

PhD Application Talks: Feb 16th, 2010 in MPII - room 024

Deutsche
Forschungsgemeinschaft

DFG

International Max Planck Research School
for Computer Science



1st session 8:50 - 10:40	<ul style="list-style-type: none">• <i>Construction of Virtual Worlds with Web 2.0 technology</i> Sergiy Byelozyorov• <i>On Linear Programming Relaxations of Hypergraph Matching</i> Yuk Hei Chan• <i>Fast Generalized Belief Propagation for Binocular Stereo Matching</i> Zhongjie Wang• <i>Bandwidth Minimization for Caterpillar Graph</i> Kashyap Dixit• <i>Syntactic analysis of ambiguous natural language constructions with Tomita shift-resolve parser</i> Alexander Kobzar
<i>Coffee break</i>	
2nd session 11:10 - 12:50	<ul style="list-style-type: none">• <i>Efficient Itemset Generator Discovery over a Stream Sliding Window</i> Chuancong Gao• <i>A Generic Top-Down Dynamic-Programming Approach to Prefix-Free Coding</i> Xiaoming Xu• <i>Fast Distributed Replication Algorithm in Modern Networks</i> Faraz Makari Manshadi• <i>An Inherently Stabilizing Distributed Algorithm for All Node-Disjoint Paths in Exchanged Hypercube</i> Thamer Alsulaiman• <i>Distance Measurement Systems and Their Applications Based on Image Understanding</i> Yin-Yu Lu
<i>Lunch break</i>	
3rd session 1:40 - 3:00	<ul style="list-style-type: none">• <i>Generalized Second Price Auction with Markovian Users</i> Jiajin Yu• <i>Graph sparsification for balanced partitioning</i> Syama Sundar Rangapuram• <i>Testing a Verification Environment</i> Markus Wagner• <i>Mining and Characterizing Scientific Communities in DBLP</i> Cailing Dong
<i>Coffee break</i>	

Sergiy Byelozyorov

University: Saarland University
Supervisor/Ref: Prof. Dr. Philipp Sluasallek

Title: Construction of Virtual Worlds with Web 2.0 technology

Abstract: Current Web technologies allow developers to create rich Web-applications. Unlike desktop applications Web 2.0 allows to create programs 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-applications similar to Second Life. I have wrapped 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 fully-featured 3D world application right inside of the browser.

Yuk Hei Chan

University: The Chinese University of Hong Kong
Supervisor/Ref: Lap Chi Lau / Dr. Nikhil Bansal

Title: On Linear Programming Relaxations of Hypergraph Matching

Abstract: A hypergraph is a generalization of a graph where each hyperedge can contain an arbitrary number of vertices. The hypergraph matching problem is to find a largest collection of disjoint hyperedges. While matching on general graphs is polynomial time solvable, hypergraph matching is NP-hard and there is no good approximation algorithm for the problem in its most general form. We study the restricted case where every hyperedge consists of k vertices, known as k -Set Packing.

We show that the integrality gap (the maximum ratio of the fractional solution to the integral solution over all hypergraphs) is $k-1+1/k$, and that the integrality gap analysis is tight in general. The proof is by showing a fractional colouring with a small number of colours. The colouring is done in a greedy manner with the help of a good ordering of hyperedges, which is derived from the LP solution.

As a special case, we prove that by adding a Fano plane constraint which deals with a set of intersecting hyperedges, the integrality gap of the LP for unweighted 3-Set Packing can be improved from $7/3$ to 2 . The result is by detailed analysis on the structure of a counterexample H , where we show that there is no Fano plane in H and by a result of Furedi [F81], we conclude that the integrality gap is 2 .

When the vertex set is partitioned into k sets so that each hyperedge contains one vertex from each set, we have the k -Dimensional Matching problem. In this case we show that the standard LP relaxation has an integrality gap of $k-1$. We remark that integrality gap analysis had been made by Furedi, Kahn and Seymour [FKS93] in a more general form using a different approach, but their analysis does not directly yield an approximation algorithm. We obtain a $(k-1)$ -approximation algorithm by combining the good ordering on the hyperedges with an algorithmic framework called local-ratio. This improves the previous result for 3-Dimensional Matching from $2+\epsilon$ to 2 .

Zhongjie Wang

University: Zhejiang University of Technology
Supervisor/Ref: Shengyong Chen, PhD

Title: Fast Generalized Belief Propagation for Binocular Stereo Matching

Abstract: This project works on algorithmic research for the well-known binocular stereo matching problem, especially develops a novel method called “fast generalized belief propagation”. Stereo matching is one of the most challenging topics in computer vision. It has been comprehensively studied for several decades, but the results are still way from accuracy and efficiency. This project contributes mainly in two aspects to improve the accuracy the result of stereo matching.

On one hand, generalized belief propagation (GBP) is highly adapted and successfully introduced into stereo matching problem. GBP is a region based belief propagation algorithm which features a good convergence. This paper proposes an idea of min-sum scheme in GBP to replace the traditional max-product messaging scheme and applies it to solve the stereo matching problem. A caching technique is used to improve the access efficiency. Furthermore, two strategies are found to speed up GBP processing. One is the direction set method which is used to reduce the complexity of computing clique messages from quartic to cubic. The other is hierarchical state space reduction which is proposed to decrease redundant labels in every hierarchical level. Combining these strategies can greatly increase the processing speed. Beside the paper mainly develops the GBP method itself to a new stage, it is also the first attempt to apply GBP for solving the stereo matching problem. Experiments show that the proposed algorithm can speed up by 80+ times for typical stereo images and lead to a more accurate result than the canonical GBP. Furthermore, considering the potential parallel computing capability which is an advantage of such message passing based algorithms, the speed can be further increased even to real-time.

On the other hand, this paper combines the existing global based and local based methods to strengthen their own advantages and weaken their disadvantages. The global based approach mostly makes the result too smooth to include more details, while the local based method makes the result noisier. This paper proposes a novel method which combines the global and local traits as much as possible to fill the gap between them. Firstly, layered result is obtained by graph cut based algorithms. Secondly, several sub-pixel local results are obtained under global result’s guidance with window correlation method. Thirdly, these local results are repeatedly merged in the Markov Random Fields (MRFs) model with a second order smooth prior energy function. Experiments show that this method can achieve better results than that of either global or local based method individually.

Kashyap Dixit

University: Indian Institute of Technology, Kanpur
Supervisor/Ref: Prof. Shashank K. Mehta

Title: Bandwidth minimization for Caterpillar graph

Abstract: Bandwidth Minimization Problem (BMP) is to find distinct labels for the vertices of a given graph so that the maximum difference between the end vertices of any edge is minimum. This problem has applications in numerical analysis and hardware architecture. BMP on caterpillar graph is of particular interest then since it is related to the multi-processor scheduling

In this work, we analysed the bandwidth minimization for the caterpillar graph. We define the constrained bush bandwidth problem and proved that it is NP-hard. Later, we proposed the labeling scheme which gives the best possible approximation result of additive factor 1. Then we defined the Shortest Distance Root Label finding problem and proposed a polynomial time exact algorithm to solve it.

We also analyzed the algorithm proposed by Haralabidas et. al. (HMM) for labeling caterpillar graphs and proposed an algorithm which can be proved to perform at least as good as HMM algorithm but gave constant approximation of 2 for the proposed tight examples (respectively $\log n$).

Alexander Kobzar

University: Kharkiv National University of Radioelectronics
Supervisor/Ref: Natalya Valenda

Title: Syntactic analysis of ambiguous natural language constructions with Tomita shift-resolve parser

Abstract: Most context-free (CF) grammar parsing algorithms developed up to now are inapplicable to natural language (NL) parsing due to its narrow coverage of CF-grammars describing NL syntax or unpractical time complexity. In this presentation we consider the design of such an algorithm directed by Russian language as well as a scheme of formalization of Russian constructions. The algorithm is based on a combination of shift-resolve parsing and Tomita parser, and its design is divided into two phases:

1. Design of a deterministic algorithm. NL constructions and its formal representations by recursive and epsilon-grammars that make most parsing algorithms unable to process them are considered. An improvement of shift-resolve parsing algorithm is carried out with respect to the ambiguity.

2. Design of a nondeterministic algorithm. Properties of shift-resolve parsing which are different to LR-parsing and formal conditions of applying of Tomita parser to shift-resolve parsing are considered. Designed parser is capable to carry out parsing with better accuracy and time complexity compared to LR-parsing implementation. Next, we consider formalization of Russian word combinations and simple sentences by attribute CF-grammar. Designed scheme and grammar make it possible to formalize these constructions completely. Next, we consider a developed NL parser that implement designed algorithm and make it possible to design NL grammars. Finally, we consider some open problems, namely grammar classes that are remain unhandled for the algorithm.

Chuancong Gao

University: Tsinghua University , Beijing
Supervisor/Ref: Jianyong Wang

Title: Efficient Itemset Generator Discovery over a Stream Sliding Window

Abstract: Mining generator patterns has raised great research interest in recent years. The main purpose of mining itemset generators is that they can form equivalence classes together with closed itemsets, and can be used to generate simple classification rules according to the MDL principle. In this paper, we devise an efficient algorithm called StreamGen to mine frequent itemset generators over a stream sliding window. We adopt a novel enumeration tree structure to help keep the information of mined generators and the border between generators and non-generators, and propose some optimization techniques to speed up the mining process. We further extend the algorithm to directly mine a set of high quality classification rules over stream sliding windows while keeping high performance. The extensive performance study shows that our algorithm outperforms other state-of-the-art algorithms which perform similar tasks in terms of both runtime and memory usage efficiency, and has high utility in terms of classification.

Xiaoming Xu

University: Fudan University, Shanghai
Supervisor/Ref: Prof. Dr. Rudolf Fleischer

Title: A Generic Top-Down Dynamic-Programming Approach to Prefix-Free Coding

Abstract: Given a probability distribution over a set of n words to be transmitted, the Huffman Coding problem is to find a minimal-cost prefix free code for transmitting those words. The basic Huffman coding problem can be solved in $O(n \log n)$ time but variations are more difficult. One of the standard techniques for solving these variations utilizes a top-down dynamic programming approach. In this paper we show that this approach is amenable to dynamic programming speedup techniques, permitting a speedup of an order of magnitude for many algorithms in the literature for such variations as mixed radix, reserved length and one-ended coding. These speedups are immediate implications of a general structural property that permits batching together the calculation of many DP entries.

Faraz Makari Manshadi

University: Saarland University
Supervisor/Ref: Dr. Mauro Sozio

Title: Fast Distributed Replication in Modern Networks

Abstract: 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 such that at each peer some minimum threshold for the probability of successful search is satisfied while minimizing the total number of replicated data items in network. 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 and the challenging point is to optimize this global function only based on the local information. The literature contains huge amount of theoretical work; Most of the proposed algorithms are, due to their complexity, far from being practically applicable or based on some simplifying assumptions, e.g. unbounded message sizes.

In this thesis, we present a fast self-stabilizing distributed algorithm for replication based on the LP relaxation technique, which provides a near-optimal solution and uses only small message. The algorithm can also be applied in dynamic settings where nodes join and leave the system. These properties are very desirable in distributed systems with unreliable components. The most appealing features of our algorithm are its simplicity and polylogarithmic convergence. We also provide some experimental results of the performance of the algorithm on different network topologies such as random graphs and internet-style topologies, and introduce some application areas of our algorithm, while proving the theoretical guarantees.

Thamer Alsulaiman

University: Kuwait University
Supervisor/Ref: Dr. Mehmet Hakan Karaata

Title: An Inherently Stabilizing Distributed Algorithm for All Node-Disjoint Paths in Exchanged Hypercube

Abstract: Two paths between a source and a destination nodes in a network are node-disjoint if they do not share any node except the endpoints. Node-disjoint paths have numerous applications in distributed systems including ways to deal with lost, damaged or altered messages during delivery. For instance, sending copies of messages over several disjoint paths increases the reliability of communication in systems prone to message omission. Moreover, data can be split into several shares and sent over disjoint paths to increase network throughput. Many topologies such as hypercube, star networks, and their variants have been proposed providing multiple disjoint paths between a pair of endpoints. Exchanged hypercube is a new topology that is obtained by systematically removing edges from a binary hypercube. Exchanged hypercube topology aims to increase scalability while reducing the number of edges, i.e., reducing relative network cost. This paper presents an inherently stabilizing distributed algorithm to find all node-disjoint paths in an exchanged hypercube network. Due to being inherently stabilizing, the system is immune to transient faults by construction

Yin-Yu Lu

University: National Central University, Taiwan
Supervisor/Ref: Kuo-Chin Fan, PhD

Title: Distance Measurement Systems and Their Applications Based on Image Understanding

Abstract: When taking images with a camera, light from a physical scene projects on the CCD (Charge-coupled device), where the light pulses are transformed into digital signal to produce 2-dimensional images. During this transformation, one dimension of the data is lost. It is extremely difficult to retrieve the lost information. To address this problem, there are several approaches available to deal with the depth retrieval problem, among which basic image formation geometry and stereo vision are the two major approaches. In the first approach, the depth between CCD and object can be calculated by the focus length and the physical dimension of the grids of CCD pixels, which are tinny in size and hard to measure. As a result, measuring error of the system is generally unconquerable because of the micrometer scale of the grids of CCD pixels. The stereo vision based methods, on the other hand, provide better measuring accuracy, but difficulties arise to solve the matching problem, which is the foundation of these methods.

It is found that investigation on the depth of the image is not highly regarded in the realm of image processing and pattern recognition. As an attempt to solve this problem, I have researched and obtained several image understanding based approaches for solving the depth retrieval problems. The main difference between the proposed systems and the abovementioned (existing) systems is the mechanism structure adopted and information from image understanding. There are three types of depth retrieval systems that I have developed so far, including: 1. Distance measurement based on laser-projected CCD Images, 2. Distance measurement based on arbitrarily designated two points in 3-D space, and 3. Distance measurement based on pixel variation of CCD images due to camera movement. These three types of depth retrieval systems serve different purposes and occasions. Type one is generally used in three-dimensional localization and long-distance landslide surveillance. Type two has advantages for use in intelligent transportation systems, where I have applied the method on smart vehicles by designating taillights as the main feature for distance measuring. Type three could provide great capabilities for 3-D space reconstruction and robot navigation. The proposed methods might provide new applications for digital image processing.

Jiajin Yu

University: Fudan University, Shanghai
Supervisor/Ref: Prof. Dr. Rudolf Fleischer

Title: Generalized Second Price Auction with Markovian Users

Abstract: Sponsored search auction is used by most search engines to find ads showing on the web page of search results. The income of this targeted advertising business is a big part of the revenue of most search engines. The most widely used approach is called Generalized Second Price (GSP) auction. Most previous works about GSP auction are based on the separation assumption: the probability a user will click on an ad is composed by two independent parts: a quality factor of the ad itself and a position factor of the slot of the ad. This previous model does not include the externality a higher ad may bring to the ads below it.

We study a GSP auction in a Markovian user model where the externality is included by modeling a user's probability behavior of scanning ad list. In particular, we propose a new ranking scheme for the bidders. We prove Nash equilibrium always exists in the auction and study the efficiency of the auction by theoretical analysis and simulation. We compare our results with social optimum and previous approaches. Comparison shows that our results approximate the social optimum and improve previous approaches under various circumstances.

Syama Sundar Rangapuram

University: Saarland University
Supervisor/Ref: Prof. Dr. Matthias Hein

Title: Graph sparsification for balanced partitioning

Abstract: Balanced graph partitioning is the problem of finding an optimal partitioning of a graph according to the specified balance criteria. It is known that balanced partitioning is an NP-hard problem and hence we have to rely on relaxations or heuristics if the problem size is large. The main motivation of this thesis is to solve partitioning problems on large graphs by designing new sparsification techniques. Here, we propose a coarsening scheme that iteratively reduces the problem size by collapsing vertices and edges until the graph becomes relatively small so that the partitioning problem can be solved more satisfactorily with less computation. The important contribution in this work would be provision of certificate of guarantee that the coarsening step does not disturb the original optimal partitioning; this means that the optimal partitioning of the coarsest graph directly transfers to the optimal partitioning of the original problem. Our method tries to achieve this guarantee by deriving a lower bound on the optimal cut that goes through a given subgraph, before claiming that the subgraph is safe to be coarsened. To find this lower bound, we formulate a modified graph cut problem using spectral relaxation: given a graph G , and a subgraph, S , find the optimal partitioning of G such that the optimal cut goes through S . If we find any other partitioning of G that does not cut S , and whose cut value is smaller than the above bound, then clearly all vertices in S can be collapsed. The modified optimization problem that we derived is non-linear, non-convex and we are currently working on the theoretical proof that we converge to the global optimum. Experimental results suggest that we can solve this problem globally, thus allowing us to provide the guarantee certificate.

Markus Wagner

University: University of Koblenz -Landau
Supervisor/Ref: Dr. Walter D. Potter

Title: Testing a Verification Environment

Abstract: The goal of software verification is the proof of correctness of a program with respect to its specification. Normally, the tools used in a verification process are not verified, and their correctness is taken for granted. One approach to check the quality of these tools without verifying them is by testing.

In the BMBF-project "Verisoft XT", the Verifying C Compiler (VCC) from Microsoft Research is used for the verification of C code. It is impossible to come up with a correctness proof of VCC's sound functioning and to keep it up-to-date because of two reasons. First, VCC is being developed during the project, which results in frequent changes of the implementation and the verification formalism. Second, the compiler is used as a black box for nonexperts.

The goal of this thesis is to develop a systematic testing process for black-box verification systems. The theoretical part consists of the application of the concepts of black-box software testing in the context of verification systems. This includes the definition of the subject under test and the definition of a suitable test metric. This process is applied to VCC, resulting in a specialized test suite and a test harness. The test harness will run and monitor the tests without user interaction, enabling the regression testing of both VCC and an arbitrary code base.

Cailing Dong

University: Shandong University
Supervisor/Ref: Prof. Dr. Yilong Yin

Title: Mining and Characterizing Scientific Communities in DBLP

Abstract: As a prominent computer science bibliographical database, DBLP (Digital Bibliography & Library Project) has been widely used in community mining. Among the huge amount of structured data that it provides, there are many communities defined by various relationships. In this work, we focus on conferences in DBLP and analyze communities in computer science in two levels. At the first level we construct communities around specific conferences based on either common topics or common authors. Our findings show that the authors provide a better ground for characterizing communities than topics alone. At the second level we decompose the computer science area into 14 research fields based on previous researches and study various properties of the corresponding communities. We characterize each research community in terms of publication growth rate, collaboration trends and population stability. All these features influence each other to some extent and together show the difference in activity rate and interdisciplinarity of each subarea.