Yearbook 2009/2010




International Max Planck Research School for Computer Science

Organized by the Max Planck Institutes for Informatics and Software Systems

In cooperation with the Computer Science Department at Saarland University

International Max Planck Research School
for Computer Science

Yearbook 2009/2010

CONTENTS

IMPRS-CS PORTRAIT

Educating and training junior scientists is of primary importance for the future of science, research, and innovation. The Max Planck Society, in cooperation with German universities, has launched the International Max Planck Research Schools (IMPRS) initiative, offering especially gifted students from Germany and abroad the opportunity to earn a doctorate within a structured program and providing excellent conditions for education and research.

The IMPRS for Computer Science (IMPRS-CS) is highly research-oriented. Admitted students have the unique opportunity to learn from, and work closely with, world-renowned researchers in their respective fields of specialization.

Our main goal at the IMPRS-CS is to attract international students to Germany for their Master's or PhD studies, to acquaint them with research institutions in this country, and thus arouse their interest in working at, or in cooperation with, German research institutions in the future. More than 50 % of our PhD students are from abroad, for example from Bulgaria, China, India, and Romania. All funded students in the Master's program are from abroad.

The International Max Planck Research School for Computer Science (IMPRS-CS) offers various programs in cooperation with Saarland University and its Saarbrücken Graduate School of Computer Science:

The PhD program leads to a doctoral degree and is open to students who already hold a research-oriented Master's degree in Computer Science (or an equivalent degree).

The PhD qualification program offers students, who hold a four-year Bachelor's degree in computer science with excellent academic performance and a strong research background or research interest, the possibility to continue directly with their PhD. Admitted students undergo a preparatory phase with course work before starting the research and dissertation phase.

The Master's program is open to students who have a Bachelor's degree in computer science with very good grades and want to pursue a Master's degree in computer science or visual computing.

All IMPRS-CS programs are organized in collaboration with the MPI for Informatics, the MPI for Software Systems, and the Department of Computer Science at Saarland University, all three of which rank among the very best scientific research institutions in computer science in Europe.

The projects are supervised by researchers of the Max Planck Institutes and their colleagues in the Computer Science Department.

All students admitted to the IMPRS-CS receive a scholarship that covers all living expenses of the student and the student's family. Furthermore, we assist our students in finding accommodation, support them in all kinds of administrative issues, offer German and English classes, and organize joint activities.

Alumni: M.Sc.

Fidaa ABED

NATIONALITY: Palestinian
DEPARTMENT: Algorithms and Complexity

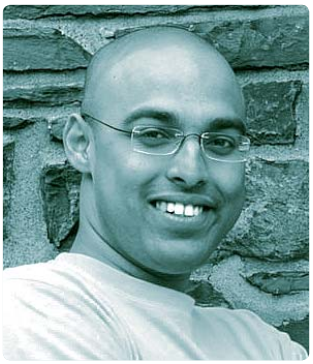
THESIS TITLE:

Coordination Mechanisms for Unrelated Machine Scheduling

ABSTRACT OF MASTER'S THESIS:

We investigate load balancing games in the context of unrelated machines. In such a game, there are a number of jobs and a number of machines, and each job needs to be scheduled on one machine. A collection of values p_{ij} are given, where p_{ij} indicates the processing time of job i on machine j. Moreover, each job is controlled by a selfish player who only wants to minimize the completion time of his job while disregarding other players' welfare. The outcome schedule is a *Nash equilibrium* if no player can unilaterally change his machine and reduce the completion of his job. It can be expected that in an equilibrium, the performance of the system can be far from optimal. The degradation of the system performance in Nash equilibria is defined as the *price of anarchy* (PoA): the ratio of the cost of the worst *Nash Equilibrium* to the cost of the optimal scheduling. Clever scheduling policies can be designed to reduce PoA. These scheduling policies are called *coordination mechanisms*.

It has been posed as an open question "what is the best possible lower bound when *coordination mechanisms* use preemption". In this thesis we consider three cases: when the jobs have IDs, when *coordination mechanisms* order the jobs randomly, and when the jobs have no IDs. We prove a tight lower bound for the first two cases and we show our progress in the third case.



Avishek ANAND

NATIONALITY: Indian

DEPARTMENT: Databases and Information Systems

THESIS TITLE:

Efficient Indexing and Query Processing for Search-Enabled Web Archives

ABSTRACT OF MASTER'S THESIS:

The World Wide Web has become a key source of knowledge pertaining to almost every walk of life. Unfortunately, much of data on the Web is highly ephemeral in nature, with more than 50 – 80 % of content change within a short time, leading to loss of potentially valuable parts of the Web. Continuing the pioneering efforts of many national (digital) libraries, organizations like the International Internet Preservation Consortium (IIPC), the Internet Archive (IA) and the European Archive (EA) have been tirelessly working towards preserving the ever changing Web.

While most of these efforts have paid significant interest towards long term preservation of Web data they pay lesser emphasis on acquisition of data, and performing historical analysis over the collected data. Given the dynamic nature of the Web, it is essential to effectively and efficiently capture near-complete version history of popular and important pages – not just their snapshots in time as most standard crawlers do. Next, in order to realize the full potential of a Web archive for performing elaborate historical analysis, we believe that advanced query methods such as time-travel queries are necessary.

Laura Maria ANDREESCU

NATIONALITY: **Romanian**
DEPARTMENT: **Databases and Information Systems**

THESIS TITLE:

Pricing Information Goods in an Agent-based Information Filtering System

ABSTRACT OF MASTER'S THESIS:

Most approaches to information filtering taken so far have the underlying hypothesis of potentially delivering notifications from every information producer to subscribers. This exact information filtering or publish/subscribe model creates an efficiency and scalability bottleneck that in many applications might not be desirable or even considered disruptive.

In this thesis we present ABIS, an Agent-Based Information Retrieval System that will prevent the information overhead and scalability bottlenecks. In ABIS, the user subscribes to and monitors only carefully selected data sources, and receives interesting events from these sources only. In this way scalability is enhanced by trading recall for lower overhead of message exchange. In this thesis we define the rules of an agent-based architecture, especially designed for approximate information filtering, that regulates the interactions between agents. We introduce a new agent selection mechanism based on resource selection, predicted publishing behavior, and cost of information to improve data source selection. We feel that this new mechanism offers a number of interesting opportunities for further research.

Aleksandr ANDREYCHENKO

NATIONALITY: Russian
DEPARTMENT: Analysis of Markovian Models

THESIS TITLE:

Uniformization for Time-Inhomogeneous Markov Population Models

ABSTRACT OF MASTER'S THESIS:

Time is one of the main factors in any kind of real-life systems. When a certain system is analysed one is often interested in its evolution with respect to time. Various phenomena can be described using a form of time-dependency. The difference between load in call-centres is the example of time-dependency in queueing systems. The process of migration of biological species in autumn and spring is another illustration of changing the behaviour in time. The ageing process in critical infrastructures (which can result in the system component failure) can also be considered as another type of time-dependent evolution.

Considering the variability in time for chemical and biological systems one comes to the general tasks of *systems biology* [9]. It is an inter-disciplinary study field which investigates complex interactions between components of biological systems and aims to explore the fundamental laws and new features of them. Systems biology is also used for referring to a certain type of research cycle. It starts with the creation a model. One tries to describe the behaviour in a most intuitive and informative way which assumes convenience and visibility of future analysis. The traditional approach is based on *deterministic* models where the evolution can be predicted with certainty. This type of model usually operates at a macroscopic scale and if one considers chemical reactions the state of the system is represented by the concentrations of species and a continuous deterministic change is assumed. A set of ordinary differential equations (ODE) is one of the ways to describe such kind of models. To obtain a solution numerical methods are applied. The choice of a certain ODE-solver depends on the type of the ODE system.

Another option is a full description of the chemical reaction system where we model each single molecule explicitly operating with their properties and positions in space. Naturally it is difficult to treat big systems in a such way and it also creates restrictions for computational analysis.

However it reveals that the deterministic formalism is not always sufficient to describe all possible ways for the system to evolve. For instance, the Lambda phage decision circuit [1] can be a motivational example of such system. When the lambda phage virus infects the E. coli bacterium it can evolve in two different ways. The first one is lysogeny where the genome of the virus is integrated into the genome of the bacterium. Virus DNA is then replicated in descendant cells using the replication mechanism of the host cell. Another way is entering the lytic cycle, which means that new phages are synthesized directly in the host cell and finally its membrane is destroyed and new phages are released. A deterministic model is not appropriate to describe this process of choosing between two pathways as this decision is *probabilistic* and one needs a *stochastic* model to give an appropriate description.

Another important issue which has to be addressed is the fact that the state of the system changes *discretely*. It means that one considers not the continuous change of chemical species concentrations but discrete events occuring with different probabilities (they can be time-dependent as well).

We will use the continuous-time Markov Population Models (MPMs) formalism in this thesis to describe discrete-state stochastic systems. They are indeed continuous-time Markov processes, where the state of the system represents populations and it is expressed by the vector of natural numbers. Such systems can have *infinitely* many states. For the case of chemical reactions network it results in the fact that one cannot provide strict upper bounds for the population of certain species. When analyzing these systems one can estimate measures of interest (like expectation and variance for the certain species populations at a given time instant). Besides this, probabilities for certain events to occur can be important (for instance, the probability for population to reach the threshold or the probability for given species to extinct).

The usual way to investigate properties of these systems is *simulation* [8] which means that a large amount of possible sample trajectories are generated and then analysed. However it can be difficult to collect a sufficient number of trajectories to provide statistical estimations of good quality. Besides the simulation, approaches based on the uniformization technique have been proven to be computationally efficient for analysis of time-independent MPMs. In the case of time-dependent processes only few results concerning the performance of numerical techniques are known [2].

Here we present a method for conducting an analysis of MPMs that can have possibly *infinitely* many states and their dynamics is *time-dependent*. To cope with the problem we combine the ideas of on-the-y uniformization [5] with the method for treating time in homogeneous behaviour presented by Bucholz.

Sergiy BYELOZYOROV

NATIONALITY: **Ukrainian**
DEPARTMENT: **Computer Graphics**

THESIS TITLE:

Construction of Virtual Worlds with Web 2.0 Technology

ABSTRACT OF MASTER'S THESIS:

Current Web technologies allow developers to create rich Web-applications. Unlike desktop applications Web 2.0 programs are created by easily linking several existing components. This approach, also known as mashup, allows to use JavaScript to connect web-services and browser components together.

I have extended this development method by bringing 3D and virtual world networking components into the browser. This allowed me to create Virtual Worlds Web-application similar to Second Life. I have wrapped the open-source Sirikata platform for virtual worlds into a Web-service component, created XML3D rendering component, combined them with other browser services and thus created a fully-featured 3D world application right inside of the browser.

Piotr DANILEWSKI

NATIONALITY: Polish
DEPARTMENT: Computer Graphics

THESIS TITLE:

Binned Kd-tree Construction with SAH on the GPU

ABSTRACT OF MASTER'S THESIS:

Our main goal is to create realistically looking animation in real-time. To that end, we are interested in fast ray tracing. Ray tracing recursively traces photon movement from the camera (backward) or light sources (forward). To find where the first intersection between a given ray and the objects in the scene is we use acceleration structures, for example kd-trees. Kd-trees are considered to perform best in the majority of cases, however due to their large construction times are often avoided for dynamic scenes. In this work we try to overcome this obstacle by building the kd-tree in parallel on many cores of a GPU.

Our algorithm builds the kd-tree in a top-down breath-first fashion, with many threads processing each node of the tree. For each node we test 31 uniformly distributed candidate split planes along each axis and use the Surface Area cost function to estimate the best one. In order to reach maximum performance, the kd-tree construction is divided into 4 stages. Each of them handles tree nodes of different primitive count, differs in how counting is resolved and how work is distributed on the GPU.

Our current program constructs kd-trees faster than other GPU implementations, while maintaining competing quality compared to serial CPU programs. Tests have shown that execution time scales well in respect to power of the GPU and it will most likely continue doing so with future releases of the hardware.

Luis Javier DE LA GARZA TREVIÑO

NATIONALITY: **Mexican**

DEPARTMENT: **Databases and Information Systems**

THESIS TITLE:

Implementation and Evaluation of an Efficient, Distributed Replication Algorithm in a Real Network

ABSTRACT OF MASTER'S THESIS:

The main challenge for p2p applications in unstructured networks is how to perform data replication. Without data replication, the performance of the network depends solely on the availability of items (which in turn depends on the availability of peers). With data replication, it is possible to make copies of items and distribute them across the unstructured network, in order to improve the availability of items, increasing the performance of the network. Obviously, generating replicas of all items would be not only inefficient but also not feasible, since in real data networks, the number of items in the network is much larger than the number of peers connected, posing storage difficulties. An optimal replication algorithm generates the right amount of replicas of the right items. To show the advantages of data replication an implementation and evaluation of a data replication algorithm (the P2R2 Distributed Algorithm [6]) will be presented.

Minko Todorov DUDEV

NATIONALITY: **Bulgarian**
DEPARTMENT: **Databases and Information Systems**

THESIS TITLE:

Personalization of Search on Structured Data

ABSTRACT OF MASTER'S THESIS:

Recent research on semantic search methods has given rise to a number of large graph-based knowledge repositories which could form the basis of semantic search engines, fully capable of finding and ranking facts. Similar to traditional Web search, the ranking provided may prove to be satisfactory on average, however just as in other types of knowledge bases, personalization could be highly beneficial in ensuring that the result ordering is truly relevant to the specific user. While a significant amount of work has been done in the areas of personalization on the Web, XML search and databases, this thesis for the first time addresses the issue of personalizing query results in the specific setting of graph-based knowledge bases.

The two main issues in personalization are explored in detail: i) the construction of a user profile including a spreading activation methodology and a relation similarity measure for the inference of user interests and ii) the formal description of a model for personalized scoring based on that profile. We present techniques that fully exploit the properties of graph-based knowledge repositories and validate the proposed approach with initial experimental results.

Qi GAO

NATIONALITY: Chinese
DEPARTMENT: Mathematical Image Analysis

THESIS TITLE:

Low Bit Rate Video Compression Using Inpainting PDEs and Optic Flow

ABSTRACT OF MASTER'S THESIS:

Video compression can be done by a variety of methods. In this thesis we propose a motion compensated video compression scheme aimed at low bit rate applications, which employs the optic flow and the partial differential equation (PDE) based image compression techniques. We concentrate mainly on the problem of coding the optic flow data as well as coding the displaced frame difference (DFD). The dense optic flow field is estimated by the method of Brox et al. The PDE based image compression method, which encodes an image by B-tree triangular coding and regards the decoding as an inpainting problem finished by edge enhancing diffusion, is then applied into optic flow data coding. This method can well adapt to the characteristics of optic flow data and allow us to obtain an accurate and detailed motion representation with a reasonable rate. As to DFD coding, a simple but efficient block discrete cosine transform based method is implemented, which only encodes the energy most-significant positions according to the given bit rate. We also analyze the DFD coding with the help of an energy mask and implement a similar improved method. Finally we demonstrate our approach on a number of test video sequences and present some comparative results.

Gabriela GHIMPETEANU

NATIONALITY: **Romanian**
DEPARTMENT: **Mathematical Image Analysis**

THESIS TITLE:

Non-Photorealistic Rendering of 2D Face Images

ABSTRACT OF MASTER'S THESIS:

We provide an artistic filter inspired from Arcimboldo's portrait of the holy roman emperor Rudolf II painted as the roman god Vertumnus. Even though the portrait is a sophisticated construction made entirely out of fruits, vegetables and owers, the portrait subject can still be recognized. To keep this likeness for our filter, we need to match each face feature with the most similar fruit or vegetable. Therefore, we first create a method for extracting each feature from the input portrait photo. Moreover, we introduce a similarity measure for comparing two shapes (defined as matrices, or black and white images). Then we match the detected face feature with the most similar fruit, found by the similarity method in our previously created fruit database.

Marek Lech HAMERLIK

NATIONALITY: **Polish**
DEPARTMENT: **Information Security and Cryptography**

THESIS TITLE:

Anonymity and Censorship Resistance in Semantic Overlay Networks

ABSTRACT OF MASTER'S THESIS:

This thesis presents Clouds, a peer-to-peer architecture that guarantees both anonymity and censorship resistance in semantic overlay networks. The design of the underlying protocol needs to meet a number of challenging goals: allowing for the exchange of encrypted messages without assuming previously shared secrets, avoiding centralized infrastructures, like trusted servers or gateways, and guaranteeing efficiency without establishing direct connections between peers. Anonymity is achieved by cloaking the identity of protocol participants behind groups of semantically close peers. Censorship resistance is guaranteed by a cryptographic protocol securing the anonymous communication between the querying peer and the resource provider. Although we ground our technique on semantic overlay networks to exploit their retrieval capabilities, our framework is general and can be applied to any unstructured overlay network. Experimental results demonstrate the security properties of Clouds under different attacks and show the message overhead and retrieval effectiveness of the protocol.

Stefan HOLDER

NATIONALITY: German
DEPARTMENT: Databases and Information Systems

THESIS TITLE:

Replication in Unstructured Peer-to-Peer Networks with Availability Constraints

ABSTRACT OF MASTER'S THESIS:

Random walks have been proven a scalable search strategy for unstructured peer-to-peer (P2P) networks. In a random walk, a query message is forwarded at each step to a randomly chosen neighbor for a limited number of hops. If random walks are employed as a search strategy, data replication is an important method to increase the probability of successful search. In this context, we investigate the optimization problem to find a replica assignment to peers that satisfies given availability constraints and minimizes the number of replicas for each data item. Availability constraints are minimum thresholds for the probability of successful search and can be specified for each peer and each data item separately. In contrast to prior work, our approach aims to consider individual peer requirements regarding search success in the presence of temporary peer outages.

In this thesis, we present a distributed P2P replication algorithm for the above-mentioned optimization problem and prove an approximation guarantee for this algorithm. Our approach to devise such an algorithm is to first study the problem in a centralized setting (i.e. with complete knowledge of the network) and to develop a centralized algorithm. We then turn the centralized algorithm into a distributed one by exploiting local properties of the centralized algorithm. We show by simulations that the approximation solution is close to an optimum one in P2P settings. Another result of our evaluation is that the distributed algorithm efficiently performs replica assignments to peers on sparse network graphs with a large diameter. Mobile P2P networks, which are a prime example for such a graph topology, are therefore an interesting application area for our algorithm.

Olha HONCHAROVA

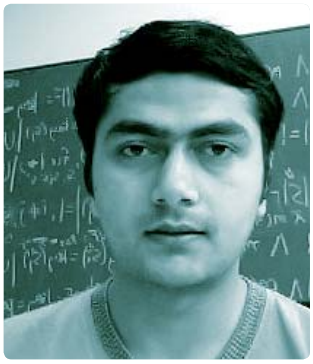
NATIONALITY: **Ukrainian**
DEPARTMENT: **Compiler Design**

THESIS TITLE:

Static Detection of Parametric Loop Bounds on C Code

ABSTRACT OF MASTER'S THESIS:

Static Worst-Case Execution Time (WCET) analysis is a technique to derive upper bounds for the execution times of programs. Such bounds are crucial when designing and verifying real-time systems. A key component for static derivation of preciseWCET estimates is upper bounds on the number of times loops can be iterated. The idea of the parametric loop bound analysis is to express the upper loop bound as a formula depending on parameters – variables and expressions staying constant within the loop body. The formula is constructed once for each loop. Then by instantiating this formula with values of parameters acquired externally (from value analysis, etc.), a concrete loop bound can be computed without high computational effort.



Alekh JINDAL

NATIONALITY: Indian
DEPARTMENT: Information Systems

THESIS TITLE:

Quality in Phrase Mining

ABSTRACT OF MASTER'S THESIS:

Phrase snippets of large text corpora like news articles or web search results offer great insight and analytical value. While much of the prior work is focussed on efficient storage and retrieval of all candidate phrases, little emphasis has been laid on the quality of the result set. In this thesis, we define phrases of interest and propose a framework for mining and post-processing interesting phrases. We focus on the quality of phrases and develop techniques to mine minimal-length maximal-informative sequences of words. The techniques developed are streamed into a post-processing pipeline and include exact and approximate match-based merging, incomplete phrase detection with filtering, and heuristics-based phrase classification. The strategies aim to prune the candidate set of phrases down to the ones being meaningful and having rich content. We characterize the phrases with heuristics- and NLP-based features. We use a supervised learning based regression model to predict their interestingness. Further, we develop and analyze ranking and grouping models for presenting the phrases to the user. Finally, we discuss relevance and performance evaluation of our techniques. Our framework is evaluated using a recently released real world corpus of New York Times news articles.

Tomasz JURKIEWICZ

NATIONALITY: **Polish**
DEPARTMENT: **Algorithms and Complexity**

THESIS TITLE:

Cycle Bases in Graphs

ABSTRACT OF MASTER'S THESIS:

Cycles in graphs play an important role in many applications, e.g., analysis of electrical networks, analysis of chemical and biological pathways, periodic scheduling, and graph drawing. Cycle bases are a compact description of the set of all cycles of a graph and cycle bases consisting of short cycles or, in weighted graphs, of small weight cycles are preferable. This thesis presents results in two following areas.

First, we will concentrate on new structural results which might help to determine if the minimum weight cycle basis problem (MCB) for integral and totally unimodular bases is in P or is NP-complete. A new potentially useful class of double cover bases is also presented. The main reason for research in this area is the fact that known algorithms for finding MCBs in general graphs return bases with rather poor theoretical properties.

Second, we will present an algorithm for computing general minimum weight cycle bases in time $O(E^{\omega})$ for general graphs; here V and E denote the number of nodes and edges, respectively, and ω is the exponent of the fastest matrix multiplication algorithm. Our algorithm is the first to run faster than $\tilde{O}(E^2 V)$. A key to our improved running time is the insight that the search for the minimum basis can be restricted to a set of candidate cycles of total length $O(VE)$.

Javor KALOJANOV

NATIONALITY: Bulgarian
DEPARTMENT: Computer Graphics

THESIS TITLE:

Parallel and Lazy Construction of Grids for Ray Tracing on Graphics Hardware

ABSTRACT OF MASTER'S THESIS:

In this thesis we investigate the use of uniform grids as acceleration structures for ray tracing on data-parallel machines such as modern graphics processors. The main focus of this work is the trade-offs between construction time and rendering performance provided by the acceleration structures, which is important for rendering dynamic scenes. We propose several parallel construction algorithms for uniform and two-level grids as well as a ray triangle intersection algorithm, which improves SIMD utilization for incoherent rays. The result of this work is a GPU ray tracer with performance for dynamic scenes that is comparable and in some cases better than the best known implementations today.

Megha KHOSLA

NATIONALITY: Indian

DEPARTMENT: Algorithms and Complexity

THESIS TITLE:

Message Passing Algorithms

ABSTRACT OF MASTER'S THESIS:

Constraint Satisfaction Problems (CSPs) are defined over a set of variables whose state must satisfy a number of constraints. We study a class of algorithms called Message Passing Algorithms, which aim at finding the probability distribution of the variables over the space of satisfying assignments. These algorithms involve passing local messages (according to some message update rules) over the edges of a factor graph constructed corresponding to the CSP. We focus on the Belief Propagation (BP) algorithm, which finds exact solution marginals for tree-like factor graphs. However, convergence and exactness cannot be guaranteed for a general factor graph. We propose a method for improving BP to account for cycles in the factor graph. We also study another message passing algorithm known as Survey Propagation (SP), which is empirically quite effective in solving random K-SAT instances, even when the density is close to the satisfiability threshold. We contribute to the theoretical understanding of SP by deriving the SP equations from the BP message update rules.

Luis Eugenio KUHN CUELLAR

NATIONALITY: **Mexican**
DEPARTMENT: **Bioinformatics and Applied Algorithmics**

THESIS TITLE:

A Probabilistic Algorithm for Matching Protein Structures - and its Application to Detecting Functionally Relevant Patterns

ABSTRACT OF MASTER'S THESIS:

In this thesis, we develop a randomized sub-graph matching method to detect common substructures between pairs of proteins. We use a representation of the protein structure that provides two different levels of abstraction. A high level of abstraction uses a labeled graph to represent the topology of the protein structure and its chemical properties. A lower abstraction level interprets every vertex in the graph as a point in three-dimensional space, it uses atomic coordinates to provide a detailed description of the geometry of the protein. This fine grained representation of protein structures captures large amounts of geometric and chemical information and allows us to move between different abstraction levels to efficiently detect structural similarity.

The method proceeds in three main steps: Feature vertex selection, local descriptor comparison and RANSAC sub-graph matching. The objective of the sub-graph matching routine is to find similar sub-graphs between two labeled graphs. This is an iterative algorithm that samples the solution space by randomly choosing an initial vertex correspondence, once an initial correspondences has been established, a greedy expansion procedure extends the match until no more suitable correspondences can be found in the vicinity.

We provide evidence of the suitability of this approach for protein substructure matching. This method was initially tested on small sets of enzymes that contain well-know structural motifs. We then used larger data sets of protein structures annotated with molecular function GO terms in order to extract collections of structural patterns.

The major contribution of this work was the implementation of a pairwise matching tool for local comparison of protein structures. Our tool is able to rapidly extract a large number of structural similarities, ranging from small structural motifs (e.g. catalytic triads and zinc fingers) to large domains. It implements a sub-graph matching method as a C++ routine that can be accessed through a Python interface, making it compatible with other python-based Bioinformatics tools. The design of our implementation aims at extensibility to other matching problems in Structural Bioinformatics.

Faraz MAKARI MANSHADI

NATIONALITY: German

DEPARTMENT: Databases and Information Systems

THESIS TITLE:

Fast Distributed Replication in Modern Networks

ABSTRACT OF MASTER'S THESIS:

In modern networks such as wireless and P2P networks and distributed search engines, replicating data is an important issue, which is often needed for availability, reliability and performance improvement. In such distributed systems, often random walks are utilized for query routing and data replication can improve the probability of successfully finding requested items. In this respect, we investigate the problem of finding a replica allocation to peers of a network such that at each peer a successful search is guaranteed and the total number of replicated data items in the network is minimized. Obviously, the number of replicated data items in the whole network is a global parameter. There exists a trade-off between the amount of local information and the quality of the solution for the global objective function. The challenging point here is to optimize this global function only based on the local information. The literature contains a huge amount of theoretical work. Most of the proposed algorithms are, due to their complexity, far from being practical or based on some unrealistic simplifying assumptions, e.g. unbounded message sizes. On the other hand, many practical works lack in theoretical guarantees. In this thesis, we present two efficient and practical distributed algorithms for replication, whose effectiveness has been shown by an extensive experimental evaluation on random graphs and internet-style topologies. Furthermore, for the first distributed algorithm, a rigorous theoretical study has also been provided. Our proposed algorithms are based on the LP-relaxation technique and uses only small messages. Moreover, they are simple to implement and can be applied in dynamic settings where nodes join and leave the system. These properties are very desirable in distributed systems with unreliable components such as P2P networks.

Yulya PATENKO

NATIONALITY: Russian
DEPARTMENT: Software Engineering

THESIS TITLE:

Improving Defect Prediction by Code Clustering

ABSTRACT OF MASTER'S THESIS:

Building high-quality software requires accurate and intensive testing before a project can be released to the customers. Limited time and human resources make it imperative that the testing efforts are focused on the most defect-prone areas. Efficient allocation of testing resources can significantly improve the process of defect prevention. Therefore, it is important to determine those parts of the code that should be tested more thoroughly.

Recently developed defect prediction models can help to discover optimal testing strategies. Such models perform well with projects that have a strong focus on single application functionality. However, they often result in lower prediction accuracy when applied to software systems with wide range of functionality.

In this thesis we explore impact of software decomposition on the efficiency of defect prediction models. To this end, we decompose software artifacts into clusters containing entities from the same application layer or with similar functionality. For every obtained clusters we find the set of metrics which best predicts the failure probability. To estimate the quality of a new entity, we first determine the cluster it belongs to and then apply the corresponding predictor.

We empirically evaluate proposed method using four decomposition techniques applied onto different software projects and compare prediction results against each other and against an overall benchmark model.

Henning PETERS

NATIONALITY: German

DEPARTMENT: Computer Graphics

THESIS TITLE:

Hardware and Software Extensions for a FTIR Multi-Touch Interface

ABSTRACT OF MASTER'S THESIS:

Even though multi-touch screens based on *frustrated total internal reflection* (FTIR) principle have gained significant popularity over the last years their overall performance is often limited by design. In particular, we identified robustness towards ambient light, high-speed tracking and precise calibration as key issues in many systems. In this thesis we propose a set of software and hardware extensions for a custom built FTIR screen that improve on these aspects by reducing the impact of ambient light, optimizing tracking performance and performing an automated calibration at high precision. Comparisons with existing approaches reveal great optimization potential: Our system performs well under intense and rapidly changing lighting conditions, its tracking runtime is up to 30x faster than Touchlib and calibration achieves sub-pixel resolution. The results are general enough to apply equally well to other camera-based technologies and might influence the next generation of multitouch processing software.

Dmytro PUZHAY

NATIONALITY: **Ukrainian**
DEPARTMENT: **Software Engineering**

THESIS TITLE:

Modeling Bug Reporter Reputation

ABSTRACT OF MASTER'S THESIS:

Tracking and resolving of software bugs are very important tasks for software developers and maintainers. Bug-tracking systems are tools which are widely used in open source projects to support these activities.

The empirical Software Engineering research community pays considerable attention to bug-tracking-related topics in order to provide bug-tracking systems users with adequate software and tool support. Bug-tracking is a highly socialized process which requires constant communication between developers and bug reporters. However, the inherent social structure of bug tracking systems and its influence on everyday bug-tracking has earlier been poorly studied.

In this work I address the role of bug reporter reputation. Using publicly available information from bug-tracking system database, I model bug reporter reputation to check whether there is any evidence of relation between reporter reputation and attention from developers his bugs get.

If reputation actually plays important role in bug-tracking activities and can relatively easily be extracted, existing prediction techniques could potentially be improved by using reputation as additional input variable; bug-tracking software could be supported with more formal notion of reporter reputation.

Roxana RAGNEALA

NATIONALITY: Romanian
DEPARTMENT: Software Engineering

THESIS TITLE:

Code History: A Useful Resource for Defect Prediction Models

ABSTRACT OF MASTER'S THESIS:

Predicting likely software defects in the future is valuable for project managers when planning resource allocation for software testing. But building prediction models using only code metrics may not be suffice for accurate results.

In this work, we investigate the value of code history metrics that can be collected from the project's version archives for the purpose of defect prediction. Our results suggest that prediction models built using code history metrics outperform those using traditional code.

Christine RIZKALLAH

NATIONALITY: Canadian
DEPARTMENT: Programming Systems

THESIS TITLE:

Proof Representations for Higher-Order Logic

ABSTRACT OF MASTER'S THESIS:

We provide a mapping from classical extensional tableau proofs of higher-order formulas to intuitionistic non extensional natural deduction proofs of semantically equivalent formulas.

We show that the Kuroda transformation, which is known to map from first-order classical logic to first-order intuitionistic logic, extends to elementary type theory.

Moreover, we introduce a transformation that we call Girard-Kuroda-Per and prove that this transformation maps from classical extensional to intuitionistic non-extensional simple type theory.

Vinay SETTY

NATIONALITY: Indian

DEPARTMENT: Databases and Information Systems

THESIS TITLE:

Efficiently Identifying Interesting Time-points in Text Archives

ABSTRACT OF MASTER'S THESIS:

Large scale text archives are increasingly becoming available on the Web. Exploring their evolving contents along both text and temporal dimensions enables us to realize their full potential. Standard keyword queries facilitate exploration along the text dimension only. Recently proposed time-travel keyword queries enable query processing along both dimensions, but require the user to be aware of the exact time point of interest. This may be impractical if the user does not know the history of the query within the collection or is not familiar with the topic.

In this work, our aim is to efficiently identify interesting time points in Web archives with an assumption that we receive a result list for a given query in standard relevance-order from an existing retrieval system. We consider two forms of Web archives: (i) one where documents have a publication time-stamp and never change (such as news archives), and (ii) the archives where documents undergo revisions, and are thus versioned. In both settings, we define interestingness as the change in top-k result set of two consecutive time-points. The key step in our solution is the maintenance of top-k results valid at each time-point of the archive, which can then be used to compute the interestingness scores for the time-points. We propose two techniques to realize efficient identification of interesting time points: (i) For the case when documents once published never change, we have a simple but effective technique. (ii) For the more general case with versioned documents, we develop an extension to the segment tree which makes it rank-aware and dynamic. To further improve efficiency, we propose an early termination technique which is proven to be very effective. Our methods are shown to be effective in efficiently finding interesting time points in a set of experiments using the New York Times news archive and the Wikipedia versioned archive.

Tomasz TYLENDA

NATIONALITY: **Polish**
DEPARTMENT: **Databases and Information Systems**

THESIS TITLE:

Time-aware Link Prediction in Evolving Social Networks

ABSTRACT OF MASTER'S THESIS:

Data sets in the form of networks occur in many domains such as sociology, biology, engineering, etc. One of the tasks that can be performed on them is the prediction of links, both new as well as recurring ones. Link prediction appears in many applications, e.g. in prediction of new friendship links in online communities such as Facebook. The problem can be tackled by using the graph structure alone or with a combination of node and edge attributes. The attributes which are utilized are mainly domain specific properties of networks. Surprisingly, prior work pays little attention to temporal information.

We investigate in this work the value of incorporating into link prediction methods the history information available on the interactions (or links) of the current social network state. Our results unequivocally show that timestamps of past interactions significantly improve the prediction accuracy of new and recurrent links compared to rather sophisticated methods proposed recently. Furthermore, we introduce a novel testing method necessary to evaluate the benefits of incorporating time-awareness into link prediction.

Mohammad Reza YOUSEFI

NATIONALITY: Iranian

DEPARTMENT: Computer Graphics

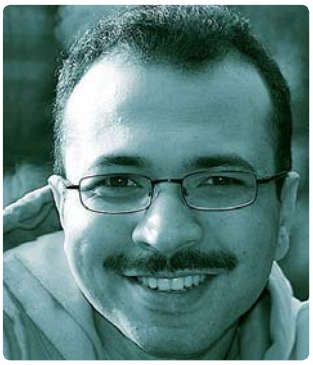
THESIS TITLE:

Generating Detailed Face Models by Controlled Lighting

ABSTRACT OF MASTER'S THESIS:

This thesis describes the design and implementation of a system for generating detailed face models by controlled lighting. The first part of the thesis is devoted to implementation of a scanner based on structured light to capture the geometry. The performance of the structured light scanner is improved through a subpixel precision step based on interpolation of codewords in the reconstruction phase. The second part details the construction of a light stage for performing photometric stereo, and basically computing the surface normals. The quality of photometric stereo highly depends upon the accurate control of the illumination condition. Our light stage consists of 150 LED sources, each of which can be illuminated and controlled over a wide range of resolution, and hence provides the required level of precision. Besides being highly accurate, it is also considered for the light stage design to be inexpensive to build and easy to produce and maintain.

Alumni: PhD



Dr. Naveed AHMAD

NATIONALITY: **Pakistani**
DEPARTMENT: **Computer Graphics**

DISSERTATION TITLE:

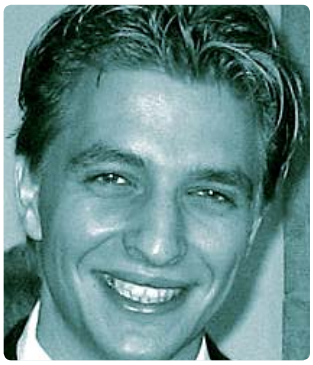
High Quality Dynamic Reflectance and Surface Reconstruction from Video

ABSTRACT OF PHD THESIS:

The creation of high quality animations of real-world human actors has long been a challenging problem in computer graphics. It involves the modeling of the shape of the virtual actors, creating their motion, and the reproduction of very fine dynamic details. In order to render the actor under arbitrary lighting, it is required that reflectance properties are modeled for each point on the surface. These steps, that are usually performed manually by professional modelers, are time consuming and cumbersome.

In this thesis, we show that algorithmic solutions for some of the problems that arise in the creation of high quality animation of real-world people are possible using multi-view video data. First, we present a novel spatio-temporal approach to create a personalized avatar from multi-view video data of a moving person. Thereafter, we propose two enhancements to a method that captures human shape, motion and reflectance properties of a moving human using eight multi-view video streams. Afterwards we extend this work, and in order to add very fine dynamic details to the geometric models, such as wrinkles and folds in the clothing, we make use of the multi-view video recordings and present a statistical method that can passively capture the fine-grain details of time-varying scene geometry. Finally, in order to reconstruct structured shape and animation of the subject from video, we present a dense 3D correspondence finding method that enables spatio-temporally coherent reconstruction of surface animations directly from multi-view video data.

These algorithmic solutions can be combined to constitute a complete animation pipeline for acquisition, reconstruction and rendering of high quality virtual actors from multi-view video data. They can also be used individually in a system that requires the solution of a specific algorithmic sub-problem. The results demonstrate that using multi-view video data it is possible to find the model description that enables realistic appearance of animated virtual actors under different lighting conditions and exhibits high quality dynamic details in the geometry.



Dr. Andre ALTMANN

NATIONALITY: German
DEPARTMENT: Computational Biology & Applied Algorithmics

DISSERTATION TITLE:

Bioinformatical Approaches to Ranking of Anti-HIV Combination Therapies and Planning of Treatment Schedules

ABSTRACT OF PHD THESIS: :

The human immunodeficiency virus (HIV) pandemic is one of the most serious health challenges humanity is facing today. Combination therapy comprising multiple antiretroviral drugs resulted in a dramatic decline in HIV-related mortality in the developed countries. However, the emergence of drug resistant HIV variants during treatment remains a problem for permanent treatment success and seriously hampers the composition of new active regimens.

In this thesis we use statistical learning for developing novel methods that rank combination therapies according to their chance of achieving treatment success. These depend on information regarding the treatment composition, the viral genotype, features of viral evolution, and the patient's therapy history. Moreover, we investigate different definitions of response to antiretroviral therapy and their impact on prediction performance of our method. We address the problem of extending purely data-driven approaches to support novel drugs with little available data. In addition, we explore the prospect of prediction systems that are centered on the patient's treatment history instead of the viral genotype. We present a framework for rapidly simulating resistance development during combination therapy that will eventually allow application of combination therapies in the best order.

Finally, we analyze surface proteins of HIV regarding their susceptibility to neutralizing antibodies with the aim of supporting HIV vaccine development.

Dr. Ralitsa ANGELOVA

NATIONALITY: **Bulgarian**
DEPARTMENT: **Databases and Information Systems**

DISSERTATION TITLE:

Graph-Based Classification and Clustering of Entities in Heterogeneous Networks

ABSTRACT OF PHD THESIS:

We address the problem of multi-label classification of relational graphs by proposing a framework that models the input graph as a first order Markov random field and devises a relaxation labeling procedure to find its maximally likely labeling. We apply this framework to classification as well as clustering problems in homogeneous networks and show significant performance gains in comparison to state-of-the-art techniques.

We also address the problem of multi-label classification in heterogeneous networks where every data point is associated with a node type and has to be labeled with one or more classes from a type-specific finite set of classes. Our algorithm is based on a random walk model. We present detailed empirical studies of our model and compare it with state-of-art techniques on two social networks.

All newly proposed algorithms are robust to scarce training data and diverse linkage patterns. They improve classification or clustering quality in homogeneous and heterogeneous networks.

Dr. Tunc Ozan AYDIN

NATIONALITY: Turkish
DEPARTMENT: Computer Graphics

DISSERTATION TITLE:

Human Visual System Models in Computer Graphics

ABSTRACT OF PHD THESIS:

At the receiving end of visual data are humans; thus it is only natural to take into account various properties and limitations of the human visual system while designing new image and video processing methods. In this dissertation we build multiple models of human vision with different focuses and complexities, and demonstrate their use in computer graphics context.

The human visual system models we present perform two fundamental tasks: predicting the visual significance, and the detection of visual features. We start by showing that a perception based importance measure for edge strength prediction results in qualitatively better outcomes compared to commonly used gradient magnitude measure in multiple computer graphics applications. Another more comprehensive model including mechanisms to simulate maladaptation is used to predict the visual significance of images shown on display devices under dynamically changing lighting conditions.

The detection task is investigated in the context of image and video quality assessment. We present an extension to commonly used image quality metrics that enables HDR support while retaining backwards compatibility with LDR content. We also propose a new “dynamic range independent” image quality assessment method that can compare HDR-LDR (and vice versa) reference-test image pairs, in addition to image pairs with the same dynamic range. Furthermore, the design and validation of a dynamic range independent video quality assessment method, that models various spatiotemporal aspects of human vision, is presented along with pointers to a wide range of application areas including comparison of rendering qualities, HDR compression and temporal tone mapping operator evaluation.

Dr. Klaus BERBERICH

NATIONALITY: German
DEPARTMENT: Databases and Information Systems

DISSERTATION TITLE:

Temporal Search in Web Archives

ABSTRACT OF PHD THESIS:

Web archives include both archives of contents originally published on the Web (e.g., the Internet Archive) but also archives of contents published long ago that are now accessible on the Web (e.g., the archive of The Times). Thanks to the increased awareness that web-born contents are worth preserving and to improved digitization techniques, web archives have grown in number and size. To unfold their full potential, search techniques are needed that consider their inherent special characteristics.

This work addresses three important problems toward this objective and makes the following contributions:

- We present the Time-Travel Inverted indeX (TTIX) as an efficient solution to time-travel text search in web archives, allowing users to search only the parts of the web archive that existed at a user's time of interest.
- To counter negative effects that terminology evolution has on the quality of search results in web archives, we propose a novel query-reformulation technique, so that old but highly relevant documents are retrieved in response to today's queries.
- For temporal information needs, for which the user is best satisfied by documents that refer to particular times, we describe a retrieval model that integrates temporal expressions (e.g., "in the 1990s") seamlessly into a language modeling approach.

Experiments for each of the proposed methods show their efficiency and effectiveness, respectively, and demonstrate the viability of our approach to search in web archives.

Dr. Gerard DE MELO

NATIONALITY: German
DEPARTMENT: Databases and Information Systems

DISSERTATION TITLE:

Graph-based Methods for Large-Scale Multilingual Knowledge Integration

ABSTRACT OF PHD THESIS:

Given that much of our knowledge is expressed in textual form, information systems are increasingly dependent on knowledge about words and the entities they represent. This thesis investigates novel methods for automatically building large repositories of knowledge that capture semantic relationships between words, names, and entities, in many different languages. Three major contributions are made, each involving graph algorithms and statistical techniques that combine evidence from multiple sources of information.

The lexical integration method involves learning models that disambiguate word meanings based on contextual information in a graph, thereby providing a means to connect words to the entities that they denote. The entity integration method combines semantic items from different sources into a single unified registry of entities by reconciling equivalence and distinctness information and solving a combinatorial optimization problem. Finally, the taxonomic integration method adds a comprehensive and coherent taxonomic hierarchy on top of this registry, capturing how different entities relate to each other.

Together, these methods can be used to produce a large-scale multilingual knowledge base semantically describing over 5 million entities and over 16 million natural language words and names in more than 200 different languages.

Dr. Christian FUCHS

NATIONALITY: German
DEPARTMENT: Computer Graphics

DISSERTATION TITLE:

Capturing and Reconstructing the Appearance of Complex 3D Scenes

ABSTRACT OF PHD THESIS:

In this thesis, we present our research on new acquisition methods for reflectance properties of real-world objects. Specifically, we first show a method for acquiring spatially varying densities in volumes of translucent, gaseous material with just a single image. This makes the method applicable to constantly changing phenomena like smoke without the use of high-speed camera equipment.

Furthermore, we investigated how two well known techniques – synthetic aperture confocal imaging and algorithmic descattering – can be combined to help looking through a translucent medium like fog or murky water. We show that the depth at which we can still see an object embedded in the scattering medium is increased. In a related publication, we show how polarization and descattering based on phase-shifting can be combined for efficient 3D scanning of translucent objects. Normally, subsurface scattering hinders the range estimation by offsetting the peak intensity beneath the surface away from the point of incidence. With our method, the subsurface scattering is reduced to a minimum and therefore reliable 3D scanning is made possible.

Finally, we present a system which recovers surface geometry, reflectance properties of opaque objects, and prevailing lighting conditions at the time of image capture from just a small number of input photographs. While there exist previous approaches to recover reflectance properties, our system is the first to work on images taken under almost arbitrary, changing lighting conditions. This enables us to use images we took from a community photo collection website.

Dr. Jürgen GALL

NATIONALITY: German
DEPARTMENT: Computer Graphics

DISSERTATION TITLE:

Filtering and Optimization Strategies for Markerless Human Motion Capture with Skeleton-Based Shape Models

ABSTRACT OF PHD THESIS:

Since more than 2000 years, people have been interested in understanding and analyzing the movements of animals and humans which lead to the development of advanced computer systems for motion capture. Although marker-based systems for motion analysis are commercially successful, capturing the performance of a human or an animal from a multi-view video sequence without the need for markers is still a challenging task. The most popular methods for markerless human motion capture are model-based approaches that rely on a surface model of the human with an underlying skeleton. In this context, markerless motion capture seeks for the pose, i.e., the position, orientation, and configuration of the human skeleton that is best explained by the image data. In order to address this problem, we discuss the two questions:

1. What are good cues for human motion capture?

Typical cues for motion capture are silhouettes, edges, color, motion, and texture. In general, a multi-cue integration is necessary for tracking complex objects like humans since all these cues come along with inherent drawbacks. Besides the selection of the cues to be combined, reasonable information fusion is a common challenge in many computer vision tasks. Ideally, the impact of a cue should be large in situations when its extraction is reliable, and small, if the information is likely to be erroneous. To this end, we propose an adaptive weighting scheme that combines complementary cues, namely silhouettes on one side and optical flow as well as local descriptors on the

other side. Whereas silhouette extraction works best in case of homogeneous objects, optical flow computation and local descriptors perform better on sufficiently structured objects. Besides image-based cues, we also propose a statistical prior on anatomical constraints that is independent of motion patterns. Relying only on image features that are tracked over time does not prevent the accumulation of small errors which results in a drift away from the target object. The error accumulation becomes even more problematic in the case of multiple moving objects due to occlusions. To solve the drift problem for tracking, we propose an analysis-by-synthesis framework that uses reference images to correct the pose. It comprises an occlusion handling and is successfully applied to crash test video analysis.

2. Is human motion capture a filtering or an optimization problem?

Model-based human motion capture can be regarded as a filtering or an optimization problem. While local optimization offers accurate estimates but often looses track due to local optima, particle filtering can recover from errors at the expense of a poor accuracy due to overestimation of noise. In order to overcome the drawbacks of local optimization, we introduce a novel global stochastic optimization approach for markerless human motion capturing that is derived from the mathematical theory on interacting particle systems. We call the method interacting simulated annealing (ISA) since it is based on an interacting particle system that converges to the global optimum similar to simulated annealing. It estimates the human pose without initial information, which is a challenging optimization problem in a high dimensional space. Furthermore, we propose a tracking framework that is based on this optimization technique to achieve both the robustness of filtering strategies and a remarkable accuracy. In order to benefit from optimization and filtering, we introduce a multi-layer framework that combines stochastic optimization, filtering, and local optimization. While the first layer relies on interacting simulated annealing, the second layer refines the estimates by filtering and local optimization such that the accuracy is increased and ambiguities are resolved over time without imposing restrictions on the dynamics.

In addition, we propose a system that recovers not only the movement of the skeleton, but also the possibly non-rigid temporal deformation of the 3D surface. While large scale deformations or fast movements are captured by the skeleton pose and approximate surface skinning, true small scale deformations or non-rigid garment motion are captured by fitting the surface to the silhouette. In order to make automatic processing of large data sets feasible, the skeleton-based pose estimation is split into a local one and a lower dimensional global one by exploiting the tree structure of the skeleton.

Our experiments comprise a large variety of sequences for qualitative and quantitative evaluation of the proposed methods, including a comparison of global stochastic optimization with several other optimization and particle filtering approaches.

Dr. Rolf HARREN

NATIONALITY: German
DEPARTMENT: Algorithms and Complexity

DISSERTATION TITLE:

Two-dimensional Packing Problems

ABSTRACT OF PHD THESIS:

In this thesis we consider the two-dimensional bin packing problem and the strip packing problem, which are popular geometric generalizations of the classical bin packing problem.

In both problems, a list of rectangles has to be packed into a designated area such that no two rectangles overlap and all rectangles are packed axis-parallel. For the strip packing problem, the given items have to be packed into a strip of unit width and minimal height, whereas in the two-dimensional bin packing problem a packing has to be found into a minimal number of unit-sized bins.

We investigate approximation algorithms and online algorithms for these problems and consider variants where rotations of the rectangles are forbidden and where rotations by 90 degrees are allowed. In particular, we present two approximation algorithms for strip packing with approximation ratios 1.9396 and arbitrarily close to 5/3, respectively. These results are the first improvements upon the approximation ratio of 2 that was established 16 years ago. Moreover, we show an improved lower bound of 2.589 on the competitive ratio of online strip packing along with an upper bound of 2.618 for restricted input instances. For two-dimensional bin packing we derive best-possible approximation algorithms for the variants with and without rotations.

Dr. Nils HASLER

NATIONALITY: German
DEPARTMENT: Computer Graphics

DISSERTATION TITLE:

Modelling Human Pose and Shape from a Database of Human 3D Scans

ABSTRACT OF PHD THESIS:

Generating realistic human shapes and motion is an important task both in the motion picture industry and in computer games. In feature films, high quality and believability are the most important characteristics. Additionally, when creating virtual doubles the generated characters have to match as closely as possible to given real persons. In contrast, in computer games the level of realism does not need to be as high but real-time performance is essential. It is desirable to meet all these requirements with a general model of human pose and shape.

In addition, many markerless human tracking methods applied, e.g., in biomedicine or sports science can benefit greatly from the availability of such a model because most methods require a 3D model of the tracked subject as input, which can be generated on-the-fly given a suitable shape and pose model.

In this thesis, an encompassing procedure is presented to generate different general models of human pose. A database of 3D scans spanning the space of human pose and shape variations is introduced. Then, four different approaches for transforming the database into a general model of human pose and shape are presented, which improve the current state of the art. Experiments are performed to evaluate and compare the proposed models on real-world problems, i.e., characters are generated given semantic constraints and the underlying shape and pose of humans given 3D scans, multi-view video, or uncalibrated monocular images is estimated.

Dr. Robert HERZOG

NATIONALITY: German
DEPARTMENT: Computer Graphics

DISSERTATION TITLE:

Exploiting Coherence in Lighting and Shading Computations

ABSTRACT OF PHD THESIS:

Computing global illumination (GI) in virtual scenes becomes increasingly attractive even for real-time applications nowadays. GI delivers important cues in the perception of 3D virtual scenes, which is important for material and architectural design. Therefore, for photo-realistic rendering in the design and even the game industry, GI has become indispensable. While the computer simulation of realistic global lighting is well-studied and often considered as solved, computing it efficiently is not. Saving computation costs is therefore the main motivation of current research in GI. Efficient algorithms have to take various aspects into account, such as the algorithmic complexity and convergence, its mapping to parallel processing hardware, and the knowledge of certain lighting properties including the capabilities of the human visual system.

In this dissertation we exploit both low-level and high-level coherence in the practical design of GI algorithms for a variety of target applications ranging from high-quality production rendering to dynamic real-time rendering. We also focus on automatic rendering-accuracy control to approximate GI in such a way that the error is perceptually unified in the resulting images, thereby taking not only into account the limitations of the human visual system but also later video compression with an MPEG encoder. In addition, this dissertation provides many ideas and supplementary material, which complements published work and could be of practical relevance.

Dr. Thomas HILLENBRAND

NATIONALITY: German
DEPARTMENT: Automation of Logic

DISSERTATION TITLE:

Superposition and Decision Procedures – Back and Forth

ABSTRACT OF PHD THESIS:

Two apparently different approaches to automating deduction are mentioned in the title; they are the subject of a debate on “big engines vs. little engines of proof”. The contributions in this thesis advocate that these two strands of research can interplay in subtle and sometimes unexpected ways, such that mutual pervasion can lead to intriguing results:

Firstly, superposition can be run on top of decision procedures. This we demonstrate for the class of Shostak theories, incorporating a little engine into a big one. As another instance of decision procedures within superposition, we show that ground confluent rewrite systems, which decide entailment problems in equational logic, can be harnessed for detecting redundancies in superposition derivations.

Secondly, superposition can be employed as proof-theoretic means underneath combined decision procedures: We re-establish the correctness of the Nelson-Oppen procedure as an instance of the completeness of superposition. Thirdly, superposition can be used as a decision procedure for many interesting theories, turning a big engine into a little one. For the theory of bits and of fixed-size bitvectors, we suggest a rephrased axiomatization combined with a transformation of conjectures, based on which superposition decides the universal fragment. Furthermore, with a modification of lifting, we adapt superposition to the theory of bounded domains and give a decision procedure, which captures the Bernays-Schönfinkel class as well.

Dr. Matthias HORBACH

NATIONALITY: German

DEPARTMENT: Automation of Logic

DISSERTATION TITLE:

Saturation-Based Decision Procedures for Fixed Domain and Minimal Model Validity

ABSTRACT OF PHD THESIS:

Superposition is an established decision procedure for a variety of first-order logic theories. A satisfiable theory, saturated by superposition, implicitly defines a minimal Herbrand model for the theory. This raises the question in how far superposition calculi can be employed for reasoning about such minimal models.

In this thesis, I propose the first superposition calculus that can compute with respect to a minimal model and a fixed domain semantics and that is sound and complete in the limit. Since Skolemization is incomplete for these semantics, existentially quantified variables are explicitly represented. Based on the calculus, I present a number of decidability results for queries with one quantifier alternation.

Dr. Anna HUBER

NATIONALITY: German
DEPARTMENT: Algorithms and Complexity

DISSERTATION TITLE:

Randomized Rounding and Rumor Spreading with Stochastic Dependencies

ABSTRACT OF PHD THESIS:

Randomness is an important ingredient of modern computer science. The present thesis is concerned with two uses of randomness viz. randomized roundings and randomized rumor spreading algorithms.

The theorem of Beck and Fiala (1981) asserts that for every hypergraph and every set of vertex weights there is a rounding of the vertex weights such that the additive rounding error for all hyperedges is bounded by the maximum degree. This theorem will be extended to randomized roundings, that is, to roundings that are efficiently generated at random in such a way that each value is rounded up with probability equal to its fractional part.

The larger part of this thesis deals with randomized rumor spreading algorithms. These are protocols for disseminating information on graphs.
The classical randomized rumor spreading was introduced and first investigated by Frieze and Grimmett on the complete graph (1985). We show a generalization of their results both in terms of the model used and in terms of the underlying graph.

We also investigate a quasirandom rumor spreading protocol introduced by Doerr, Friedrich, and Sauerwald (2008). We present a detailed analysis of its evolution and show that its performance and robustness match performance and robustness of the randomized rumor spreading protocol.

The unifying idea is to use dependencies so as to obtain results that are superior or equal to those obtained via independent randomness.

Dr. Matthias HULLIN

NATIONALITY: German
DEPARTMENT: Computer Graphics

DISSERTATION TITLE:

Reconsidering Light Transport – Acquisition and Display of Real-World Reflectance and Geometry

ABSTRACT OF PHD THESIS:

In this thesis, we cover three scenarios that violate common simplifying assumptions about the nature of light transport.

We begin with the first ingredient to any 3D rendering: a geometry model. Most 3D scanners require the object-of-interest to show diffuse reflectance. The further a material deviates from the Lambertian model, the more likely these setups are to produce corrupted results. By placing a traditional laser scanning setup in a participating (in particular, fluorescent) medium, we have built a light sheet scanner that delivers robust results for a wide range of materials, including glass.

Further investigating the phenomenon of fluorescence, we notice that, despite its ubiquity, it has received moderate attention in computer graphics. In particular, to date no datadriven reflectance models of fluorescent materials have been available. To describe the wavelength-shifting reflectance of fluorescent materials, we define the bispectral bidirectional reflectance and reradiation distribution function (BRRDF), for which we introduce an image-based measurement setup as well as an efficient acquisition scheme.

Finally, we envision a computer display that shows materials instead of colours, and present a prototypical device that can exhibit anisotropic reflectance distributions similar to common models in computer graphics.

Dr. Georgiana IFRIM

NATIONALITY: Romanian
DEPARTMENT: Databases and Information Systems

DISSERTATION TITLE:

Statistical Learning Techniques for Text Categorization with Sparse Labeled Data

ABSTRACT OF PHD THESIS:

Many applications involve learning a supervised classifier from very few explicitly labeled training examples, since the cost of manually labeling the training data is often prohibitively high. For instance, we expect a good classifier to learn our interests from a few example books or movies we like, and recommend similar ones in the future, or we expect a search engine to give more personalized search results based on whatever little it learned about our past queries and clicked documents.

There is thus a need for classification techniques capable of learning from sparse labeled data, by exploiting additional information about the classification task at hand (e.g., background knowledge) or by employing more sophisticated features (e.g., n-gram sequences, trees, graphs). In this thesis, we focus on two approaches for overcoming the bottleneck of sparse labeled data.

We first propose the Inductive/Transductive Latent Model (ILM/TLM), which is a new generative model for text documents. ILM/TLM has various building blocks designed to facilitate the integration of background knowledge (e.g., unlabeled documents, ontologies of concepts, encyclopedia) into the process of learning from small training data. Our method can be used for inductive and transductive learning and achieves significant gains over state-of-the-art methods for very small training sets.

Second, we propose Structured Logistic Regression (SLR), which is a new coordinate-wise gradient ascent technique for learning logistic regression in the space of all (word or character) sequences in the training data. SLR exploits the inherent structure of the n-gram feature space in order to automatically provide a compact set of highly discriminative n-gram features. Our detailed experimental study shows that while SLR achieves similar classification results to those of the state-of-the-art methods (which use all n-gram features given explicitly), it is more than an order of magnitude faster than its opponents.

The techniques presented in this thesis can be used to advance the technologies for automatically and efficiently building large training sets, therefore reducing the need for spending human computation on this task.

Dr. Carsten IHLEMANN

NATIONALITY: German
DEPARTMENT: Automation of Logic

DISSERTATION TITLE:

Reasoning in Combinations of Theories

ABSTRACT OF PHD THESIS:

Verification problems are often expressed in a language which mixes several theories. A natural question to ask is whether one can use decision procedures for individual theories to construct a decision procedure for the union theory. In the cases where this is possible one has a powerful method at hand to handle complex theories effectively.

The setup considered in this thesis is that of one base theory which is extended by one or more theories. The question is if and when a given ground satisfiability problem in the extended setting can be effectively reduced to an equi-satisfiable problem over the base theory. A case where this reductive approach is always possible is that of so-called local theory extensions.

The theory of local extensions is developed and some applications concerning monotone functions are given. Then the theory of local theory extensions is generalized in order to deal with data structures that exhibit local behavior. It will be shown that a suitable fragment of both the theory of arrays and the theory of pointers is local in this broader sense. Finally, the case of more than one theory extension is discussed. In particular, a modularity result is given that under certain circumstances the locality of each of the extensions lifts to locality of the entire extension.

The reductive approach outlined above has become particularly relevant in recent years due to the rise of powerful solvers for background theories common in verification tasks. These so-called SMT-solvers effectively handle theories such as real linear or integer arithmetic. As part of this thesis, a program called H-PILoT was implemented which carries out reductive reasoning for local theory extensions. H-PILoT found applications in mathematics, multiple-valued logics, data-structures and reasoning in complex systems.

Dr. Swen JACOBS

NATIONALITY: German
DEPARTMENT: Automation of Logic

DISSERTATION TITLE:

Hierarchic Decision Procedures for Verification

ABSTRACT OF PHD THESIS:

Information-handling systems are becoming ever more complex. They may be pure hardware or software systems, or complex systems of hardware and software that act in a real-world environment.

Verification is a method to ensure that systems behave in the expected way, which is a necessity for safety-critical applications like automatic railway control. The size of such systems makes manual verification impossible. Therefore, we need automatic or computer-aided verification procedures.

Automated reasoning is already widely used in the analysis and verification of systems. For a restricted class of systems, the resulting verification problems are inherently finite and can be solved efficiently. For complex systems, such finiteness cannot be expected. To express and prove properties of these systems, we need a formal language and reasoners that can deal with universal quantification, arithmetic expressions and unbounded data structures at the same time.

Thus, in recent years there has been new interest in the handling of first-order formulas modulo a given background theory. The problem is known to be undecidable in general, and research focuses mostly on methods that solve many problem instances quickly, but sacrifice completeness. We take a different approach and focus on instances of this problem that we can show to be decidable. In this way we can solve the resulting problems efficiently and guarantee termination.

This work is based on research by Sofronie-Stokkermans on local theory extensions and on work by Ganzinger and Korovin on instantiation-based first-order theorem proving. We extend the existing work on local theory extensions, giving new examples of axioms which satisfy a locality condition and using ideas from instantiation-based first-order theorem proving to make local reasoning more efficient. Furthermore, we show that local theory extensions allow us to decide certain verification problems for parameterized systems and develop increasingly complex system models of an automatic train controller on which we demonstrate how to use local reasoning to verify safety properties of such systems.

Dr. Daniel JOHANNSEN

NATIONALITY: German
DEPARTMENT: Algorithms and Complexity

DISSERTATION TITLE:

Random Combinatorial Structures and Randomized Search Heuristics

ABSTRACT OF PHD THESIS:

This thesis is concerned with the probabilistic analysis of random combinatorial structures and the runtime analysis of randomized search heuristics.

On the subject of random structures, we investigate two classes of combinatorial objects. The first is the class of planar maps and the second is the class of generalized parking functions. We identify typical properties of these structures and show strong concentration results on the probabilities that these properties hold. To this end, we develop and apply techniques based on exact enumeration by generating functions. For several types of random planar maps, this culminates in concentration results for the degree sequence. For parking functions, we determine the distribution of the defect, the most characteristic parameter.

On the subject of randomized search heuristics, we present, improve, and unify different probabilistic methods and their applications. In this, special focus is given to potential functions and the analysis of the drift of stochastic processes. We apply these techniques to investigate the runtimes of evolutionary algorithms. In particular, we show for several classical problems in combinatorial optimization how drift analysis can be used in a uniform way to give bounds on the expected runtimes of evolutionary algorithms.

Dr. Gjergji KASNECI

NATIONALITY: German

DEPARTMENT: Databases and Information Systems

DISSERTATION TITLE:

Searching and Ranking in Entity-Relationship Graphs

ABSTRACT OF PHD THESIS:

The Web bears the potential to become the world's most comprehensive knowledge base. Organizing information from the Web into entity-relationship graph structures could be a first step towards unleashing this potential. In a second step, the inherent semantics of such structures would have to be exploited by expressive search techniques that go beyond today's keyword search paradigm. In this realm, we have developed NAGA (Not Another Google Answer), a semantic search engine which provides expressive means for querying, searching and ranking at entity-relationship level. For knowledge discovery tasks, like finding broad or interesting relations between k (≥ 2) given entities NAGA comes with two efficient algorithms: STAR (Steiner Tree Approximation in Relationship Graphs), and MING (Mining Informative Graphs). NAGA is a fully implemented prototype system and is part of the YAGO-NAGA project.

Dr. Michael KERBER

NATIONALITY: German
DEPARTMENT: Algorithms and Complexity

DISSERTATION TITLE:

Geometric Algorithms for Algebraic Curves and Surfaces

ABSTRACT OF PHD THESIS:

This work presents novel geometric algorithms dealing with algebraic curves and surfaces of arbitrary degree. These algorithms are exact and complete they return the mathematically true result for all input instances. Efficiency is achieved by cutting back expensive symbolic computation and favoring combinatorial and adaptive numerical methods instead, without spoiling exactness in the overall result.

We present an algorithm for computing planar arrangements induced by real algebraic curves. We show its efficiency both in theory by a complexity analysis, as well as in practice by experimental comparison with related methods. For the latter, our solution has been implemented in the context of the CGAL library. The results show that it constitutes the best current exact implementation available for arrangements as well as for the related problem of computing the topology of one algebraic curve. The algorithm is also applied to related problems, such as arrangements of rotated curves, and arrangements embedded on a parameterized surface.

In $\mathbb{R}$, we propose a new method to compute an isotopic triangulation of an algebraic surface. This triangulation is based on a stratification of the surface, which reveals topological and geometric information. Our implementation is the first for this problem that makes consequent use of numerical methods, and still yields the exact topology of the surface.

Dr. Stefan KRATSCH

NATIONALITY: German
DEPARTMENT: Algorithms and Complexity

DISSERTATION TITLE:

Kernelization of Generic Problems: Upper and Lower Bounds

ABSTRACT OF PHD THESIS:

This thesis addresses the kernelization properties of generic problems, defined via syntactical restrictions or by a problem framework. Polynomial kernelization is a formalization of data reduction, aimed at combinatorially hard problems, which allows a rigorous study of this important and fundamental concept. The thesis is organized into two main parts.

In the first part we prove that all problems from two syntactically defined classes of constant factor approximable problems admit polynomial kernelizations. The problems must be expressible via optimization over first-order formulas with restricted quantification; when relaxing these restrictions we find problems that do not admit polynomial kernelizations. Next, we consider edge modification problems, and we show that they do not generally admit polynomial kernelizations.

In the second part we consider three types of Boolean constraint satisfaction problems. We completely characterize whether these problems admit polynomial kernelizations, i.e., given such a problem our results either provide a polynomial kernelization, or they show that the problem does not admit a polynomial kernelization. These dichotomies are characterized by properties of the permitted constraints.

Dr. Edda KRONENWERTH

NATIONALITY: German
DEPARTMENT: Algorithms and Complexity

DISSERTATION TITLE:

Analyses of Evolutionary Algorithms

ABSTRACT OF PHD THESIS:

Evolutionary algorithms (EAs) are a highly successful tool commonly used in practice to solve algorithmic problems. This remarkable practical value, however, is not backed up by a deep theoretical understanding. Such an understanding would facilitate the application of EAs to further problems. Runtime analyses of EAs are one way to expand the theoretical knowledge in this field.

This thesis presents runtime analyses for three prominent problems in combinatorial optimization. Additionally, it provides probability theoretical tools that will simplify future runtime analyses of EAs.

The first problem considered is the Single Source Shortest Path problem. The task is to find in a weighted graph for a given source vertex shortest paths to all other vertices. Developing a new analysis method we can give tight bounds on the runtime of a previously designed and analyzed EA for this problem.

The second problem is the All-Pairs Shortest Path problem. Given a weighted graph, one has to find a shortest path for every pair of vertices in the graph. For this problem we show that adding a crossover operator to a natural EA using only mutation provably decreases the runtime. This is the first time that the usefulness of a crossover operator was shown for a combinatorial problem.

The third problem considered is the Sorting problem. For this problem, we design a new representation based on trees. We show that the EA naturally arising from this representation has a better runtime than previously analyzed EAs.

Dr. Evangelia PYRGA

NATIONALITY: Cypriot

DEPARTMENT: Algorithms and Complexity

DISSERTATION TITLE:

Algorithmic Game Theory and Networks

ABSTRACT OF PHD THESIS:

In this thesis we are studying three different problems that belong to the intersection of Game Theory and Computer Science. The first concerns the design of efficient protocols for a Contention Resolution problem regarding selfish users who all need to transmit information over a common single–access channel. We will provide efficient solutions for different variants of the problem, depending on the feedback that the users can receive from the channel. The second problem concerns the Price of Stability of a fair cost sharing Network Design problem for undirected graphs. We consider the general case for which the best known upper bound is the Harmonic number $H(n)$, where n is the number of players, and the best known lower bound is $12/7 \approx 1.778$. We improve the value of the previously best lower bound to $42/23 \approx 1.8261$. Furthermore, we study two and three players instances. Our upper bounds indicate a separation between the Price of Stability on undirected graphs and that on directed graphs, where $H(n)$ is tight. Previously, such a gap was only known for the cases where all players shared a terminal, and for weighted players. Finally, the last problem applies Game Theory as an evaluation tool for a computer system: we will employ the concept of Stochastic Stability from Evolutionary Game Theory as a measure for the efficiency of different queue policies that can be employed at an Internet router.

Dr. Oliver SCHALL

NATIONALITY: German
DEPARTMENT: Computer Graphics

DISSERTATION TITLE:

Robust and Efficient Processing Techniques for Static and Dynamic Geometric Data

ABSTRACT OF PHD THESIS:

Generating high quality geometric representations from real-world objects is a fundamental problem in computer graphics which is motivated by manifold applications. They comprise image synthesis for movie production or computer games but also industrial applications such as quality assurance in mechanical engineering, the preservation of cultural heritage and the medical adaptation of prostheses or orthoses. Common demands of these applications on their underlying algorithms are robustness and efficiency. In addition, technological improvements of scanning devices and cameras which allow for the acquisition of new data types such as dynamic geometric data, create novel requirements which rise new challenges for processing algorithms. This dissertation focuses on these aspects and presents different contributions for flexible, efficient and robust processing of static and time-varying geometric data. Two techniques focus on the problem of denoising. A statistical filtering algorithm for point cloud data building on non-parametric density estimation is introduced as well as a neighborhood filter for static and time-varying range data which is based on a novel non-local similarity measure. The third contribution unifies partition of unity decomposition and a global surface reconstruction algorithm based on the Fast Fourier Transform which results in a novel, robust and efficient reconstruction technique. Concluding, two flexible and versatile tools for designing scalar fields on meshes are presented which are useful to facilitate a controllable quadrangular remeshing.

Dr. Andreas SCHLICKER

NATIONALITY: German

DEPARTMENT: Computational Biology and Applied Algorithmics

DISSERTATION TITLE:

Ontology-based Similarity Measures and their Application in Bioinformatics

ABSTRACT OF PHD THESIS:

Genome-wide sequencing projects of many different organisms produce large numbers of sequences that are functionally characterized using experimental and bioinformatics methods. Following the development of the first bio-ontologies, knowledge of the functions of genes and proteins is increasingly made available in a standardized format. This allows for devising approaches that directly exploit functional information using semantic and functional similarity measures. This thesis addresses different aspects of the development and application of such similarity measures.

First, we analyze semantic and functional similarity measures and apply them for investigating the functional space in different taxa. Second, a new software program and a new database are described, which overcome limitations of existing tools and simplify the utilization of similarity measures for different applications.

Third, we delineate two applications of our functional similarity measures. We utilize them for analyzing domain and protein interaction datasets and derive thresholds for grouping predicted domain interactions into low- and high-confidence subsets. We also present the new MedSim method for prioritization of candidate disease genes, which is based on the observation that genes and proteins contributing to similar diseases are functionally related. We demonstrate that the MedSim method performs at least as well as more complex state-of-the-art methods and significantly outperforms current methods that also utilize functional annotation.

Dr. Thomas SCHULTZ

NATIONALITY: German
DEPARTMENT: Computer Graphics

DISSERTATION TITLE:

Feature Extraction for Visual Analysis of DW-MRI Data

ABSTRACT OF PHD THESIS:

Diffusion Weighted Magnetic Resonance Imaging (DW-MRI) is a recent modality to investigate the major neuronal pathways of the human brain. However, the rich DW-MRI datasets cannot be interpreted without proper preprocessing. In order to achieve understandable visualizations, this dissertation reduces the complex data to relevant features.

The first part is inspired by topological features in flow data. Novel features reconstruct fuzzy fiber bundle geometry from probabilistic tractography results. The topological properties of existing features that extract the skeleton of white matter tracts are clarified, and the core of regions with planar diffusion is visualized.

The second part builds on methods from computer vision. Relevant boundaries in the data are identified via regularized eigenvalue derivatives, and boundary information is used to segment anisotropy isosurfaces into meaningful regions. A higher-order structure tensor is shown to be an accurate descriptor of local structure in diffusion data.

The third part is concerned with fiber tracking. Streamline visualizations are improved by adding features from structural MRI in a way that emphasizes the relation between the two types of data, and the accuracy of streamlines in high angular resolution data is increased by modeling the estimation of crossing fiber bundles as a low-rank tensor approximation problem.

Dr. Pascal SCHWEITZER

NATIONALITY: **Luxembourgish/German**
DEPARTMENT: **Algorithms and Complexity**

DISSERTATION TITLE:

Problems of Unknown Complexity: Graph Isomorphism and Ramsey Theoretic Numbers

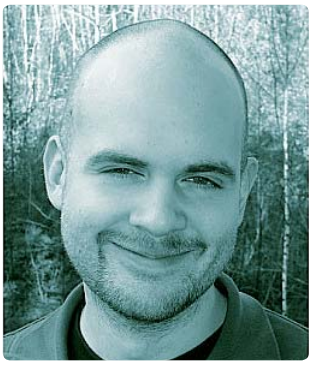
ABSTRACT OF PHD THESIS:

We consider three computational problems with unknown complexity status: The graph isomorphism problem, the problem of computing van der Waerden numbers and the problem of computing Ramsey numbers. For each of the problems, we devise an algorithm that we analyze with theoretical and practical means by a comparison with contemporary algorithms that solve the respective problems.

The ScrewBox algorithm solves the graph isomorphism problem by a random sampling process. Given two graphs, the algorithm randomly searches an invariant that may be randomly evaluated quickly and that shows significant statistical difference on the input graphs. This invariant is gradually and adaptively constructed depending on the difficulty of the input. Isomorphism is certified by supplying an isomorphism. Non-isomorphism is certified by the ScrewBox, the invariant whose statistical behavior deviates on the input graphs, together with the appropriate statistical test.

The wildcards algorithm for van der Waerden numbers solves the second problem. Its key technique is to treat colorings of integers avoiding monochromatic arithmetic progressions simultaneously by allowing ambiguity. This, together with a specific preprocessing step, forms the algorithm that is used to compute previously unknown van der Waerden numbers.

The wildcards algorithm for Ramsey numbers combines the techniques and algorithms with which we approach the first two problems to solve the third problem.



Dr. Carsten STOLL

NATIONALITY: German
DEPARTMENT: Computer Graphics

DISSERTATION TITLE:

Template Based Shape Processing

ABSTRACT OF PHD THESIS:

As computers can only represent and process discrete data, information gathered from the real world always has to be sampled. While it is nowadays possible to sample many signals accurately and thus generate high-quality reconstructions (for example of images and audio data), accurately and densely sampling 3D geometry is still a challenge. The signal samples may be corrupted by noise and outliers, and contain large holes due to occlusions. These issues become even more pronounced when also considering the temporal domain. Because of this, developing methods for accurate reconstruction of shapes from a sparse set of discrete data is an important aspect of the computer graphics processing pipeline.

In this thesis we propose novel approaches to including semantic knowledge into reconstruction processes using template based shape processing. We formulate shape reconstruction as a deformable template fitting process, where we try to fit a given template model to the sampled data. This approach allows us to present novel solutions to several fundamental problems in the area of shape reconstruction. We address static problems like constrained texture mapping and semantically meaningful hole-filling in surface reconstruction from 3D scans, temporal problems such as mesh based performance capture, and finally dynamic problems like the estimation of physically based material parameters of animated templates.

Dr. Josiane XAVIER PARREIRA

NATIONALITY: **Brazilian**

DEPARTMENT: **Databases and Information Systems**

DISSERTATION TITLE:

Decentralized Link Analysis in Peer-to-Peer Web Search Networks

ABSTRACT OF PHD THESIS:

Analyzing the authority or reputation of entities that are connected by a graph structure and ranking these entities is an important issue that arises in the Web, in Web 2.0 communities, and in other applications. The problem is typically addressed by computing the dominant eigenvector of a matrix that is suitably derived from the underlying graph, or by performing a full spectral decomposition of the matrix. Although such analyses could be performed by a centralized server, there are good reasons that suggest running theses computations in a decentralized manner across many peers, like scalability, privacy, censorship, etc. There exist a number of approaches for speeding up the analysis by partitioning the graph into disjoint fragments. However, such methods are not suitable for a peer-to-peer network, where overlap among the fragments might occur. In addition, peer-to-peer approaches need to consider network characteristics, such as peers unaware of other peers' contents, susceptibility to malicious attacks, and network dynamics (so-called churn).

In this thesis we make the following major contributions. We present JXP, a decentralized algorithm for computing authority scores of entities distributed in a peer-to-peer (P2P) network that allows peers to have overlapping content and requires no a priori knowledge of other peers' content. We also show the benefits of JXP in the Minerva distributed Web search engine. We present an extension of JXP, coined TrustJXP that contains a reputation model in order to deal with misbehaving peers. We present another extension of JXP that handles dynamics on peer-to-peer networks, as well as an algorithm for estimating the current number of entities in the network. This thesis also presents novel methods for embedding JXP in peer-to-peer networks and applications. We present an approach for creating links among peers, forming semantic overlay networks, where peers are free to decide which connections they create and which they want to avoid based on various usefulness estimators. We show how peer-to-peer applications, like the JXP algorithm, can greatly benefit from these additional semantic relations.

Dr. Hongbo ZHU

NATIONALITY: Chinese
DEPARTMENT: Computational Biology and Applied Algorithmics

DISSERTATION TITLE:

Characterization, Classification and Alignment of Protein-Protein Interfaces

ABSTRACT OF PHD THESIS:

Protein structural models provide essential information for the research on protein-protein interactions. In this dissertation, we describe two projects on the analysis of protein interactions using structural information. The focus of the first is to characterize and classify different types of interactions. We discriminate between biological obligate and biological non-obligate interactions, and crystal packing contacts. To this end, we defined six interface properties and used them to compare the three types of interactions in a hand-curated dataset. Based on the analysis, a classifier, named NOXclass, was constructed using a support vector machine algorithm in order to generate predictions of interaction types. NOXclass was tested on a non-redundant dataset of 243 protein-protein interactions and reaches an accuracy of 91.8 %. The program is beneficial for structural biologists for the interpretation of protein quaternary structures and to form hypotheses about the nature of protein-protein interactions when experimental data are yet unavailable.

In the second part of the dissertation, we present Galinter, a novel program for the geometrical comparison of protein-protein interfaces. The Galinter program aims at identifying similar patterns of different non-covalent interactions at interfaces. It is a graph-based approach optimized for aligning non-covalent interactions. A scoring scheme was developed for estimating the statistical significance of the alignments. We tested the Galinter method on a published dataset of interfaces. Galinter alignments agree with those delivered by methods based on interface residue comparison and backbone structure comparison. In addition, we applied Galinter on four medically relevant examples of protein mimicry. Our results are consistent with previous human-curated analysis. The Galinter program provides an intuitive method of comparative analysis and visualization of binding modes and may assist in the prediction of interaction partners, and the design and engineering of protein interactions and interaction inhibitors.

Current Students

Name	Supervisor	Department	Nationality
ALCARAZ MILLMAN, Nicolás	Jennifer Gerling	IMPRS-CS	*Mexican*
ARVANITOPOULOS-DARGINIS, Nikolaos	Jennifer Gerling	IMPRS-CS	*Greek*
BELET, Razvan	Jennifer Gerling	IMPRS-CS	*Romanian*
CIOLACU, Ines Lucia	Jennifer Gerling	IMPRS-CS	*Romanian*
CROITORU, Cosmina	Jennifer Gerling	IMPRS-CS	*Romanian*
DINU, Lavinia	Jennifer Gerling	IMPRS-CS	*Romanian*
EBRAHIMI, Mahdi	Jennifer Gerling	IMPRS-CS	*Iranian*
GONCHAROV, Ilya	Jennifer Gerling	IMPRS-CS	*Russian*
IQBAL, Javeria	Jennifer Gerling	IMPRS-CS	*Pakistani*
IVANOVA, Violeta	Jennifer Gerling	IMPRS-CS	*Bulgarian*
KARGIN, Yagiz	Jennifer Gerling	IMPRS-CS	*Turkish*
KHOSHNEVIS, Hossein	Jennifer Gerling	IMPRS-CS	*Iranian*
KUZEY, Erdal	Jennifer Gerling	IMPRS-CS	*Turkish*
LEVINKOV, Evgeny	Jennifer Gerling	IMPRS-CS	*Russian*
MAHMOUD ABDEL MONEIM, Dina	Jennifer Gerling	IMPRS-CS	*Egyptian*
MALINOWSKI, Mateusz	Jennifer Gerling	IMPRS-CS	*Polish*
PARKS, Justin	Jennifer Gerling	IMPRS-CS	*US-American*
PIRMORADIAN, Ebad	Jennifer Gerling	IMPRS-CS	*Iranian*
PRYTKOVA, Natalia	Jennifer Gerling	IMPRS-CS	*Russian*
SALIBA, Bahjat	Jennifer Gerling	IMPRS-CS	*Syrian*
SIMONOVSKY, Martin	Jennifer Gerling	IMPRS-CS	*Czech*
TANDON, Niket	Jennifer Gerling	IMPRS-CS	*Indian*
TEFLIOUDI, Christina	Jennifer Gerling	IMPRS-CS	*Greek*
TERIS, Liviu	Jennifer Gerling	IMPRS-CS	*Romanian*
TRAN, Anh Tuan	Jennifer Gerling	IMPRS-CS	*Vietnamese*
VASUDEVAN, Aravind	Jennifer Gerling	IMPRS-CS	*Indian*
VENKATACHALAPATHY, Mahendiran	Jennifer Gerling	IMPRS-CS	*Indian*
VIDRIALES ESCOBAR, Mónica Paola	Jennifer Gerling	IMPRS-CS	*Mexican*
ZHENG, Qinqing	Jennifer Gerling	IMPRS-CS	*Chinese*

Name	Supervisor	Department	Nationality
ABDEL MAKSOUD, Mohamed	Reinhard Wilhelm Gerhard Weikum	Programming Languages and Compiler Construction	*Egyptian*
ABED, Fidaa	Kurt Mehlhorn	Algorithms and Complexity	*Palestinian*
ALEXA, Adrian	Thomas Lengauer	Computational Biology and Applied Algorithmics	*Romanian*
ALVANAKI, Foteini	Sebastian Michel	Databases and Information Systems	*Greek*
ANAND, Avishek	Gerhard Weikum	Databases and Information Systems	*Indian*
ASSENOV, Yassen	Thomas Lengauer	Computational Biology and Applied Algorithmics	*Bulgarian*
AWADALLAH, Rawia	Gerhard Weikum	Databases and Information Systems	*Palestinian*
BAAK, Andreas	Meinard Müller	Computer Graphics	*German*
BEGGEL, Bastian	Thomas Lengauer	Computational Biology and Applied Algorithmics	*German*
BERNER, Alexander	Michael Wand	Computer Graphics	*German*
BLANKENBURG, Hagen	Thomas Lengauer	Computational Biology and Applied Algorithmics	*German*
BOGOJESKA, Jasmina	Thomas Lengauer	Computational Biology and Applied Algorithmics	*Macedonian*
BOKELOH, Martin	Michael Wand	Computer Graphics	*German*
BOZEK, Kasia	Thomas Lengauer	Computational Biology and Applied Algorithmics	*Polish*
BROSCHART, Andreas	Ralf Schenkel Gerhard Weikum	Databases and Information Systems	*German*
BYELOZYOROV, Sergiy	Philipp Slusallek Thorsten Thormählen	Computer Graphics	*Ukrainian*
CHUONG, Nguyen	Karol Myszkowski	Computer Graphics	*Vietnamese*

Name	Supervisor	Department	Nationality
CIOBOTARU, Oana	Michael Backes	Information Security and Cryptography	*Romanian*
CRECELIUS, Tom	Gerhard Weikum Ralf Schenkel	Databases and Information Systems	*German*
DANILEWSKI, Piotr	Philipp Slusallek Christian Theobalt	Computer Graphics	*Polish*
DENEV, Dimitar	Gerhard Weikum	Databases and Information Systems	*Bulgarian*
DIDYK, Piotr	Karol Myszkowski	Computer Graphics	*Polish*
DIETZ, Laura	Gerhard Weikum Tobias Scheffer	Databases and Information Systems	*German*
DIETZEN, Matthias	Thomas Lengauer	Computational Biology and Applied Algorithmics	*German*
DIMITROVA, Rayna	Bernd Finkbeiner Gerhard Weikum	Reactive Systems	*Bulgarian*
DONG, Zhao	Hans-Peter Seidel	Computer Graphics	*Chinese*
DYLLA, Maximilian	Martin Theobald	Databases and Information Systems	*German*
EBERT, Sandra	Bernt Schiele	Computer Vision and Multimodal Computing	*German*
ELBASSUONI, Shady	Gerhard Weikum	Databases and Information Systems	*Egyptian*
EMELIYANENKO, Pavel	Michael Sagraloff	Algorithms and Complexity	*Russian*
EMIG, Dorothea	Mario Albrecht	Computational Biology and Applied Algorithmics	*German*
FEUERBACH, Lars	Thomas Lengauer	Computational Biology and Applied Algorithmics	*German*
FIETZKE, Arnaud Luc	Christoph Weidenbach	Automation of Logic	*German*
GASSE, Francis	Viorica Sofronie-Stokkermans	Automation of Logic	*Canadian*

Name	Supervisor	Department	Nationality
GEORGIEV, Iliyan	Philipp Slusallek Bodo Rosenhahn	Computer Graphics	*Bulgarian*
GERLING, Sebastian	Michael Backes	Information Security and Cryptography	*German*
GRANADOS VELAZQUEZ, Miguel Andrés	Christian Theobalt	Computer Graphics	*Colombian*
GROSCHE, Peter	Meinard Müller	Computer Graphics	*German*
GROSS, Florian	Andreas Zeller Gerhard Weikum	Software Engineering	*German*
HAGEMANN, Willem	Uwe Waldmann	Automation of Logic	*German*
HALACHEV, Konstantin	Thomas Lengauer	Computational Biology and Applied Algorithmics	*Bulgarian*
HATEFIARDAKANI, Hassan	Holger Hermanns Christoph Weidenbach	Dependable Systems and Software	*Iranian*
HELTEN, Thomas	Meinard Müller	Computer Graphics	*German*
HOFFART, Johannes	Gerhard Weikum	Databases and Information Systems	*German*
HRITCU, Catalin	Michael Backes	Information Security and Cryptography	*Romanian*
JAIN, Arjun	Thorsten Thormählen Hans-Peter Seidel	Computer Graphics	*Indian*
JURKIEWICZ, Tomasz	Kurt Mehlhorn	Algorithms and Complexity	*Polish*
KALOJANOV, Javor	Philipp Slusallek Michael Wand	Computer Graphics	*Bulgarian*
KERBER, Jens	Michael Wand	Computer Graphics	*German*
KHATTAB, Dina	Christian Theobalt	Computer Graphics	*Egyptian*
KHOSLA, Megha	Kurt Mehlhorn	Algorithms and Complexity	*Indian*
KONDREDDI, Sarath Kumar	Gerhard Weikum	Databases and Information Systems	*Indian*

Name	Supervisor	Department	Nationality
Konietzny, Sebastian	Alice McHardy	Computational Genomics and Epidemiology	*German*
KONZ, Verena	Meinard Müller	Computer Graphics	*German*
KURZ, Christian	Thorsten Thormählen	Computer Graphics	*Austrian*
LU, Tianxiang	Christoph Weidenbach	Automation of Logic	*Chinese*
MAKARI MANSHADI, Faraz	Mauro Sozio	Databases and Information Systems	*German*
MANJUNATH, Madhusudan	Kurt Mehlhorn	Algorithms and Complexity	*Indian*
MARSALEK, Lukas	Philipp Slusallek Thorsten Thormählen	Computer Graphics	*Czech*
METZGER, Steffen	Ralf Schenkel Gerhard Weikum	Databases and Information Systems	*German*
MILEVA, Yana	Andreas Zeller Gerhard Weikum	Software Engineering	*Bulgarian*
NAKASHOLE, Ndapandula	Gerhard Weikum	Databases and Information Systems	*Namibian*
PAN, Hanglin	Gerhard Weikum	Databases and Information Systems	*Chinese*
PATIL, Kaustubh	Alice McHardy	Computational Genomics and Epidemiology	*Indian*
PETER, Hans-Jörg	Bernd Finkbeiner Christoph Weidenbach	Reactive Systems	*German*
PISHCHULIN, Leonid	Bernt Schiele	Computer Vision and Multimodal Computing	*Russian*
POPOV, Stefan	Philipp Slusallek Hans-Peter Seidel	Computer Graphics	*Bulgarian*
POURMIRI, Ali	Kurt Mehlhorn	Algorithms and Complexity	*Iranian*
QU, Lizhen	Gerhard Weikum	Databases and Information Systems	*Chinese*

Name	Supervisor	Department	Nationality
RAMEZANI, Fahimeh	Kurt Mehlhorn	Algorithms and Complexity	*Iranian*
RAMIREZ, Fidel	Mario Albrecht	Computational Biology and Applied Algorithmics	*Colombian*
RANGAPURAM, Syama Sundar Yadav	Matthias Hein Kurt Mehlhorn	Machine Learning	*Indian*
RAUF, Imran	Khaled Elbassioni Kurt Mehlhorn	Algorithms and Complexity	*Pakistani*
RESHETOUSKI, Ilya	Ivo Ihrke	Computer Graphics	*Byelorussian*
RIZKALLAH, Christine	Rob van Stee	Algorithms and Complexity	*Canadian*
ROHRBACH, Marcus	Bernt Schiele	Computer Vision and Multimodal Computing	*German*
RÖTTGER, Richard	Jan Baumbach	Computational Biology and Applied Algorithmics	*German*
SCHELHORN, Sven Eric	Mario Albrecht	Computational Biology and Applied Algorithmics	*German*
SCHERBAUM, Kristina	Thorsten Thormählen	Computer Graphics	*German*
SEUFERT, Stephan	Gerhard Weikum	Databases and Information Systems	*German*
SHAHEEN, Mohammed	Robert Strzodka Hans-Peter Seidel	Computer Graphics	*Palestinian*
STEINBRÜCK, Lars	Alice McHardy	Computational Genomics and Epidemiology	*German*
STUPAR, Aleksandr	Sebastian Michel Ralf Schenkel	Databases and Information Systems	*Serbian*
SUDA, Martin	Christoph Weidenbach	Automation of Logic	*Czech*
TANEVA, Bilyana	Gerhard Weikum	Databases and Information Systems	*Bulgarian*
TEVS, Art	Christian Theobalt	Computer Graphics	*German*

Name	Supervisor	Department	Nationality
TOLOSI, Laura	Thomas Lengauer	Computational Biology and Applied Algorithmics	*Romanian*
TUSCHE, Christina	Alice McHardy	Computational Genomics and Epidemiology	*German*
TYLENDA, Tomasz	Gerhard Weikum	Databases and Information Systems	*Polish*
VAN DER WEELE, Vincent	Rob van Stee	Algorithms and Complexity	*Dutch*
WAGNER, Markus	Frank Neumann	Algorithms and Complexity	*German*
WANG, Yafang	Gerhard Weikum	Databases and Information Systems	*Chinese*
WINZEN, Carola	Benjamin Doerr	Algorithms and Complexity	*German*
WISCHNEWSKI, Patrick	Christoph Weidenbach	Automation of Logic	*German*
WU, Chenglai	Christian Theobalt	Computer Graphics	*Chinese*
YAHYA, Mohamed	Gerhard Weikum	Databases and Information Systems	*Palestinian*
YOSEF, Mohamed Amir	Gerhard Weikum	Databases and Information Systems	*Egyptian*
ZIMMER, Henning Lars	Joachim Weickert Hans-Peter Seidel	Mathematical Image Analysis	*German*

Staff

Krista AMES

NATIONALITY: US-American
POSITION: German and English Teacher
E-MAIL: ames@mpi-inf.mpg.de

EDUCATION: German
UNIVERSITY: Willamette University, Salem, OR; LMU München

Patrick DEMPSEY

NATIONALITY: Irish
POSITION: English Teacher
E-MAIL: dempsey@mpi-inf.mpg.de

EDUCATION: Comparative Literature & Cinema studies
UNIVERSITY: San Francisco State University

Irena DOTCHEVA

NATIONALITY: Bulgarian
POSITION: English Teacher
E-MAIL: irena@mpi-inf.mpg.de

EDUCATION: English Linguistics & Computational Linguistics
UNIVERSITY: Saarland University

Aikaterini GEORGIOU

NATIONALITY: Greek
POSITION: German Teacher
E-MAIL: georgiou@mpi-inf.mpg.de

EDUCATION: Translating, German as a Foreign Language
UNIVERSITY: Ionian University, Saarland University

Jennifer GERLING

NATIONALITY: German
POSITION: IMPRS-CS Coordinator
E-MAIL: jgerling@mpi-inf.mpg.de

EDUCATION: Applied Linguistics and Translating (Spanish, English, and German)
UNIVERSITY: Saarland University

Olga GRIGORIEVA

NATIONALITY: Moldavian
POSITION: German Teacher
E-MAIL: ogrigori@mpi-inf.mpg.de

EDUCATION: German, as a Foreign Language
UNIVERSITY: Alecu Russo State University of Bălţi, Saarland University

Stephanie JÖRG

NATIONALITY: German
POSITION: Assistant to the Coordinator
E-MAIL: sjoerg@mpi-inf.mpg.de

EDUCATION: Translator for English and Spanish
UNIVERSITY: Saarland University

Heike PRZYBYL

NATIONALITY: German
POSITION: Student Assistant
E-MAIL: hprzybyl@mpi-inf.mpg.de

EDUCATION: Translating / Interpreting for English and French
UNIVERSITY: Saarland University

Katja SCHÖNBORN

NATIONALITY: German
POSITION: Webmaster
E-MAIL: schoenb@mpi-inf.mpg.de

EDUCATION: Sociology, Psychology, Political Science, English
UNIVERSITY: Saarland University

Elena TREGUBOVA

NATIONALITY: Russian
POSITION: German Teacher
E-MAIL: etregubo@mpi-inf.mpg.de

EDUCATION: Philology, German, Translating/Interpreting, Education
UNIVERSITY: Voronezh State University; Voronezh State Technical University

Prof. Dr.-Ing. Gerhard WEIKUM

NATIONALITY: German
POSITION: Spokesperson
E-MAIL: weikum@mpi-inf.mpg.de

EDUCATION: Computer Science
UNIVERSITY: University of Darmstadt

Awards

AWARDS

In 2009 and 2010, the following IMPRS students received awards for their work:

03 Dec 2010	*Richard* Röttger	Siemens Awards (for Master's Thesis)
22 Nov 2010	*Sebastian* Gerling	Förderpreis im Rahmen des GDD-Wissenschaftspreises (for Master's Thesis)
03 Nov 2010	*Gerard* de Melo	CIKM Best Paper Award 2010
27 Oct 2010	*Carola* Winzen	Google Europe Doctoral Fellowship in Randomized Algorithms
21 Oct 2010	*Christoph* Bock	Eduard Martin Preis 2010 (for PhD Thesis)
16 Jun 2010	*Fabian* Suchanek	Otto-Hahn-Medaille 2009 (for PhD Thesis)
16 Jun 2010	*Thomas* Schultz	Otto-Hahn-Medaille 2009 (for PhD Thesis)
29 Jan 2010	*Fabian* Müller	Günter-Hotz-Medaille 2010 (for Master's Thesis)
11 Jun 2009	*Christoph* Bock	Otto-Hahn Medaille 2008 (for PhD Thesis)
11 Jun 2009	*Yana* Mileva	2009 Google Anita Borg Scholarship Finalist
06 Mar 2009	*Sebastian* Michel	GI DBIS Dissertationspreis 2009 (for PhD Thesis)
01 Feb 2009	*Dorothea* Emig	Travel grant from the Boehringer Ingelheim Fonds, Foundation for Basic Research in Medicine, funding the research stay at UC Berkeley, San Francisco, CA, USA, February – April 2009

Publisher

International Max Planck Research School
for Computer Science

International Max Planck Research School
for Computer Science
Max Planck Institute for Informatics
Campus E1 4 | 66123 Saarbrücken | Germany
Telefon +49 681 9325 1800
Fax +49 681 9325 1899
E-mail imprs@mpi-inf.mpg.de
Internet www.imprs-cs.de

www.imprs-cs.de

max planck institut informatik

Max Planck Institute for Software Systems

SAARLAND UNIVERSITY
COMPUTER SCIENCE