to come. The thesis is focused on complex interactions between all these visual cues on digital displays. The first part investigates challenges of the stereoscopic 3D and motion combination. Motion of objects in the 3D world as well as self-motion of an observer are both considered. The second part focuses on the role of high dynamic range imaging for stereoscopic displays. We go beyond the current display capabilities by considering the full perceivable luminance range and we simulate the real world experience in such adaptation conditions. The core of our research methodology is perceptual modeling supported by our own experimental studies to overcome limitations of current display technologies and improve the viewer experience by enhancing perceived depth, reducing visual artifacts or improving viewing comfort.

Myroslav BACHYNSKYI
Biomechanical Models for Human-Computer Interaction
(Advisor: Prof. Jürgen Steimle)
Fri, 04.11.2016, 12:00h, building E1 4, room 019

Post-desktop user interfaces, such as smartphones, tablets, interactive tabletops, public displays and mid-air interfaces, already are a ubiquitous part of everyday human life, or have the potential to be. One of the key features of these interfaces is the reduced number or even absence of input movement constraints imposed by a device form-factor. This freedom is advantageous for users, allowing them to interact with computers using more natural limb movements; however, it is a source of 4 issues for research and design of post-desktop interfaces which make traditional analysis methods inefficient: the new movement space is orders of magnitude larger than the one analyzed for traditional desktops; the existing knowledge on post-desktop input methods is sparse and sporadic; the movement space is non-uniform with respect to performance; and traditional methods are ineffective or inefficient in tackling physical ergonomics pitfalls in post-desktop interfaces. These issues lead to the research problem of efficient assessment, analysis and design methods for high-throughput ergonomic post-desktop interfaces. To solve this research problem and support researchers and designers, this thesis proposes efficient experiment- and model-based assessment methods for post-desktop user interfaces. We achieve this through the following contributions:
- adopt optical motion capture and biomechanical simulation for HCI experiments as a versatile source of both performance and ergonomics data describing an input method;
- identify applicability limits of the method for a range of HCI tasks;
- validate the method outputs against ground truth recordings in typical HCI setting;
- demonstrate the added value of the method in analysis of performance and ergonomics of touchscreen devices; and
- summarize performance and ergonomics of a movement space through a clustering of physiological data.

Oktober

Syama Sundar Yadav RANGAPURAM
Graph-based methods for unsupervised and semi-supervised data analysis
(Advisor: Prof. Matthias Hein)
Mon, 11.10.2016, 15:00h, building E1 4, room 024

Clustering and community detection are two important problems in data analysis with applications in various disciplines. Often in practice, there exists prior knowledge that helps the process of data analysis. In this talk we present graph-based methods for these data analysis problems both in unsupervised and semi-supervised settings. The main advantage of our methods is that they provide a common framework for integrating soft as well as hard prior knowledge. In the latter case, ours is the first method to have provable guarantees on the satisfaction of the given prior knowledge. The foundation of our methods is the exact continuous relaxation result that we derive for a class of combinatorial optimization problems.

More specifically, we show that the (constrained) minimization of a ratio of non-negative set functions can be equivalently rewritten as a continuous optimization problem. Building on this result, we derive a novel continuous relaxation for the multiclass clustering problem which makes the integration of hard prior knowledge possible in the multiclass setting as well. We also present efficient algorithms for solving the continuous relaxations. While the global optimality is not guaranteed, in practice our methods consistently outperform the corresponding convex or spectral relaxations by a large margin. Moreover, our method has an additional guarantee that the solution respects the prior knowledge.

September

Laszlo KOZMA
Binary Search Trees, Rectangles and Patterns
(Advisor: Prof. Raimund Seidel)
Fri, 16.09.2016, 13:30h, building E1 7, room 0.01

The topic of the dissertation is the classical problem of searching for a sequence of keys in a binary search tree, allowing the re-arrangement of the tree after every search. Our current understanding of the power and limitations of this model is incomplete, despite decades of research. Sleator and Tarjan conjectured in 1983 that Splay tree, a simple online strategy for tree re-arrangement is as good, up to a constant factor, as the theoretical optimum, for every input. This is the famous dynamic optimality conjecture. Our main contributions to this topic are the following.
1. A generalization of Splay to a broad class of algorithms that share many of its known efficiency-properties. The class of algorithms is defined implicitly, through a set of sufficient (and, in a limited sense, necessary) conditions on the performed tree re-arrangements.
2. A study of search sequences in terms of their avoided patterns. It is shown that pattern-avoiding sequences can be served using binary search trees much faster than what the logarithmic worst-case guarantees would suggest. The result generalizes and complements known structural properties from the literature.
3. A novel interpretation of the binary search tree problem in terms of rectangulations, a well-studied structure in combinatorics and geometry. The connection yields new insight about rectangulations, as well as about other geometric structures such as manhattan networks, and brings new techniques to the study of dynamic optimality.

Beata TURONOVA
Progressive Stochastic Reconstruction Technique for Cryo Electron Tomography
(Advisor: Prof. Philipp Slusallek)
Thur, 15.09.2016, 15:00h, building E1 4, room 0.19

Cryo Electron Tomography (cryoET) plays an essential role in Structural Biology, as it is the only technique that allows us to study the structure of large macromolecular complexes in their close to native environment in-situ. The reconstruction process is a challenging task as the single-tilt acquisition scheme imposes severe limitation on the input projections. High-resolution protocols such as Subtomogram Averaging (SA) can alleviate some of these limitations. Results of these protocols are highly dependent on the quality of the reconstruction. State-of-the-art methods deliver low-contrast and noisy reconstructions, which complicates their processing during the SA.
In this work we focus on improvement of the quality of tomograms in cryoET in order to facilitate their subsequent processing. We propose a Progressive Stochastic Reconstruction Technique (PSRT) – a novel iterative approach to tomographic reconstruction in cryoET that is based on Monte Carlo random walks. We design a progressive scheme to suit conditions present in cryoET and integrate PSRT into the SA pipeline, where it delivers high-contrast reconstructions that significantly improve template-based localization without any loss of high-resolution structural information. Furthermore, we perform a systematic study of the geometry-related acquisition artifacts and draw recommendations regarding the mutual influence of these artifacts and implications to the interpretation of both cryoET and SA experiments.

Wei-Chen CHIU
Bayesian Non-Parametrics for Multi-Modal Segmentation
(Advisor: Dr. Mario Fritz)
Tues, 13.09.2016, 17:45, building E1 4, room 0.24

Segmentation is a fundamental and core problem in computer vision research which has applications in many tasks, such as object recognition, content-based image retrieval, and semantic labelling. To partition the data into groups coherent in one or more characteristics such as semantic classes, is often a first step towards understanding the content of data. As information in the real world is generally perceived in multiple modalities, segmentation performed on multi-modal data for extracting the latent structure usually encounters a challenge: how to combine features from multiple modalities and resolve accidental ambiguities.
This thesis tackles three main axes of multi-modal segmentation problems: video segmentation and object discovery, activity segmentation and discovery, and segmentation in 3D data.
For the first two axes, we introduce non-parametric Bayesian approaches for segmenting multi-modal data collections, including groups of videos and context sensor streams. The proposed method shows benefits on: integrating multiple features and data dependencies in a probabilistic formulation, inferring the number of clusters from data and hierarchical semantic partitions, as well as resolving ambiguities by joint segmentation across videos or streams.
The third axis focuses on the robust use of 3D information for various applications, as 3D perception provides richer geometric structure and holistic observation of the visual scene. The studies covered in this thesis for utilizing various types of 3D data include: 3D object segmentation based on Kinect depth sensing improved by cross-modal stereo, matching 3D CAD models to objects on 2D image plane by exploiting the differentiability of the HOG descriptor, segmenting stereo videos based on adaptive ensemble models, and fusing 2D object detectors with 3D context information for an augmented reality application scenario.

August

Niket TANDON
Commonsense Knowledge Acquisition and Applications
(Advisor: Prof. Gerhard Weikum)
Fri, 19.08.2016, 16:00h, building E1 4, room 024

Computers are increasingly expected to make smart decisions based on what humans consider commonsense. This would require computers to understand their environment, including properties of objects in the environment (e.g., a wheel is round), relations between objects (e.g., two wheels are part of a bike, or a bike is slower than a car) and interactions of objects (e.g., a driver drives a car on the road). The goal of this dissertation is to investigate automated methods for acquisition of large-scale, semantically organized commonsense knowledge. Prior state-of-the-art methods to acquire commonsense are either not automated or based on shallow representations. Thus, they cannot produce large-scale, semantically organized commonsense knowledge.
To achieve the goal, we divide the problem space into three research directions, constituting our core contributions:
1. Properties of objects: acquisition of properties like has Size, has Shape, etc. We develop WebChild, a semi-supervised method to compile semantically organized properties.
2. Relationships between objects: acquisition of relations like largerThan, partOf, memberOf, etc. We develop CMPKB, a linear-programming based method to compile comparative relations, and, we develop PWKB, a method based on statistical and logical inference to compile part-whole relations.
3. Interactions between objects: acquisition of activities like drive a car, park a car, etc., with attributes such as temporal or spatial attributes. We develop Knowlywood, a method based on semantic parsing and probabilistic graphical models to compile activity knowledge.
Together, these methods result in the construction of a large, clean and semantically organized Commonsense Knowledge Base that we call WebChild KB.

Marvin KÜNNEMANN
Tight(er) Bounds for Similarity Measures, Smoothed Approximation and Broadcasting
(Advisor: Prof. Benjamin Doerr, now Paris)
Wed, 03.08.2016, 13:00h, building E1 4, room 024

In this thesis, we prove upper and lower bounds on the complexity of sequence similarity measures, the approximability of geometric problems on realistic inputs, and the performance of randomized broadcasting protocols.
The first part approaches the question why a number of fundamental polynomial-time problems - specifically, Dynamic Time Warping, Longest Common Subsequence (LCS), and the Levenshtein distance – resists decades-long attempts to obtain polynomial improvements over their simple dynamic programming solutions. We prove that any (strongly) subquadratic algorithm for these and related sequence similarity measures would refute the Strong Exponential Time Hypothesis (SETH). Focusing particularly on LCS, we determine a tight running time bound (up to lower order factors and conditional on SETH) when the running time is expressed in terms of all input parameters that have been previously exploited in the extensive literature.
In the second part, we investigate the approximation performance of the popular 2-Opt heuristic for the Traveling Salesperson Problem using the smoothed analysis paradigm. For the Fréchet distance, we design an improved approximation algorithm for the natural input class of c-packed curves, matching a conditional lower bound.
Finally, in the third part we prove tighter performance bounds for processes that disseminate a piece of information, either as quickly as possible (rumor spreading) or as anonymously as possible (cryptogenography).

July

Nico PERSCH
Physically-Based Image Reconstruction
(Advisor: Prof. Joachim Weickert)
Tues, 26.07.2016, 16:15h, building E1 7, room 0.01

The reconstruction of perturbed or lost data is one of the fundamental challenges in image processing and computer vision. In this work, we focus on the physical imaging process and approximate it in terms of so-called forward operators. We consider the inversion of these mathematically sound formulations to be the goal of reconstruction. We approach this task with variational techniques where we tailor our methods to the specific physical limitations and weaknesses of different imaging processes.
The first part of this work is related to image processing. We propose an advanced reconstruction method for 3-D confocal and stimulated emission depletion STED microscopy imagery. To this end, we unify image denoising, deconvolution and anisotropic inpainting. The second part is related to computer vision: We propose a novel depth-from-defocus method and design a novel forward operator that preserves important physical properties. Our operator fits well into a variational framework. Moreover, we illustrate the benefits of a number of advanced concepts such as a joint depth-from-defocus and denoising approach as well as robustification strategies. Besides, we show the advantages of the multiplicative Euler-Lagrange formalism compared to the additive one. Synthetic and real-world experiments within the main chapters confirm the applicability and the performance of our methods.

Anne-Christin HAUSCHILD
Computational Methods for Breath Metabolomics in Clinical Diagnostics
(Advisor: Prof. Volkhard Helms)
Fri, 15.07.2016, 10:00h, building E1 4, room 0.24

For a long time, human odors and vapors have been known for their diagnostic power. Therefore, the analysis of the metabolic composition of human breath and

odors creates the opportunity for a non-invasive tool for clinical diagnostics. Innovative analytical technologies to capture the metabolic profile of a patient's breath are available, such as, for instance, the ion mobility spectrometry coupled to a multicapilary collumn. However, we are lacking automated systems to process, analyse and evaluate large clinical studies of the human exhaled air. To fill this gap, a number of computational challenges need to be addressed.

For instance, breath studies generate large amounts of heterogeneous data that requires automated preprocessing, peak-detection and identification as a basis for a sophisticated follow up analysis. In addition, generalizable statistical evaluation frameworks for the detection of breath biomarker profiles that are robust enough to be employed in routine clinical practice are necessary. In particular since breath metabolomics is susceptible to specific confounding factors and background noise, similar to other clinical diagnostics technologies.

Moreover, specific manifestations of disease stages and progression, may largely influence the breathomics profiles. To this end, this thesis will address these challenges to move towards more automatization and generalization in clinical breath research. In particular I present methods to support the search for biomarker profiles that enable a non-invasive detection of diseases, treatment optimization and prognosis to provide a new powerful tool for precision medicine.

Christopher SCHROERS
Variational Surface Reconstruction
(Advisor: Prof. Joachim Weickert)
Fri, 08.07.2016, 14:15h, building E1 7, room 001

The demand for capturing 3D models of real world objects or scenes has steadily increased in the past. Today, there are numerous developments that indicate an even greater importance in the future: Computer generated special effects are extensively used and highly benefit from such data, 3D printing is starting to become more affordable, and the ability to conveniently include 3D content in websites has quite matured. Thus, 3D reconstruction has been and still is one of the most important research topics in the area of computer vision.

Here, the reconstruction of a 3D model from a number of colour images with given camera poses is one of the most common tasks known as multi-view stereo. We contribute to the two main stages that arise in popular strategies for solving this problem: The estimation of depth maps from multiple views and the integration of multiple depth maps into a single watertight surface. Subsequently, we relax the constraint that the camera poses have to be known and present a novel pipeline for 3D reconstruction from image sequences that solely relies on dense ideas. It proves to be an interesting alternative to popular sparse approaches and leads to competitive results. When relying on sparse features, this only allows to estimate an oriented point cloud instead of a surface. To this end, we finally propose a general higher order framework for the surface reconstruction from oriented points.

Sebastian OTT
Algorithms for Classical and Modern Scheduling Problems
(Advisor: Prof. Kurt Mehlhorn)
Mon, 04.07.2016, 9:30h, building E1 4, room 0.24

Subject of this thesis is the design and the analysis of algorithms for scheduling problems. In the first part, we focus on energy-efficient scheduling, where one seeks to minimize the energy needed for processing certain jobs via dynamic adjustments of the processing speed (speed scaling). We consider variations and extensions of the standard model introduced by Yao, Demers, and Shenker in 1995 [79], including the addition of a sleep state, the avoidance of preemption, and variable speed limits. In the second part, we look at classical makespan scheduling, where one aims to minimize the time in which certain jobs can be completed. We consider the restricted assignment model, where each job can only be processed by a specific subset of the given machines. For a special case of this problem, namely when heavy jobs can go to at most two machines, we present a combinatorial algorithm with approximation ratio strictly better than two.

June

Andrey KUPRIYANOV
Causality-based Verification
(Advisor: Prof. Bernd Finkbeiner)
Tues, 28.06.2016, 11:00h, building E1 7, room 001

Two aspects make program verification particularly challenging: concurrent program execution on parallel processors, and large, or even infinite, state spaces of data-manipulating programs.

In this thesis we propose a new approach to the verification of infinite-state concurrent programs. We call it causality-based, because it captures in an automatic proof system the "cause-effect" reasoning principles, which are often used informally in manual correctness proofs. Our method is based on a new concurrency model, called concurrent traces, which are the abstractions of the history of a concurrent program to some key events and the relationships between them.

We study the syntactic and language-based properties of concurrent traces, and characterize the complexity of such operations as emptiness checking and language inclusion. Regarding the program correctness, we develop proof systems for the broad classes of safety and liveness properties, and provide algorithms for the automatic construction of correctness proofs. We demonstrate that for practically relevant classes of programs, such as multi-threaded programs with binary semaphores, the constructed proofs are of polynomial size, and can be also checked in polynomial time.

Timo BOLKART
Dynamic and Groupwise Statistical Analysis of 3D Faces
(Advisor: Dr. Stefanie Wuhrer, now Grenoble)
Tues, 14.06.2016, 14:00h, building E1 4, room 019
This thesis proposes several methods to statistically analyze static and dynamic 3D face data. First, we present a fully-automatic method to robustly register entire facial motion sequences. The representation of the 3D facial motion sequences obtained by the registration allows us to perform statistical analysis of 3D face shapes in

motion. We then introduce a new localized multilinear model that is able to capture fine-scale details while being robust to noise and partial occlusions. To obtain a suitable registration for multilinearly distributed data, we introduce a groupwise correspondence optimization method that jointly optimizes a multilinear model and the registration of the 3D scans used for training. To robustly learn a multilinear model from 3D face databases with missing data, corrupt data, wrong semantic correspondence, and inaccurate vertex correspondence, we propose a robust model learning framework that jointly learns a multilinear model and fixes the data. Finally, we present one application of our registration methods, namely to obtain a sizing system that incorporates the shape of an identity along with its motion. We introduce a general framework to generate a sizing system for dynamic 3D motion data.

Hristo PENTCHEV
Sound Semantics of a High-Level Language with Interprocessor Interrupts
(Advisor: Prof. Wolfgang Paul)
Mon, 13.06.2016, 16:15h, building E1 7, room 0.01

Pervasive formal verification guarantees highest reliability of complex multi-core computer systems. This is required especially for safety critical applications in automotive, medical and military technologies. A crucial part of formal verification is the profound understanding of all system layers and the correct specification of their computational models and the interaction between software and hardware. The underlying architecture and the semantics of the higher-level programs cannot be considered in isolation. In particular, when the program execution relies on specific hardware features, these features have to be integrated into the computational model of the programing language.
In this thesis, we present an integration approach for interprocessor interrupts provided by multi-core architectures in the pervasive verification of system software written in C. We define an extension to the semantics of a high-level language, which considers interprocessor interrupts. We prove simulation between a multi-core hardware model and the high-level semantics with interrupts. In this simulation, we assume interrupts to occur on the boundary between statements. We justify that assumption by stating and proving an order reduction theorem, to reorder the interprocessor interrupt service routines to dedicated consistency points.

Nadezhda DONCHEVA
Network Biology Methods for Functional Characterization and Integrative Prioritization of Disease Genes and Proteins
(Advisor: Dr. Mario Albrecht)
Mon, 06.06.2016, 17:00h, building E1 5, room 029

Nowadays, large amounts of experimental data have been produced by high-through- put techniques, in order to provide more insight into complex phenotypes and cellular processes. The development of a variety of computational and, in particular, network-based approaches to analyze these data have already shed light on previously unknown mechanisms. However, we are still far from a comprehensive understanding of human diseases and their causes as well as appropriate preventive measures and successful therapies.
This thesis describes the development of methods and user-friendly software tools for the integrative analysis and interactive visualization of biological networks as well as their application to biomedical data for understanding diseases. We design an integrative phenotype-specific framework for prioritizing candidate disease genes and functionally characterizing similar phenotypes. It is applied to the identification of several disease-relevant genes and processes for inflammatory bowel diseases and primary sclerosing cholangitis as well as for Parkinson's disease.
Since finding the causative disease genes does often not suffice to understand diseases, we also concentrate on the molecular characterization of sequence mutations and their effect on protein structure and function. We develop a software suite to support the interactive, multi-layered visual analysis of molecular interaction mechanisms such as protein binding, allostery and drug resistance. To capture the dynamic nature of proteins, we also devise an approach to visualizing and analyzing ensembles of protein structures as, for example, generated by molecular dynamics simulations.

Christian RUß
Agentenbasierte marktliche Koordination von Wertschöpfungsnetzwerken
(Advisor: Prof. Jörg Siekmann)
Mon, 06.06.2016, 14:00h, building D3 2 (DFKI), Reuse meeting room

Die Arbeit stellt die Multiagenten-Supply-Chain-Simulationsumgebung MACSIMA vor, die das Design und die Simulation von großen Wertschöpfungsnetzwerken ermöglicht, die aus einer Vielzahl an autonomen Geschäftsagenten bestehen. MACSIMA stattet die Agenten, die miteinander über einen elektronischen Markt interagieren, mit einem adaptiven Verhandlungsmodul für bilaterale Verhandlungen aus, das mittels evolutionärer Algorithmen realisierte Lernfähigkeiten aufweist. Dadurch können Agenten ihre Verhandlungsstrategien an unterschiedliche Verhandlungspartner und sich dynamisch ändernde Marktbedingungen adaptieren. MACSIMA erlaubt es, die Lernfähigkeiten der Agenten sowie ihre Möglichkeiten zum Austausch und zur Nutzung von Informationen detailliert einzustellen und das sich daraus ergebende Systemverhalten zu simulieren und zu evaluieren. MACSIMA erweitert bekannte Ansätze um neue Konzepte, Methoden und Einstellmöglichkeiten. Simulationsergebnisse werden auf einer Makro-, Meso- und Mikroebene gemessen, mit Hilfe einer Evaluationsmethodik aufbereitet und schließlich im Hinblick auf die Systemperformanz und die Co-Evolution von Verhandlungsstrategien quantitativ und qualitativ analysiert. Die Evaluationsergebnisse zeigen, dass das Ergebnis der Selbstkoordination der Agenten stark von ihren Lern- und Informationsaustausch-Einstellungen abhängt und die Systemperformanz durch die Erweiterungen, die MACSIMA bereitstellt, im Vergleich zu bisherigen Ansätzen deutlich gesteigert werden kann.

Mai

Leonid PISHCHULIN
Articulated People Detection and Pose Estimation in Challenging Real World Environments

(Advisor: Prof. Bernt Schiele)
Tues, 31.05.2016, 16:15h, building E1 4, room 0.21

In this thesis we are interested in the problem of articulated people detection and pose estimation being key ingredients towards understanding visual scenes containing people. First, we investigate how statistical 3D human shape models from computer graphics can be leveraged to ease training data generation. Second, we develop expressive models for 2D single- and multi-person pose estimation. Third, we introduce a novel human pose estimation benchmark that makes a significant advance in terms of diversity and difficulty. Thorough experimental evaluation on standard benchmarks demonstrates significant improvements due to the proposed data augmentation techniques and novel body models, while detailed performance analysis of competing approaches on our novel benchmark allows to identify the most promising directions of improvement.

Cheng LI
Building Fast and Consistent (Geo-)Replicated Systems: from Principles to Practice
(Advisor: Prof. Rodrigo Rodrigues, now Lisbon)
Mon, 30.05.2016, 16:30h, building E1 5, room 029

Distributing data across replicas within a data center or across multiple data centers plays an important role in building Internet-scale services that provide a good user experience, namely low latency access and high throughput. This approach often compromises on strong consistency semantics, which helps maintain application-specific desired properties, namely, state convergence and invariant preservation. To relieve such inherent tension, in the past few years, many proposals have been designed to allow programmers to selectively weaken consistency levels of certain operations to avoid costly immediate coordination for concurrent user requests. However, these fail to provide principles to guide programmers to make a correct decision of assigning consistency levels to various operations so that good performance is extracted while the system behavior still complies with its specification.
The primary goal of this thesis work is to provide programmers with principles and tools for building fast and consistent (geo-) replicated systems by allowing programmers to think about various consistency levels in the same framework. The first step we took was to propose RedBlue consistency, which presents sufficient conditions that allow programmers to safely separate weakly consistent operations from strongly consistent ones in a coarse-grained manner. Second, to improve the practicality of RedBlue consistency, we built SIEVE - a tool that explores both Commutative Replicated Data Types and program analysis techniques to assign proper consistency levels to different operations and to maximize the weakly consistent operation space. Finally, we generalized the tradeoff between consistency and performance and proposed Partial Order-Restrictions consistency (or short, PoR consistency) - a generic consistency definition that captures various consistency levels in terms of visibility restrictions among pairs of operations and allows programmers to tune the restrictions to obtain a fine-grained control of their targeted consistency semantics.

Geng CHEN
Store Buffer Reduction Theorem and Application
(Advisor: Prof. Wolfgang Paul)
Wed, 11.05.2016, 14:00h, building E1 7, room 0.01

The functional correctness of multicore systems can be shown through pervasive formal verification, which proves the simulation between the system software computation and the corresponding hardware computation. In the implementation of the system software, the sequential consistency (SC) of memory is usually assumed by the system programmers. However, most modern processors (x86, Sparc) provide the total store order (TSO) memory model for greater efficiency. A store buffer reduction theorem was presented by Cohen and Schirmer to bridge the gap between the SC and the TSO. Nevertheless, the theorem is not applicable to programs hat edit their own page tables. The reason is that the MMU can be treated neither as a part of the program thread nor as a separate thread. This thesis contributes to generalize the Cohen-Schirmer reduction theorem by adding the MMUs.
As the first contribution of this thesis, we present a programming discipline which guarantees sequential consistency for the TSO machine with MMUs. Under this programming discipline, we prove the store buffer reduction theorem with MMUs. For the second contribution of this thesis, we apply the theorem to the ISA level and the C level. By proving a series of simulation theorems, we apply our store buffer reduction theorem with MMU to the ISA named MIPS-86. After that, we introduce the multicore compiler correctness theorem to map the programming discipline to the parallel C level.

Sabine JANZEN
Mixed Motive Dialogues in Equilibrium: Kooperative Planung von Antworten in Dialogen mit kongruenten und inkongruenten Teilnehmermotiven
(Advisor: Prof. Wolfgang Wahlster)
Wed, 04.05.2016, 10:00h, building D3 1 (DFKI), room Reuse

A mixture of congruent as well as incongruent, partially conflictive motives characterizes cooperative mixed motive dialogues, e.g., sales conversations, where interlocutors make concessions to establish a compromise between selfishness and fair play. So far, no dialogue system is available that supports dialogues of this type. Despite of the overall presence of mixed motive dialogues in everyday life, little attention has been given to this topic in text planning in contrast to scrutinized collaborative and non-collaborative dialogues. In this thesis, a model is introduced that formalizes answer planning as psychological game combined with text planning approaches as well as explicit motive representations for generating dialogues perceived as fair by all interlocutors. Cooperative mixed motive dialogues will be captured theoretically and analyzed by empirical studies before modeling and formalizing results that will be evaluated by means of a prototypical dialogue system. For solving the conflict in mixed motives, a game theoretical equilibrium approach is applied to simulate human cooperative behavior in mixed motive dialogues. Based on established interdisciplinary approaches, this thesis represents an initiative

contribution for planning little investigated mixed motive interactions supported by dialogue systems.

April

Pascal Tobias PETER
Understanding and Advancing PDE-based Image Compression
(Advisor: Prof. Joachim Weickert)
Fri, 29.04.2016, 14:15h, building E1 1, room 4.07

This thesis is dedicated to the emerging field of image compression with partial differential equations (PDEs). PDE-based codecs store only a small amount of image points and propagate their information into the unknown image areas during the decompression step. It has already been shown that PDE-based compression can outperform the current quasi-standard, JPEG2000, for certain classes of images. However, the reasons for this success are not yet fully understood, and PDE-based compression is still in a proof-of-concept stage. We contribute to a deeper insight into design principles for PDE-based codecs by
proposing a probabilistic justification for anisotropic diffusion. Moreover, we analyse the interaction between efficient storage methods and image reconstruction with diffusion, which allows us to rank different PDEs according to their practical value in compression. Based on these new observations, we advance PDE-based compression towards practical viability: First, we present a new hybrid codec that combines PDE- and patch-based interpolation to deal with highly textured images. Furthermore, a new video player demonstrates the real-time capacities of PDE-based image interpolation. In addition, we introduce perceptive coding: A region of interest coding algorithm allows to specify important image areas that are represented with higher accuracy.
Finally, we propose a new framework for diffusion-based image colourisation that allows us to build an efficient codec for colour images. Experiments on natural image databases show that our new method is competitive to the current state-of-the-art codecs on real world data.

Bimal VISWANATH
Towards Trustworthy Social Computing Systems
(Advisor: Dr. Krishna Gummadi)
Thu, 28.04.2016, 9:00h, building E1 5, room 002

The rising popularity of social computing systems has managed to attract rampant forms of service abuse that negatively affects the sustainability of these systems and degrades the quality of service experienced by their users. The main factor that enables service abuse is the weak identity infrastructure used by most sites, where identities are easy to create with no verification by a trusted authority. Attackers are exploiting this infrastructure to launch Sybil attacks, where they create multiple fake (Sybil) identities to take advantage of the combined privileges associated with the identities to abuse the system. In this thesis, we present techniques to mitigate service abuse by designing and building defense schemes that are robust and practical. We use two broad defense strategies: (1) Leveraging the social network: We first analyze existing social network-based Sybil detection schemes and present their practical limitations when applied on real world social networks. Next, we present an approach called Sybil Tolerance that bounds the impact an attacker can gain from using multiple identities; (2) Leveraging activity history of identities: We present two approaches, one that applies anomaly detection on user social behavior to detect individual misbehaving identities, and a second approach called Stamper that focuses on detecting a group of Sybil identities. We show that both approaches in this category raise the bar for defense against adaptive attackers.

Jörg SCHAD
Understanding and Managing the Performance Variation and Data Growth in Cloud Computing
(Advisor: Prof. Jens Dittrich)
Fri, 15.04.2016, 16:00h, building E1 1, room 3.06

The topics of Cloud Computing and Big Data Analytics dominate today's IT landscape. This dissertation considers the combination of both and the challenges which result. In particular, it addresses executing data intensive jobs efficiently on public cloud infrastructure, with respect to response time, cost, and reproducibility. We present an extensive study of performance variance in public cloud infrastructures covering various dimensions including micro-benchmarks and application-benchmarks, different points in time, and different cloud vendors. Next, this dissertation addresses the challenge of efficiently processing dynamic datasets which change over time.
This dissertation presents a framework to deal efficiently with dynamic datasets inside MapReduce using different techniques depending on the characteristics of the dataset. This dissertation concludes with a discussion on how new technologies such as container virtualization will affect the challenges presented here.

Mohamed YAHYA
Question Answering and Query Processing for Extended Knowledge Graphs
(Advisor: Prof. Gerhard Weikum)
Fri, 15.04.2016, 10:00h, building E1 4, room 0.24

Knowledge graphs have seen wide adoption, in large part owing to their schemaless nature that enables them to grow seamlessly, allowing for new relationships and entities as needed. With this rapid growth, several issues arise:
(i) how to allow users to query knowledge graphs in an expressive, yet user-friendly, manner, which shields them from all the underlying complexity,
(ii) how, given a structured query, to return satisfactory answers to the user despite possible mismatches between the query vocabulary and structure and the knowledge graph, and
(iii) how to automatically acquire new knowledge, which can be fed into a knowledge graph. In this dissertation, we make the following contributions to address the above issues:
- We present DEANNA, a framework for question answering over knowledge graphs, allowing users to easily express complex information needs using natural language

and obtain tuples of entities as answers thereby taking advantage of the structure in the knowledge graph.
- We introduce TriniT, a framework that compensates for unsatisfactory results for structured queries over knowledge graphs either due to mismatches with the knowledge graph or the knowledge graph's inevitable incompleteness. TriniT tackles the two issues by extending the knowledge graph using information extraction over textual corpora, and supporting query relaxation where a user's query is rewritten in a manner transparent to the user to compensate for any mismatches with the data.
- We present ReNoun, an open information extraction framework for extracting binary relations mediated by noun phrases and their instances from text. Our scheme extends the state of the art in open information extraction which has thus far focused on relations mediated by verbs.
Our experimental evaluations of each of the above contributions demonstrate the effectiveness of our methods in comparison to state-of-the-art approaches.

Christina TEFLIOUDI
Algorithms for Shared-Memory Matrix Completion and Maximum Inner Product Search
(Advisor: Prof. Rainer Gemulla, now Uni Mannheim)
Thu, 14.04.2016, 16:00h, building E1 4, room 0.24

In this thesis, we propose efficient and scalable algorithms for shared-memory matrix factorization and maximum inner product search. Matrix factorization is a popular tool in the data mining community due to its ability to quantify the interactions between different types of entities. It typically maps the (potentially noisy) original representations of the entities into a lower dimensional space, where the "true" structure of the dataset is revealed. Inner products of the new vector representations are then used to measure the interactions between different entities. The strongest of these interactions are usually of particular interest in applications and can be retrieved by solving a maximum inner product search problem.
For large real-life problem instances of matrix factorization and maximum inner product search, efficient and scalable methods are necessary. We first study large-scale matrix factorization in a shared-memory setting and we propose a cache-aware, parallel method that avoids fine-grained synchronization or locking. In more detail, our approach partitions the initial large problem into small, cache-fitting sub-problems that can be solved independently using stochastic gradient descent. Due to the low cache-miss rate and the absence of any locking or synchronization, our method achieves superior performance in terms of speed (up to 60% faster) and scalability than previously proposed techniques.
We then proceed with investigating the problem of maximum inner product search and design a cache-friendly framework that allows for both exact and approximate search. Our approach reduces the original maximum inner product search problem into a set of smaller cosine similarity search problems that can be solved using existing cosine similarity search techniques or our novel algorithms tailored for use within our framework. Experimental results show that our approach is multiple orders of magnitude faster than naive search, consistently faster than alternative methods for exact search, and achieves better quality-speedup tradeoff (up to 3.9x faster for similar recall levels) than state-of-the-art approximate techniques.

Ruslan AKULENKO
Data Mining Techniques for improving and enriching cancer epigenetics
(Advisor: Prof. Volkhard Helms)
Fri, 01.04.2016, 14:00h, building E2 1, room 001

The aim of epigenetic cancer research is to connect tumor development with processes involving gene regulations, which are not directly encoded in the DNA. Recent advanced techniques for genome-wide mapping of epigenetic information are significantly improving this knowledge by generating large amounts of high resolution data. But still many dependencies remain unknown due to the presence of technical errors such as batch effect. This thesis describes the associations discovered during DNA methylation data mining as well as technical challenges which should be considered during data processing. Moreover, the same computational methods applied to epigenetic data were adapted to process bacterial genome data.

March

Christian DOZCKAL
A Machine-Checked Constructive Metatheory of Computation Tree Logic
(Advisor: Prof. Gert Smolka)
Mon, 21.03.2016, 11:00h, building E1 7, room 0.01

This thesis presents a machine-checked constructive metatheory of computation tree logic (CTL) and its sublogics K and K* based on results from the literature. We consider models, Hilbert systems, and history-based Gentzen systems and show that for every logic and every formula s the following statements are decidable and equivalent: s is true in all models, s is provable in the Hilbert system, and s is provable in the Gentzen system. We base our proofs on pruning systems constructing finite models for satisfiable formulas and abstract refutations for unsatisfiable formulas. The pruning systems are devised such that abstract refutations can be translated to derivations in the Hilbert system and the Gentzen system, thus establishing completeness of both systems with a single model construction. All results of this thesis are formalized and machine-checked with the Coq interactive theorem prover. Given the level of detail involved and the informal presentation in much of the original work, the gap between the original paper proofs and constructive machine-checkable proofs is considerable. The mathematical proofs presented in this thesis provide for elegant formalizations and often differ significantly from the proofs in the literature.

Kim PECINA
Trustworthy and Privacy-Preserving Processing of Personal Information - Cryptographic Protocols, Constructions, and Tools
(Advisor: Prof. Matteo Maffei)
Thu, 10.03.2016, 11:00h, building E1 7, room 0.01

Internet services collect lots of information that users deem highly sensitive such as the browsing history and email addresses, often without users noticing this conduct. The collected information is used for personalizing services and it is monetized, e.g., in the form of targeted advertisements. Realizing modern web services that maintain their functionality and satisfy the seemingly conflicting properties of trustworthiness and privacy is challenging: in a social network, how to enforce that only friends can post comments, if users are unwilling to reveal their identity? in online behavioral advertising, how to serve personalized ads, if users insist on keeping their interests private?

In this thesis, we propose techniques for the trustworthy and privacy-preserving processing of personal information. First, we present an API for the trustworthy and privacy-preserving release of personal information. The API enables the declarative specification of distributed systems that satisfy sophisticated security properties, including authorization, anonymity, and accountability. We use this API to implement an anonymous evaluation system, anonymous webs of trust, and a secure social network. Second, we present a methodology for the trustworthy and privacy-preserving retrieval of information. We exemplify our approach by presenting ObliviAd, an architecture for online behavioral advertising that provably protects user profiles and delivers highly-personalized advertisements.

February

Krzysztof TEMPLIN
Depth, Shading, and Stylization in Stereoscopic Cinematography
(Advisor: Prof. Hans-Peter Seidel)
Fri, 26.02.2016, 9:00h, building E1 4, room 0.19

Due to the constantly increasing focus of the entertainment industry on stereoscopic imaging, techniques and tools that enable precise control over the depth impression and help to overcome limitations of the current stereoscopic hardware are gaining in importance. In this dissertation, we address selected problems encountered during stereoscopic content production, with a particular focus on stereoscopic cinema. First, we consider abrupt changes of depth, such as those induced by cuts in films. We derive a model predicting the time the visual system needs to adapt to such changes and propose how to employ this model for film cut optimization. Second, we tackle the issue of discrepancies between the two views of a stereoscopic image due to view-dependent shading of glossy materials. The suggested solution eliminates discomfort caused by non-matching specular highlights while preserving the perception of gloss. Last, we deal with the problem of film grain management in stereoscopic productions and propose a new method for film grain application that reconciles visual comfort with the idea of medium-scene separation.

Oliver KLEHM
User-Guided Scene Stylization using Efficient Rendering Techniques
(Advisor: Prof. Hans-Peter Seidel)
Fri, 12.02.2016, 16:00h, building E1 4, room 0.19

In this dissertation, we propose new techniques to display and manipulate virtual worlds to support visual design.The real-time constraint of applications such as games limits the accuracy at which rendering algorithms can simulate light transport. Our first method focuses on the efficient rendering of surfaces under natural illumination, extending previous work that ignores directional-dependent effects in the lighting. In a second work, we present an approach for the efficient computation of scattering in homogeneous participating media. The main challenge is the accumulation of visibility along view rays, which we solve using an efficient filtering scheme. In the second part of the dissertation, we investigate methods that provide artists with approaches to stylize and manipulate the appearance of volumetric scattering. First, we focus on the effect of light shafts that one can typically observe on hazy days due to openings in the clouds. While the effect is often used in games and movies, it is difficult to manipulate. We propose tools to directly manipulate parameters of the rendering, effectively providing control over the creation, shape, and color of these light shafts. In another work, we abstract from direct parameter changes and propose a goal-based design approach to manipulate the appearance of heterogeneous media such as clouds. We use inverse rendering to infer volume parameters from user paintings to achieve a desired look. The problem is expressed as an optimization procedure for which we show an efficient execution on the GPU. We show in several examples that these novel methods enable intuitive, expressive, and effective control over the stylization of volumetric scattering.

Oskar ELEK
Efficient Methods for Physically-based Rendering of Participating Media
(Advisor: Dr. Tobias Ritschel, now University College London)
Tues, 02.02.2016, 11:00h, building E1 4, room 0.19

This thesis proposes several novel methods for realistic synthesis of images containing participating media. This is a challenging problem, due to the multitude and complexity of ways how light interacts with participating media, but also an important one, since such media are ubiquitous in our environment and therefore are one of the main constituents of its appearance. The main paradigm we follow is designing efficient methods that provide their user with an interactive feedback, but are still physically plausible.

The presented contributions have varying degrees of specialisation and, in a loose connection to that, their resulting efficiency. First, the screen-space scattering algorithm simulates scattering in homogeneous media, such as fog and water, as a fast image filtering process. Next, the amortised photon mapping method focuses on rendering clouds as arguably one of the most difficult media due to their high scattering anisotropy. Here, interactivity is achieved through adapting to certain conditions specific to clouds. A generalisation of this approach is principal-ordinates propagation, which tackles a much wider class of heterogeneous media. The resulting method can handle almost arbitrary optical properties in such media, thanks to a custom finite-element propagation scheme. Finally, spectral ray differentials aim at an efficient reconstruction of chromatic dispersion phenomena, which occur in transparent media such as water, glass and gemstones. This method is based on analytical ray differentiation and as such can be incorporated to any ray-based rendering framework, increasing the efficiency of reproducing dispersion by

about an order of magnitude. All four proposed methods achieve efficiency primarily by utilizing high-level mathematical abstractions, building on the understanding of the underlying physical principles that guide light transport. The methods have also been designed around simple data structures, allowing high execution parallelism and removing the need to rely on any sort of preprocessing. Thanks to these properties, the presented work is not only suitable for interactively computing light transport in participating media, but also allows dynamic changes to the simulated environment, all while maintaining high levels of visual realism.

January

Markus RABE
A Temporal Logic Approach to Information-flow Control
(Advisor: Prof. Bernd Finkbeiner)
Thu, 28.01.2016, 14:00h, building E1 7, room 0.01

Information leaks and other violations of information security pose a severe threat to individuals, companies, and even countries. The mechanisms by which attackers threaten information security are extremely diverse and to show their absence thus proved to be a challenging problem. Information-flow control is a principled approach to prevent security incidents in programs and other technical systems. In information-flow control we define information-flow properties, which are sufficient conditions for when the system is secure in a particular attack scenario.
In this thesis, we propose a temporal logic approach to information-flow control to provide a simple formal basis for the specification and enforcement of information-flow properties. We extend temporal logics to enable the expression of information-flow properties and study the resulting verification problems. The approach can be used to enforce a wide range of previously known and also new information-flow properties with a single verification algorithm. We also shed light on the relation of information-flow properties to the linear-time branching-time spectrum of properties.

Robert NESSELRATH
SiAM-dp: An open development platform for massively multimodal dialogue systems in cyber-physical environments
(Advisor: Prof. Wolfgang Wahlster)
Wed, 27.01.2016, 16:00h, building D3 2, room -2.17 (Reuse)

Cyber-physical Environments (CPEs) enhance natural environments of daily life such as homes, factories, oces, and cars by connecting the cybernetic world of computers and communication with the real physical world. While under the keyword of Industrie 4.0, CPEs will take a relevant role in the next industrial revolution, and they will also appear in homes, oces, workshops, and numerous other areas. In this new world, classical interaction concepts where users exclusively interact with a single stationary device, PC or smartphone become less dominant and make room for new occurrences of interaction between humans and the environment itself. Furthermore, new technologies and a rising spectrum of applicable modalities broaden the possibilities for interaction designers to include more natural and intuitive non-verbal and verbal communication. The dynamic characteristic of a CPE and the mobility of users confronts developers with the challenge of developing systems that are flexible concerning the connected and used devices and modalities. This implies new opportunities for cross-modal interaction that go beyond dual modalities interaction as is well known nowadays.
This thesis addresses the support of application developers with a platform for the declarative and model based development of multimodal dialogue applications, with a focus on distributed input and output devices in CPEs.

Christian ENGELS
Why are certain polynomials hard?
(Advisor: Prof. Markus Bläser)
Wed, 27.01.2016, 12:15h, building E1 7, room 0.01

In this thesis we will try to answer the question why specific polynomials have no small suspected arithmetic circuits. We will look at this general problem in three different ways. First, we study non-commutative computation. Here we show matching upper and lower bounds for the non-commutative permanent for various restricted graph classes. Our main result gives algebraic branching program upper and lower bounds for graphs with connected component size 6 as well as a #P hardness result. We introduce a measure that characterizes this complexity on these instances. Secondly, we introduce a new framework for arithmetic circuits, similar to fixed parameter tractability in the boolean setting. This framework shows that specific polynomials based on graph problems have the expected complexity as in the counting FPT case. We introduce classes $\bvw\{t\}$ which are close to the boolean setting but hardly use the power of arithmetic circuits. We then introduce $\vw\{t\}$ with modified problems to remedy this situation. Thirdly, we study polynomials defined by graph homomorphisms and show various dichotomy theorems. This shows that even restrictions on the graphs can already give us hard instances. Finally, we stray from our main continuous thread and handle simple heuristics for metric graphs. Instead of studying specific metrics we look at a randomized process giving us shortest path metric instances.

Luciano DEL CORRO
Methods for Open Information Extraction and Sense Disambiguation on Natural Language Text
(Advisor: Prof. Rainer Gemulla, now Uni Mannheim)
Mon, 11.01.2016, 16:15h, building E1 5, room 0.29

Natural language text has been the main and most comprehensive way of expressing and storing knowledge. A long standing goal in computer science is to develop systems that automatically understand textual data, making this knowledge accessible to computers and humans alike. We conceive automatic text understanding as a bottom-up approach, in which a series of interleaved tasks build upon each other. Each task achieves more understanding over the text than the previous one. In this regard, we present three methods that aim to contribute to the primary stages of this setting. Our first contribution, ClausIE, is an open information extraction method intended to recognize textual expressions of potential facts in text

(e.g. "Dante wrote the Divine Comedy") and represent them with an amenable structure for computers [("Dante", "wrote", "the Divine Comedy")]. Unlike previous approaches, ClausIE separates the recognition of the information from its representation, a process that understands the former as universal (i.e., domain-independent) and the later as application-dependent. ClausIE is a principled method that relies on properties of the English language and thereby avoids the use of manually or automatically generated training data. Once the information in text has been correctly identified, probably the most important element in a structured fact is the relation which links its arguments, a relation whose main component is usually a verbal phrase. Our second contribution, Werdy, is a word entry recognition and disambiguation method. It aims to recognize words or multi-word expressions (e.g., "Divine Comedy" is a multi-word expression) in a fact and disambiguate verbs (e.g., what does "write" mean?). Werdy is also an unsupervised approach, mainly relying on the semantic relation established between a verb sense and its arguments. The other key components in a structured fact are the named entities (e.g., "Dante") that often appear in the arguments. FINET, our last contribution, is a named entity typing method. It aims to understand the types or classes of those names entities (e.g., "Dante" refers to a writer). FINET is focused on typing named entities in short inputs (like facts). Unlike previous systems, it is designed to find the types that match the entity mention context (e.g., the fact in which it appears). It uses the most comprehensive type system of any entity typing method to date with more than 16k classes for persons, organizations and locations. These contributions are intended to constitute constructive building blocks for deeper understanding tasks in a bottom-up automatic text understanding setting.