Research I Foundation

Annual Report

2011-2012

Contents

Expenditure Sheet for 2011-2012

Expenditure Sheet for April 2011- March 2012

Expenditure incurred towards Travel Allowance	Rs. 1305310
Expenditure incurred towards Salary	Rs. 1884597
Expenditure incurred towards Honorarium	Rs. 70000
Expenditure incurred towards Equipments	Rs. 7800
Expenditure incurred towards Contingency	Rs. 286796
Total Expenditure incurred	**Rs. 3554485**

List of Conference Papers presented

List of Conference Papers presented with Travel Grant awards[1]

1. Purushottam Kar and Harish Karnick, "Random Feature Maps for Dot Product Kernels", 15th International Conference on Artificial Intelligence and Statistics (AISTATS 2012), La Palma, Canary Islands, April 21-23, 2012.

2. Surya Prakash and Phalguni Gupta, "An Efficient Technique for Ear Detection in 3D: Invariant to Rotation and Scale", 5th IAPR/IEEE International Conference on Biometrics (ICB 2012), New Delhi, India, March 29 - April 1, 2012.

3. Saurabh Joshi, Shuvendu Lahiri and Akash Lal, "Underspecified harness and interleaved bugs", 39th ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages (POPL 2012), Philadelphia, USA, January 25-27, 2012.

4. Surender Baswana, Utkarsh Lath, and Anuradha Mehta, "Near optimal distance oracle for planar digraphs avoiding any failed node or link", ACM-SIAM Symposium on Discrete Algorithms (SODA 2012), Kyoto, Japan, January 17-19, 2012.

5. Balwinder Sodhi and T.V. Prabhakar, "A Cloud Architecture Using Smart Nodes", 6th IEEE Asia-Pacific Services Computing Conference (APSCC 2011), Jeju Island, Korea, December 12-15, 2011.

6. Ajitha Shenoy K B, Somenath Biswas and Piyush P Kurur, "Metropolis Algorithm For Solving Shortest Lattice Vector Problem (SVP)", IEEE 11th International Conference on Hybrid Intelligent Systems (HIS 2011), Malacca, Malaysia, December 5-8, 2011.

7. Aditya Nigam and Phalguni Gupta, "Knuckleprint Recognition using Feature Tracking", 6th Chinese Conference on Biometric Recognition (CCBR 2011), Beijing, China, December 3-4, 2011.

8. Balwinder Sodhi and T.V. Prabhakar, "A design pattern to decouple data from markup", 12th International Conference on Electronic Commerce and Web Technologies (EC-Web 2011), Toulouse, France, August 29 - September 2 2011.

9. Shashank Mehta and Deepanjan Kesh, "A Saturation Algorithm for Homogeneous Binomial Ideals", 5th Annual International Conference on Combinatorial Optimization and Applications (COCOA 2011), Zhangjiajie, China, August 4-6, 2011.

10. Sagarmoy Dutta and Piyush Kurur, "Quantum Cyclic Code of length dividing $p^t + 1$", IEEE International Symposium on Information Theory (ISIT 2011), St. Petersburg, Russia, July 31st - August 5th, 2011.

11. Umarani Jayaraman, Amit Kumar Gupta, Surya Prakash ,Phalguni Gupta, "An Enhanced Geometric Hashing", 10th IEEE International Conference on Communications (ICC 2011), Kyoto, Japan, June 5-9, 2011.

12. Abhinav Mishra and Arnab Bhattacharya, "Finding the bias and prestige of nodes in networks based on trust scores", International World Wide Web Conference (WWW 2011), Hyderabad, 28th March - 1st April 2011.

13. Ruturaj Dhekane and Brion Vibber, "Talash: Friend Finding in Federated Social Networks", Workshop on Linked Data on the Web (colocated with WWW 2011), Hyderabad, March 29th, 2011.

[1] A complete list of department publications can be accessed at `http://www.cse.iitk.ac.in/research/star/2011-12.html`

Research Activities of Faculty Members

Prof. Amey Karkare

Research Work

1. Intelligent Tutoring System: In this project, we are working on building intelligent tutoring systems (for primary and secondary maths and science subjects, logic problems, etc.)

2. Automated End-user Programming: This project involves synthesis of text manipulating programs for spreadsheets. The work is submitted to the Symposium on Principles of Programming Languages (POPL 2013).

3. Debugging and Verifying Functional Programs: In this project, We are working on developing tools to simplify understanding, debugging and verification of functional programs. The work was presented at 2012 Symposium on Trends in Functional Programming (TFP 2012), and a paper is submitted to the post-symposium proceedings.

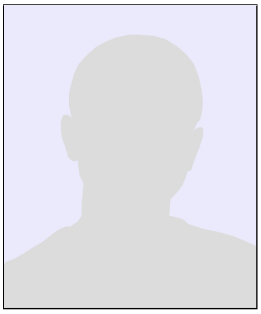
4. Shape Analysis: Understanding the properties of heap structures is important for any optimizing compiler. We have developed an effective shape analysis technique to discover sharing patterns in heap structures. Papers based on this work are accepted for publication in International Conference on Parallel Computing (ParCo 2011) and ACM Symposium On Applied Computing (SAC 2012). an extended version of SAC-2012 paper is submitted to the journal "Innovations in Systems and Software Engineering".

Publications

[1] Siddharth Agarwal and Amey Karkare. Functional SMT solving with Z3 and Racket. In *Symposium on Trends in Functional Programming (TFP 2012)*. University of St Andrews, UK, 2012.

[2] Sandeep Dasgupta Barnali Basak and Amey Karkare. Heap Dependence Analysis for Sequential Programs. In *International Conference on Parallel Computing (ParCo 2011)*, Ghent, Belgium, August 30 - September 2 2011.

[3] Sandeep Dasgupta and Amey Karkare. Precise Shape Analysis using Field Sensitivity. In *27th ACM Symposium On Applied Computing (SAC 2012)*, March 25-29 2012.



Prof. Krithika Venkataramani

Research projects with Students

1. **Optimal Fusion of Statistically Dependent Classifier Ensembles**

   (a) Automatic optimal classifier ensemble generation
   Feb-April 2011: Anurag Awasthi - course project for CS674
   When binary class data is present in multiple clusters, classifiers are generated to separate a cluster from one class from a cluster of another class. These classifiers are then combined, or discarded, in an iterative manner using decision fusion rules based on the classifier diversity. This provides an automatic manner of generating optimal classifier ensembles along with the best decision fusion strategy.

   (b) Theoretical calculation of optimal decision fusion rules
   Aug-Dec 2011: Shish Basu Palit - CS497
   The theoretically optimal decision fusion rule for a set of binary classifiers whose scores are jointly normally distributed is calculated. The dependence of the optimal rule and the corresponding minimum error rate on the correlation structure of the scores is investigated. The characterization of the optimal rules for the multi-class classification problem, by dividing it into a set of binary classification problems, is also discussed.

   (c) Ensemble fusion of continuous classifier outputs
   Aug-Dec 2011: Asheesh Agrawal - CS497
   Analysis of score fusion using diverse classifier ensembles is done here to find which fusion approach is optimal. The relation between number of classifiers and classifier diversity on the optimal fusion rule and fusion error is investigated.

2. **Fingerprint Recognition**

   (a) Fingerprint image quality quantification
   Jan-May 2011 and June-Aug 2012: Anurag Awasthi - CS697
   Different image quality metrics from literature are analyzed to select the best set. These are applied to quality impairment assessment of dry and wet fingers, thus providing information for any necessary fingerprint re-scanning to an Automatic Fingerprint Identification System (AFIS). Submitted a paper based on this work in August 2012.

   (b) Fingerprint indexing using SURF features
   May-July 2011: Satvik Chauhan - SURGE internship, co-supervision with Prof. Arnab Bhattacharya
   A hierarchical 3-stage indexing of SURF features is done for identification in large databases. The SURF points around the core point are used for indexing in increasingly larger areas in each higher stage. This is applied to a 1000 finger subset of the NIST 14 database.

   (c) Fingerprint indexing using minutiae neighborhood
   Sep 2010-Sep 2011: Abhay Sachan - M.Tech Thesis
   A rotation invariant representation of a minutiae neighborhood is created using the radial and

angular positions of neighboring minutiae from the central minutia location and orientation, respectively. This is indexed using Locally Sensitive Hashing. The core-point location is used for registration. This is applied on NIST 14, a large database of 27000 fingers.

3. **Decoding Cognitive States from Brain fMRIs**
   Feb-April 2011: Chittibabu Namballa and Rahul E. - course project for CS674
   A dimensionality reduction technique to find the "most differentiating fMRI voxels" for predicting the cognitive states of viewing an image and reading a related sentence is designed. Published a paper based on this work in March 2012.

4. **Urban area classification using integrated high-resolution multi-spectral images and LiDAR data**
   June 2010 - June 2011: Anand Mehta - M.Tech. Thesis, Co-supervision with Prof. Onkar Dikshit
   This work is applied on urban area data from two different study sites. First, the LiDAR point cloud is used to generate Digital Surface Model (DSM) and intensity raster. After that, DSM and intensity raster are registered with multi-spectral images. Then watershed segmentation is carried out over imagery, DSM and intensity raster. Finally, Gaussian Mixture Model (GMM) and rule-based classification methods are applied for obtaining the dominant classes in urban areas.

M. Tech. Students Graduated

1. Anand Mehta, "Integration of high-resolution imagery and LIDAR data for classification of urban area," M. Tech. Civil Engg., June 2011, co-supervision with Prof. Onkar Dikshit

2. Abhay Sachan, "Fingerprint Identification using Orientation in the Minutiae Neighborhood," M. Tech., CSE, Sept. 2011

Published/Submitted Papers

[1] Chittibabu Namballa, Rahul Erai, and Krithika Venkataramani. Decoding Cognitive States from Brain fMRIs: the most differentiating voxels way. In *ACIIDS*, volume 7196 of *LNAI (Part I)*, pages 369–376, March 2012.

[2] Anurag Awasthi, Krithika Venkataramani, and Avani Nandini. Image Quality Quantification for Fingerprints Using Quality-Impairment Assessment. submitted to WACV 2013 in August 2012.

Prof. Raghunath Tewari

Research Summary (since December, 2011)

I am currently interested in understanding the power and limitations of space bounded computations (typically logarithmic). The notion of unambiguous computation is a natural and well-studied restriction of general nondeterminism. In the logarithmic space framework, I along with some students, are trying to understand the power of unambiguity with respect to general nondeterministic computations. We have some partial progress on this problem. Our techniques have also yielded some interesting upper bounds for the seemingly unrelated perfect matching problem as well.

We are also looking at the deterministic end of the spectrum and investigating certain restricted versions of the graph reachability problem to obtain a better deterministic space upper bound on them. Here also we have obtained some initial results and are currently trying to extend them further.

The list of publications, in reverse chronological order, since December 2011, is given below.

Publications

[1] Raghunath Tewari and N. V. Vinodchandran. Green's theorem and isolation in planar graphs. *Information and Computation*, 215(0):1–7, June 2012.

[2] Samir Datta, Raghav Kulkarni, Raghunath Tewari, and N. V. Vinodchandran. Space Complexity of Perfect Matching in Bounded Genus Bipartite Graphs. *Journal of Computer and System Sciences*, 78(3):765–779, May 2012.

[3] Samir Datta, Arjun Gopalan, Raghav Kulkarni, and Raghunath Tewari. Improved Bounds for Bipartite Matching on Surfaces. In *29th International Symposium on Theoretical Aspects of Computer Science (STACS 2012)*, pages 254–265, March 2012.

Prof. Satyadev Nandakumar

This is a brief summary of my research and guidance activity over the past year.

Publications

Conferences

1. "Predictive Complexity and Generalized Entropy of Stationary Ergodic Games" (joint work with Mrinalkanti Ghosh), in the 23^{rd} *Conference on Algorithmic Learning Theory*, 2012.
2. "Axiomatizing Resource Bounded Measure" (joint work with Xiaoyang Gu, Jack Lutz and Jim Royer), 7^{th} *Conference on Computability in Europe*, June 2011.

Under Review

1. "Normality and Finite-State Dimension of Liouville Numbers" (joint work with Santhosh Kumar Vangepalli.)

Invited Workshops

1. Invited as a participant to Dagstuhl Seminar on "Computability, Complexity and Randomness", Schloss Dagstuhl, Germany, January 2012.

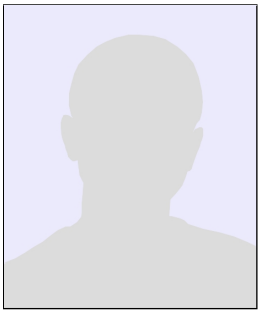
Master's Theses

The following Masters' theses were completed :

1. Mrinalkanti Ghosh, CSE department, July 2012. Thesis title "Predictive Complexity and Generalized Entropy of Stationary Ergodic Games".
2. Pulkit Bansal, Mathematics department, May 2012. Thesis title "Randomness and Differentiability".

I am currently supervising the following Master's theses :

1. Abhinav Tripathi, CSE department, expected graduation - December 2012.
2. Atanu Pal, CSE department, expected graduation - May 2013.
3. Prajyoti Waghmare, CSE department, expected graduation - May 2013.



Prof. Subhajit Roy

Short summary of research

1. Program profiling
   (a) Design of a new profiling algorithm "Pertinent Path Profiling" that allows for efficient profiling of the intra-procedural control-flow among a subset of all the basic-blocks in a program. (joint work with Mr. Ramshankar Chouhan and Prof. Surender Baswana)
   (b) Identifying frequently executed acyclic paths in a program execution using data stream algorithms (joint work with Mr. Gaurav Kumar)
2. Design of new intermediate representations to aid compiler optimizations
   Design of a new intermediate representation "the Object SSA form" that allows for efficient analysis of heap-memory (joint work with Mr. Yash Vohra)
3. Transactional memory systems
   Improving multi-versioned transactional memory systems by reducing the number of false conflicts (joint work with Mr. Tanmay Mogra)
4. Program debugging
   Using program phases and statistical bug isolation for fault localization (joint work with Mr. Varun Modi and Prof. Sanjeev K Aggarwal)

Recent Publication

[1] Arun Ramamurthi, Subhajit Roy, and Y. N. Srikant. Probabilistic dataflow analysis using path profiles on structure graphs. In *SIGSOFT FSE (New Ideas track)*, pages 512–515, 2011.

Research Activities of PhD Students

Aditya Nigam

Supervisor : Prof. Phalguni Gupta

Thesis

In this academic year we have proposed a knuckleprint based recognition system [3]. Also we have developed iris [1] and 4-slap [4] fingerprint segmentation algorithm that is the preliminary and most important step in any iris and slap based recognition system respectively. Apart from this we have also developed a demography data based de-duplication algorithm [2] using some string matching techniques.

Iris [1]

It is done by combining the two techniques hough transform and integro-differential operator in such a way so that they can compliment each other. It works on enhanced and normalized iris edge maps to segment it in real time.

Demography based de-duplication [2]

The demography information can be used to de-duplicate big size multi-modal biometrics databases before going into biometric trait based de-duplication. The edit distances are used in order to calculate the distance between two name strings. The word flipping, common word binning and word phonetics are used to achieve good performance.

Knuckleprint [3]

The texture based feature in knuckleprints are observed to be very discriminative. The proposed enhancement techniques can underline and highlight these features. The corner features from enhanced knuckleprints are extracted that are later tracked using LK tracking. The tracking performance is measured using proposed dissimilarity measure so as to discriminate between genuine and imposter matching.

4-Slap Fingerprint [4]

The recent 4-slap scanners can capture all the four fingers simultaneously hence a segmentation algorithm is needed to extract the four fingers in order to match them. The proposed algorithm can segment the single 4-slap image and extract the four required roi's of index, middle, ring and little finger. Also it aligns all the fingerprints vertically by estimating the orientation of the print so as to facilitated the matching.

References

[1] Amit Bendale, Aditya Nigam, Surya Prakash, and Phalguni Gupta. Iris segmentation using improved hough transform. In *ICIC (3)*, pages 408–415, 2012.

[2] Vandana Dixit Kaushik, Amit Bendale, Aditya Nigam, and Phalguni Gupta. An efficient algorithm for de-duplication of demographic data. In *ICIC (1)*, pages 602–609, 2012.

[3] Aditya Nigam and P Gupta. Finger knuckleprint based recognition system using feature tracking. In *CCBR*, pages 125–132, 2011.

[4] Nishant Singh, Aditya Nigam, Puneet Gupta, and Phalguni Gupta. Four slap fingerprint segmentation. In *ICIC (2)*, pages 664–671, 2012.

Ajitha Shenoy K B

Supervisor : Prof. Somenath Biswas

Research Activity for the year 2011-2012

We defined search space for shortest lattice vector problem (SVP). We also proved that our search graph had polynomial number of neighbours and it has maginifaction greather than 1/2. We also applied Hasting's Generalization of Metropolis algorihtm for sovling SVP and found the results are quite encouraging. The details of my publications during this period are given below.

1. Ajitha Shenoy K B, Somenath Biswas and Piyush P Kurur, "Metropolis Algorithm For Solving Shortest Lattice Vector Problem (SVP)", Proceedings of 11th International Conference on Hybrid Intelligent Systems (HIS 2011), Malacca, Malaysia, December 5-8, 2011 (Published in IEEE).

2. Ajitha Shenoy K B, Somenath Biswas and Piyush P Kurur, "Search Space Formulation and Hasting's Generalization of Metropolis Algorithm for SVP", International Journal of Computer Information Systems and Industrial Management Application, ISSN 2150-7988,vol 5, 2013.

Amrita Chaturvedi

Supervisor : Prof. T. V. Prabhakar

We contributed further to our existing work of ontology driven MVC style. We demonstrated the utility of the proposed style through an example implementation based on the proposed style. We found that the proposed solution addresses all the drawbacks of MVC style and also enhances software quality attributes like maintainability and reusability. The implementation aspects reveal that the proposed solution gives rise to a new programming technique where the ontology drives the entire communication sequence, breaking the communication dependencies between the components. This enhances the maintainability and reusability of the components but negatively impacts the overall performance because of the indirection in the flow of control. We also found out the contexts in which the proposed ontology-driven MVC style can be used. These are as follows:

1. When the model and/or the view is expected to change in near future.
2. When the view is rich.
3. When runtime changes in the view can be made.
4. When the user interface (UI) designer wants to generate and fix various different types of user interfaces.

The general trend in software engineering is to capture the functional requirements specification in UML use case model and user interface design specifications in task models. The use of two different models for functional requirements specifications and user interface design specifications leads to inconsistencies, redundancies and overlaps in the later phase of software development life cycle. We wrote a paper to address this problem by advocating a single model (domain action ontology) for both the specifications. The domain action ontology has a tree structure to represent the domain actions. Sequence of actions, alternate actions and sub actions can be represented very easily. Unlike task model, the action ontology hides unnecessary task details by indicating only the user actions which can perform user desired application tasks i.e. the domain action ontology only includes the user actions which are actually performed on the Model component of the MVC style. Moreover, the domain action ontology purely represents the domain actions and no knowledge about domain objects is included in it which makes the model reusable. It is linked to the domain object ontology through object type property to indicate the link between the domain actions and domain entities.The paper uses MVC style to demonstrate the how domain object ontology (DOO) and domain action ontology (DAO) can be used to address problem of inconsistencies in the different functional requirements specification models . We describe how a fully functional View (user interface) can be automatically generated from the DOO and DAO which enables early validation of both functional requirements as well as the user interface design specifications. Any discrepancies can be removed by changing the single model. The DAO also drives the control flow of the MVC style thus making the components completely independent of each other.

Arpita Korwar

Supervisor : Prof. Manindra Agrawal

This year, I worked on the following topics:

1. Talk on Multi-commodity max-flow min-cut problem: This is a NP-hard variant of the max-flow min-cut theorem by Ford and Fulkerson. Since it is NP-hard, we can only hope to get an approximate answer in polynomial-time (unless P=NP). We studied two approximation algorithms, one using the Primal-dual schema and another, using LP-rounding for two variants of the problem. This talk was given as my comprehensive exam.

2. Talk on Arithmetic Circuit Lower Bounds for the Permanent: Permanent is the unwieldy cousin of the determinant. It is #P-complete and VNP-complete. Hence, a lower bound for the permanent, showing that it cannot be computed by polynomial sized arithmetic circuits would mean that VNPis not equal to VP.

   Sometime ago, it was shown that a special kind of black box identity test would imply lower bounds on the permanent. I described this approach and its development since then. This talk was my State of the Art seminar. Since then I have been working on the questions raised in this talk.

3. Poset games: These are 2-player games, played on an input poset. The players take turns, removing a vertex, and all the vertices greater than that vertex, from the poset. The first player who encounters an empty poset loses. What is the complexity of deciding, given a poset, which player wins? What should be the winning strategy? What is the complexity of computing this winning strategy (or the next move)?

   It was recently shown that deciding which player wins is PSPACE-hard. Even 3-level games are PSPACE-hard. We were studying the complexity of 2-level poset games. This study took place in University of Ulm, Germany. Prof. Thomas Thierauf had invited us.

4. India Theory Day: I attended a day of research talks by eminent researchers in Theoretical Computer Science. It was organized and sponsored by Microsoft Research and held in IISc, Bangalore. It included talks by illustrious people like Eva Tardos, Nisheeth Vishnoi and Amit Kumar. The talks were on topics like Randomized rounding, clustering problem, and Lattice based Cryptosystems.

Ashish Agrawal

Supervisor : Prof. T.V. Prabhakar

During the year 2011-2012, focus of my research area is on "Software architectural perspective of cloud platforms and its applications". Cloud computing facilitates instantaneous provisioning of resources to the user in an elastic manner. The computing model reduces the operational cost by achieving server consolidation. Using this, cloud user can request for resources in a granular manner even through programs. However, cloud computing has also affected the development of software applications. Depending upon the service level (IaaS, PaaS or SaaS), different cloud models have different programming models which directly impact the architecture of application software. For an application developer, the number of available resources are infinite. This changes the way resources are allocated to the software components. Another important thing is that by choosing a particular cloud platform, some parts of software application (its platform, which includes both software and hardware components) are already fixed. As deployment diagram is very essential part of software architecture of application, fixed components in deployment diagram impact quality of software application. Impact of cloud platforms, their architecture and services offered on architectural design and development of software application is the main focus of my research work. My research work is on finding solutions for following questions:

1. How do the components in a cloud platform impact quality of software applications deployed on it ?
2. Which cloud platform is best suitable for managing a data-center ?
3. How to exploit services of cloud platforms to achieve quality attributes in the application ?
4. How to design new cloud architectures?

Our work in the above areas have been resulted in few papers in international conferences during the mentioned year. [1] [2].

References

[1] Ashish Agrawal, Balwinder Sodhi, and T. V. Prabhakar. Lift - a mechanism for composing virtual appclusters from heterogeneous apps. In *2012 International Conference on Collaboration Technologies and Systems (CTS)*, pages 211 –217, Denver, CO, USA, May 2012.

[2] Balwinder Sodhi, Ashish Agrawal, and T. V. Prabhakar. Appification of web applications: Architectural aspects. In *2012 IEEE Workshop on Mobile Cloud Computing (MobiCC 2012)*, Bejing, China, August 2012.

Balwinder Sodhi

Supervisor : Prof. T.V. Prabhakar

Details

This year I and my supervisor (Prof. T.V. Prabhakar) continued to investigate several aspects of Cloud computing that concern software architecture and design. Complexity of contemporary business applications has increased. At the same time the computing platforms and hardware devices too have become more powerful. Adoption of virtualization and cloud based computing platforms by enterprises has seen an upward trend.

Given all of the above change and evolution, the focus of my work has been the following research questions:

1. How these changes in the modern datacenter and development environments have impacted the way we design and architect software systems and applications?
2. How can we characterize the computing environments to better understand their impact on software design and architecture activities?
3. How do various computing environments (e.g. virtual, cloud-oriented etc.) impact various non-functional quality attributes (NFQA) of a system?
4. How can we assess the suitability of a particular computing environment for a specific goal of satisfying certain NFQAs?

Results of our research investigations around the above questions have been published in the following international conferences during this period:

1. Balwinder Sodhi and T.V. Prabhakar. Cloud Platforms: Impact on Guest Application Quality Attributes. In The 2012 IEEE Asia-Pacific Services Computing Conference (APSCC 2012). December, 2012, Guilin, China. (Accepted)
2. Balwinder Sodhi and T.V. Prabhakar. Performance Characteristics of Virtualized Platforms from Applications Perspective. In International Conference on Data Management in Cloud, Grid and P2P Systems (Globe 2012). September, 2012 in Vienna, Austria. Lecture Notes in Computer Science(LNCS), Springer. (Accepted)
3. Balwinder Sodhi, Ashish Agrawal and T.V. Prabhakar. Appification of Web Applications: Architectural Aspects. In 2012 IEEE Workshop on Mobile Cloud Computing (MobiCC 2012). August 2012 in Beijing, China.
4. Balwinder Sodhi and T.V. Prabhakar. Cloud-oriented platforms: Bearing on Application Architecture and Design patterns. In proceedings of The IEEE Congress on Services (SERVICES 2012), June 2012 Honolulu, Hawaii USA.
5. Sodhi, B. and Prabhakar, T.V. (2012) An Architecture for Enterprise PC Cloud, Int. J. Computational Science and Engineering, Vol. 7, No. 4, pp.296-307.

6. Ashish Agrawal, Balwinder Sodhi and T.V. Prabhakar. Lift - A mechanism for Composing Virtual App-clusters from Heterogeneous Apps. In The Proceedings of 13th ACM/IEEE/IFIP International Conference on Collaboration Technologies and Systems (CTS 2012), May 2012, Denver, Colorado, USA.

7. Balwinder Sodhi and T.V. Prabhakar. A cloud architecture using smart nodes. In proceedings of The 2011 IEEE Asia-Pacific Services Computing Conference (APSCC 2011), December 2011, Jeju Island, Korea.

Debarati Das

Supervisor : Prof. Manindra Agrawal and Prof. Raghunath Tewari

Courses Done

During first and second of 2011-2012 I did the following eight courses:

Course No.	Course Name
CS601	Fundamentals of CSE
CS645	Design and Analysis of Algorithms
CS648	Randomized Algorithms
CS727	Topics in Internet Technology
CS640	Computational Complexity
CS646	Parallel Algorithms
CS687	Algorithmic Information Theory
CS718	Sublinear Algorithm

Projects Done

- **MedShop:SMS Based Shopping**
  Here we have made one application for SMS based shopping.We have mainly done it for a Medicine Shop.Registered customers can send the madicine name,dosage and amount and accordingly the order will be delivered and if the particaular namadecine is out of stock names of madicines with similar drug will be send to user through SMS only.This can be extended for other shopping items also.
- **Smallest Enclosing Circle**
  It was actually an programming project where we have applied the Randimized algorithm of Smallest Enclosing Circle throug coding and analysied its nature , run time with a dataset of size 4-5 million.

Diptarka Chakraborty

Supervisor : Prof. Manindra Agrawal and Prof. Raghunath Tewari

Courses Done

Course No.	Course Name
CS601	Fundamentals of CSE
CS641	Modern Cryptology
CS685	Data Mining
CS698X	Multimodal Biometrics
CS640	Computational Complexity
CS687	Algorithmic Information Theory
CS718	Sublinear Algorithm
CS685	Multiagent Systems

Projects Done

- **Prediction of Stock Exchange Share Price using ANN and PSO**
  Stock Exchange Share Price is very hard to predict system since there are no significant rules to estimate or predict that. We use Multilayer Feed-forward network as a network model for predicting stock price and to train this network model we are going to use Particle Swarm Optimization.

- **An Efficient Finger-knuckle-print based Recognition System Fusing SIFT and SURF Matching Scores**
  It was an implementation project implementing the methods discussed in the following paper
  An Efficient Finger-knuckle-print based Recognition System Fusing SIFT and SURF Matching Scores
  G S Badrinath, Aditya Nigam and Phalguni Gupta

- **Rational Secret Sharing & Game Theory**
  Consider m out of n secret sharing protocol among n players where each player is rational. Although the previously known protocol was secure from the game theoretic point of view but main problem in their protocol is unbounded number of rounds. Here I propose a new 2-round protocol for Rational Secret Sharing which is also completely secure from the game theoretic point of view.

Kamlesh Tiwari

Supervisor : Prof. Phalguni Gupta

This year we have proposed an algorithm to encrypt image database. The proposed algorithm generates a unique key for every image at the time of encryption/decryption to make encryption algorithm robust against attacks. The encrypted image appears chaotic and random; and thereby does not reveal any features of original image. It presents a mechanism for a secure image database transfer over insecure channel. It uses least significant bit insertion method to make the encrypted image visually unsuspicious. The result got published in International Conference on Intelligent Computing (ICIC-2012) [2].

We have also proposed a feature extraction technique for palmprint based identification system [1]. This system makes use of one dimensional Discrete Cosine Transform (DCT) to design an efficient palmprint based recognition system. It extracts the palmprint from the hand images which are acquired using a flat bed scanner at low resolution. It uses new techniques to correct the non-uniform brightness of the palmprint and to extract features using difference of 1D-DCT coefficients of overlapping rectangular blocks of variable size and variable orientation. Features of two palmprints are matched using Hamming distance while nearest neighbor approach is used for classification. The system has been tested on three databases, viz. IITK, CASIA and PolyU databases and is found to be better than the well known palmprint systems.

References

[1] G. S. Badrinath, Kamlesh Tiwari, and Phalguni Gupta. An efficient palmprint based recognition system using 1d-dct features. In *International Conference on Intelligent Computing (ICIC)*, pages 594–601, 2012.

[2] Kamlesh Tiwari, E A Siddiqui, and Phalguni Gupta. An efficient image database encryption algorithm. In *International Conference on Intelligent Computing (ICIC)*, pages 400–407, 2012.

Pawan Kumar Aurora

Supervisor : Prof. Shashank Mehta

While working on the maximum map problem, we realized that our formulation of that problem resembles the Lovasz Theta function for a specific perfect graph. So we decided to study the feasible space of the Semi-definite program underlying the function for that graph. We could show that there is a special polytope that lies within the feasible space. The vertices of this polytope have a one-to-one correspondence with the vertices of the Birkhoff Polytope but the facets are not as trivial as those of the Birkhoff polytope. More significantly we could show that any point of the feasible space can be expressed as a linear combination of the vertices of this polytope. Also the feasible space when restricted to the cone of completely positive matrices gives us precisely this polytope.

Along the side lines we have been working on the partial degree bounded edge packing problem with non-uniform degree constraints. The objective is to find a subset of edges such that each selected edge satisfies the degree constraint at at least one of its end vertices in the subgraph induced by these edges. Also the constraints are non-uniform i.e., the vertices can have different constraints on their degrees. The existing results are for uniform constraints only. We have managed to get a 2-approximation algorithm for this NP-hard problem for the case when the edges are unweighted and a $\log n$-approximation when the edges are weighted. We also have a polynomial time exact algorithm for the special case when the input graph is a tree. Since the LP relaxation of the natural IP formulation of the problem has a large integrality gap it is not useful in getting a good approximation. So an approximate IP is proposed whose optimum value is within a constant factor of the value of an optimal solution to the problem. Although this method does not yield a better than 2 approximation factor, we hope this approach can be useful in getting better approximation algorithms for some other hard problems. We will be submitting this work to a conference shortly.

Puneet Gupta

Supervisor : Prof. Phalguni Gupta

First semester 2011-12

I have done a course on multimodal biometrics system under Dr. Phalguni Gupta. I have done a project for this on ear detection that is invariant to geometrical distortions like scaling and rotation.

Second semester 2011-12

After I have finished course work, I gave my comprehensive exam. The title of my presentation is "Top-k Queries Over a mixture of attractive and repulsive dimensions".

Thereafter I started to look for my thesis problem by exploring various biometrics techniques and their associated problems. Based on my understanding I choose to works on fingerprints and I give my state of the art seminar on "Slap fingerprint image for Personal Authentication". In this I have discussed various problems related to the use of Slap fingerprint image in authentication of human being and their proposed solution. I have also attending a conference ICB (International conference on biometrics) 2012 which is one of the major conference in biometrics.

Now I have started to work on slap fingerprint segmentation which is the first step of using slap image in personal authentication of a system. Towards the same my paper titled "Slap fingerprint segmentation" gets accepted in BTAS (Biometrics:Theory,Applications and Systems)2012.

Purushottam Kar

Supervisor : Prof. Harish Karnick and Prof. Manindra Agrawal

The academic work undertaken in the past year (Fall 2011 - Summer 2012) was mainly of two types - collaborative research and research visits to conferences, workshops and the Microsoft Research (MSR) Lab India. I collaborated with students / faculty at IIT Kanpur as well as researchers at MSR Labs, Bengaluru. We looked at a variety of research questions in the past year, of which accounts are given below. The collaboration with the researchers at MSR India was made possible, in part, by two research visits to the lab. A detailed account of the visits is given below.

Research Visits

As mentioned earlier, I undertook two research visits to MSR Labs, Bengaluru. Of this one was a short term visit (November 2011) lasting approximately 5 weeks and the other was a summer internship from May 2012 till August 2012 (14 weeks). During these visits I collaborated and had discussions with Dr. Prateek Jain and Dr. Manik Varma of the Machine Learning and Optimization (MLO) group. Apart from this I had conversations and interacted with several other researchers, research assistants and interns at the Algorithms and MLO Research groups.

I attended two schools/workshops during this period

1. Machine Learning Summer School, Canary Islands, Spain, April 11-20, 2012.
2. MSR India, Mysore Workshop on Machine Learning, August 1-5, 2012.

At both places, I gave a short talk / poster presentation on my past work. Apart from this I attended the following two conferences to present research papers

1. 24^{th} Annual Conference on Neural Information Processing Systems (NIPS 2011), Granada, Spain, December 12-17, 2011.
2. 15^{th} International Conference on Artificial Intelligence and Statistics (AISTATS 2012), Canary Islands, Spain, April 21-23, 2012.

The paper presented at NIPS 2011 [1] was joint work with Dr. Prateek Jain. The paper presented at AISTATS 2012 [2] was joint work with Prof. Harish Karnick who is my advisor.

The conference travel to AISTATS and the Machine Learning summer school was partly funded by the Research-I foundation. Additional support was obtained from various sources such as travel grants from conference organizers and from a Microsoft Research India PhD fellowship of which the author is a recipient.

Research Work Undertaken

Here I give very a very brief overview of some of the projects I have been involved in the past year. Kernel algorithms have been one of the most influential paradigms within machine learning for the past two decades and have been responsible for state-of-the-art performance in several areas such as vision, text processing, spam detection, bioinformatics etc. Kernel algorithms require mainly two ingredients to produce their output. For example let us take the case of learning a spam filter. A kernel learning algorithm would require the following two inputs :

1. A *Training set* : A set of emails that are known to be spam and a set of emails that are known to be not spam. Typically the user of a mail client like Outlook or Thunderbird provides this data by tagging emails as spam/non-spam. Let us denote this set by $\mathbf{x}_1, \ldots, \mathbf{x}_n$ - each of which is tagged with a *label* +/- depending on whether $\mathbf{x}_i$ is a spam email or not.

2. A *Kernel* : A bivariate function that imposes a measure of similarity between two emails. Let the set of all emails be called $\mathcal{X}$, then a kernel is a function $K : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$

Kernel algorithms like SVMs can take these inputs and output a classifier $h : \mathcal{X} \to \{+, -\}$ which can label future emails as spam/non-spam. There are two constraints with vanilla SVM frameworks :

1. Requirement of a *Positive Definite* Kernel : The kernel must satisfy a formal condition known as positive semi-definiteness. This aids in formulating convex optimization problems that can be solved efficiently. However several notions of similarity are not positive semi-definite and it remains to be seen if they can also be used to train classifiers efficiently.

2. Form of the classifier : the classifier is always of the form $h(\mathbf{x}) = \text{sign}\left(\sum_{i=1}^{n} \alpha_i K(\mathbf{x}, \mathbf{x}_i)\right)$ for $\alpha_i \in \mathbb{R}$. This can result in slow test times since the new email would have to be compared to a large number of training set emails (those which have non-zero α_i terms) before we can give a verdict.

In [1] we addressed the first problem by showing how one can use the existing SVM formulations and software packages to efficiently train using arbitrary similarity measures. Our work gives theoretical guarantees for the learned classifier provided the similarity measure being used was "suitable" for the problem in a formal sense. We also provided empirical validation of our framework on benchmark datasets.

In [2] we addressed the second problem for the class of dot product kernels. We give algorithms based on classical results in harmonic analysis that allow us to provide testing times that are independent of the training set sizes. This effectively allows us to use huge training sets without worrying about the resulting classifier being slow at test time.

Publications

[1] Purushottam Kar and Prateek Jain. Learning with Similarity via Data-driven Embeddings. In *25th Annual Conference on Neural Information Processing Systems (NIPS)*, 2011.

[2] Purushottam Kar and Harish C. Karnick. Random Feature Maps for Dot Product Kernels. In *15th International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2012.

Rohit Gurjar

Supervisor : Prof. Manindra Agrawal

In year 2011-12, we mainly worked on parallel complexity of the perfect matching problem and also looked at some other related problems. Below we describe the work done.

Introduction

Given a graph $G = (V, E)$ the perfect matching problem asks if there is a set of disjoint edges in the graph which cover every vertex in it. In other words, say for a group of people we are given connections of the form 'person x and person y know each other' and the question is whether one can make pairs of persons known to each other, such that every person is covered.

The perfect matching problem is of fundamental interest in combinatorics, algorithms and complexity theory for a variety of reasons. Although there are efficient sequential algorithms known for the problem, its parallel complexity is still unclear. An efficient parallel algorithm is the one which runs on polynomially many processors and takes only polylogarithmic time. The class of problems which have efficient parallel algorithm is called NC. A very important question in the theory of parallel computation is whether all sequentially 'easy' problems are efficiently parallelizable. The same question for the perfect matching problem has been of great interest in complexity theory. The problem is known to have RNC algorithms i.e. it has efficient parallel algorithms which uses some random bits. However, a deterministic efficient parallel algorithm, i.e. which does not use any random bits, is not known. The main focus of our work is to get a deterministic parallel algorithm for some non-trivial class of graphs.

The same question can be asked for various versions of matching problem - such as maximum matching, maximum weight perfect matching, exact matching, unique perfect matching - and for various classes of graphs. Also, for each of these problems one can define a *decision* version and a *construction* version, since in general, the decision and construction are not known to be NC-equivalent. For some natural classes of graphs such as planar graphs, regular bipartite graphs, dense graphs etc. some of these problems have NC algorithms. Apart from looking at the general problem, we also worked on some special classes of graphs and different versions of the problem. The sections below describe our approaches and results.

Current Work

Planarizing Gadgets

Planar graphs are graphs which can be drawn on a plane without any edges intersecting. For planar graphs, deciding the existence of a perfect matching can be done in NC. So, one natural way of solving the matching problem would be to reduce it to the planar version. The reduction must work in NC. One such possible reduction is through planarizing gadgets. A planarizing gadget is a planar graph which would locally replace every crossing in a planar drawing of the graph. We show that no such planarizing gadget exists which would preserve the existence of a perfect matching. This eliminates one possible way of reduction from general

matching to planar matching.

We also extend this result to unique perfect matching, weighted perfect matching and exact perfect matching. We also show that there is no planarizing gadget which would preserve the nonzeroness of number of matchings modulo some number $k \neq 2$. The planarizing gadget for the last problem was unlikely as the problem is believed to be very hard for general graphs, but we showed the nonexistence without any assumptions. This resulted in a publication, see here [GKM$^+$12].

Planar Graphs

As described earlier the decision version for planar graphs has an NC algorithm. But for the construction version no NC algorithm is known. When the given graph is bipartite planar, then a perfect matching can be constructed in NC. There has been a lot of work in bipartite planar graphs and mainly there are three approaches- network flows, isolation and the matching polytope approach.

For non-bipartite planar graphs, we believe the isolation approach cannot be extended directly. But for the matching polytope approach we have some partial progress in extending it to the non-bipartite case. The idea is to start from a point inside the polytope and move towards a corner, which would be a perfect matching. For non-bipartite graphs the description of the matching polytope is a bit more complicated as compared to the bipartite case. It was not clear while moving, how to stay inside the polytope. We gave an NC separation oracle for the matching polytope for planar graphs, which would tell if a given point is inside the polytope. However, how to reach a corner in a small number of steps is still unclear.

$K_{3,3}$-free bipartite graphs

We give a logspace reduction from $K_{3,3}$-free bipartite matching to planar bipartite matching, thus putting it in complexity class SPL.

References

[GKM$^+$12] Rohit Gurjar, Arpita Korwar, Jochen Messner, Simon Straub, and Thomas Thierauf. Planarizing gadgets for perfect matching do not exist. In *MFCS*, pages 478–490, 2012.

Saurabh Bhanuprasad Joshi

Supervisor : Prof. Sanjeev Kumar Aggarwal

Research Work

In this year (2011-12) I have mostly worked on finishing my existing works and getting them published.

The first work [1] deals with finding bugs in concurrent programs even when the given harness is not precise. In the absence of a precise harness even the most precise model checkers lead to false alarms. It takes several iterations and expert manual intervention to classify whether the reported warning is a *true bug* or not. If not, the harness needs to be revised in order to captured *missed* assumptions about the environment behaviour. We automate this process of classification-revision-verification under the assumptions that the given concurrent program is correct under sequential semantics. We use the sequential behaviour as an oracle to automate the classification phase. If the warning is found to be a false alarm, a filter program is generated so as to avoid the similar warnings in the next iteration. We show the efficacy of our approach on real world programs in reducing false alarms in the absence of precise harness.

Another work [2] proposes a new method for may-happen-in-parallel analysis. We argue that our method can provide optimization opportunities which are missed by earlier works. Though, our work does not completely subsume earlier works, our algorithm offers better time complexity as compared to earlier approaches. Hence, for the best results, one can use our method along with earlier approaches.

Apart from this, I focussed on automated repair of concurrent programs. Under the assumption that the sequential behaviours of the given concurrent programs are correct, we give techniques to automatically rule out *bad interleavings* so as to avoid assertion failures. Not only we are able to synthesize atomic regions with respect to strong atomicity faster as compared to earlier approach, but we also synthesize atomic regions with respect to weak atomicity while providing soundness guarantees. We are concentrating on getting this work published at a suitable venue.

Financial Aid

I thank Research-I foundation for the followings :

- Enhanced scholarship throughout my PhD programme.
- Travel grant of RS. 1,04,800/- to present my publication [1] at the United States of America.
- Travel grant of RS. 3736/- to attend FSTTCS 2011 at Mumbai.

References

[1] Saurabh Joshi, Shuvendu K. Lahiri, and Akash Lal. Underspecified harnesses and interleaved bugs. In *Proceedings of the 39th annual ACM SIGPLAN-SIGACT symposium on Principles of programming languages*, POPL '12, pages 19–30, New York, NY, USA, 2012. ACM.

[2] Saurabh Joshi, R.K. Shyamasundar, and Sanjeev K. Aggarwal. A new method of mhp analysis for languages with dynamic barriers. In *Parallel and Distributed Processing Symposium Workshops PhD Forum (IPDPSW), 2012 IEEE 26th International*, pages 519 –528, may 2012.

M Seetha Ramaiah

Supervisor : Prof. Amitabha Mukerjee

I completed my comprehensive exam and the state-of-the-art (SOTA) seminar during this period. Topic for my comprehensive exam was Deep Learning based on [1] and related articles. For the SOTA seminar, I investigated neighbourhood selection and manifold composition aspects of manifold mearning. I proposed to work on the problem of manifold composition which is described below.

Introduction

Consider two circular mobile robots, which can only translate but not rotate, independantly moving in a 2-dimensional plane. Each of these robots can be at any point in the plane at any given time. Two coordinates are needed to specifiy the position of each robot in the 2D plane and the motion of the two robots together forms a 4-manifold.

Now consider the problem of learning this 4-manifold from a collection of images containing the two robots at various positions. Due to the empty space phenomenon, which says that the datasets become exponentially sparse as dimensionality increases, the number of examples required to reduce the motion of two robots into a 4-manifold would be much greater than what we would need if we were to reduce the motion of one robot to a 2-manifold.

So instead of learning the motion of two robots as a 4-manifold at once, it is computationally benificial to learn the two 2-manifolds of the motions of two robots independantly and then combine them to get the 4-manifold of the motion of the two robots together.

Problem Statement

Let $X = \left\{x^{(1)}, x^{(2)}, ..., x^{(n)}\right\}$ be a set of *observable* points drawn from a *smooth m-manifold* $\mathcal{O}$ embedded in $\mathbb{R}^p$ such that the observable points $x \in \mathcal{O}$ are generated from some *latent variables/parameters* $y \in \mathcal{L} \subseteq \mathbb{R}^m$ by some bijective function $g : \mathcal{L} \to \mathcal{O}$. We will call $\mathcal{O}$ the *observable space* and $\mathcal{L}$ the *latent space*.

The problem of *manifold learning* is the following: given a finite set X of p-dimensional points, discover the underlying m-manifold $\mathcal{O}$ and find an m-dimensional embedding Y of X such that $y^{(i)} \in Y$ is an m-dimensional represenation of $x^{(i)} \in X$. In other words, manifold learning is the process of finding $g^{-1} : \mathcal{O} \to \mathcal{L}$.

In many situations, such as the example of robot motion planning described in section , the observable space $\mathcal{O}$ is composed of more than one independant component observable spaces $\mathcal{O}_1, \mathcal{O}_2, \ldots, \mathcal{O}_c$, generated from the component latent spaces $\mathcal{L}_1, \mathcal{L}_2, \ldots, \mathcal{L}_c$ respectively, such that $\mathcal{O}_i$ is an m_i dimensional manifold, $\mathcal{L}_i \subseteq \mathbb{R}^{m_i}$, $i = 1, \ldots, c$ and $\sum_{i=1}^{c} m_i = m$. In such cases, instead of learning $\mathcal{O}$ directly to get $\mathcal{L}$, it is computationally benificial to learn each $\mathcal{O}_i$ to get the corresponding $\mathcal{L}_i, i = 1, \ldots, c$, and compose them to get $\mathcal{L}$.

We have been trying to formalize the notion of *composition* and the goal is to study the conditions under which two or more manifolds are *composable*.

References

[1] Yoshua Bengio. Learning deep architectures for AI. *Foundations and Trends in Machine Learning*, 2(1):1–127, 2009.

Shubhadip Mitra

Supervisor : Prof. Arnab Bhattacharya

We realized that queries involving identification of various complex real world events or phenomena invlve joins over spatial and temporal attributes. Firstly, confining our study to temporal joins, we observed that such joins can be well expressed as constraint satisfaction problems. Further we realized that constraints representing temporal join queries over point and interval events are *binary linear row convex* constraints.

Therefore we investigated the problem of constraint satisfaction over binary linear row convex constraints. Given a set of variables $\mathcal{X} = (X_1, X_2, \ldots, X_n)$, each of which is associated with a finite domain of distinct integer values, and a finite set of constraints, a Constraint Satisfaction Problem (CSP) asks whether there exists an assignment to the variables that satisfy all the constraints.

Constraint satisfaction over row convex constraints is though NP Hard, we showed that the problem is polynomially solvable for linear row convex constraints. We studied the various notions of consistency, viz arc consistency, path consistency and degree-k consistency that have already been established in the literature. We extended a generic arc consistency algorithm namely AC4, and designed a faster algorithm for the class of linear row convex constraints. Further, we proposed another algorithm that incrementally scans the variable domains and reports a solution as soon as it detects one combination of values that satisfy all the constraints.

This work was submitted at AAAI 2012 but faced rejection because the work was not well differentiated with some of the existing related works. Henceforth, we generalized the scope of our work, by considering the CSP problem over *staircase* constraints. These constraints generalize the class of linear row convex constraints. These class of constraints also generalize the class of monotone constraints.

By closely comparing with the related work, we have confirmed the novelty of our work. We studied some closure properties of staircase constraints namely, intersection, transposition and composition. We showed that arc consistency is sufficient to solve a CSP with all constraints as down staircase constraints and path consistency is sufficient to solve a CSP with all constraints as up staircase constraints.

A faster arc consistency algorithm for down staircase constraints and a faster path consistency algorithm for up staircase constraints have been designed. A further improved non-arc consistency based deterministic algorithm is being proposed for CSP with down staircase constraints.

All these results are currently being written down for a journal submission at JAIR.

In May, 2012, I cleared my comprehensive examination. Meanwhile in parallel, we initiated weekly group meetings namely SIGDATA, where we have been discussing the state of the arts in databases and data mining. During the summer vacation, we discussed the literature in spatio temporal databases.

Siddharth Kumar Rai

Supervisor : Prof. Mainak Chaudhuri

I joined the department in July 2011. During academic year 2011-12 I completed my course work and gave Comprehensive Exam and State-of-the-art seminar.

In the first semester I completed the course work. In the second semester I gave Comprehensive Exam. On 5th September I gave my State-of-the-art seminar titled 'Avoiding Trips to Slow Memory : Algorithms to Manage Last-level Caches Effectively'. I am currently working on the problem of efficient management of Last-level cache for heterogeneous architecture. In today's multi-core processors, graphic core has been integrated with general purpose core and to avoid high latency DRAM accesses a large Last-level cache have been incorporated in the memory hierarchy. But how this large LLC can be shared between graphic and general purpose core is not known. My current focus is to investigating how GPU can benefit a large LLC.

Sudhanshu Shukla

Supervisor : Prof. Mainak Chaudhuri

I am a third year Ph.D. student working in the area of computer architecture. In the past year, I have continued to work on scalability of multi-core processors.

Multi-core processors with upto sixty four cores are already commercially available and future microprocessors are expected to have hundreds/thousands of cores. The main factors limiting the scalability of multi-core processors are high storage and on-chip bandwidth requirements of cache coherence protocols. Snoopy coherence protocols introduce large power overhead due to enormous amount of cache tag probing triggered by broadcast. Directory coherence protocols introduce large storage and performance overhead due to directory store and lookups. In addition, the efficiency of coherence protocols depends on the network topology. Traditional topologies, such as Mesh, face challenge in performance scalability. For example, core-to-core communications require up to 64 hops for a 1024-core mesh network, making it prohibitively expensive to support indirection in directory protocols.

I have been exploring following ideas, which can be leveraged to design more scalable cache coherence protocols:

- Initial study of some commercial and scientific applications show that there is strong correlation between instructions and the sharing profile of memory blocks. We are trying to leverage this correlation to predict the sharing of memory blocks. The sharing information can be used to optmize directory coherence protocols and reduce the storage and on-chip bandwidth requirements.

- Certain computations do not have to be 100% accurate e.g., data mining, clustering and rendering. Many programs compute on approximate and/or probabilistic data that are highly tolerant of error. Such computations can tolerate significant corruption and still produce meaningful and useful results. We are trying to explore whether such behavior can be exploited systematically to simplify coherence? Past research has leveraged the error tolerance of these applications to provide increased fault-tolerance and design low power arithmetic logic units.

Sujith Thomas

Supervisor : Prof. Harish Karnick

During the academic year 2011-12 I have been working on proposing a computational model for the representation of 'word' concepts occurring in a boolean world. By boolean world we mean a simulated world where every object has boolean attributes. The computational model we propose must exhibit properties like 'Fast mapping' and 'Typicality effect' that are common among humans. In addition to these our model must also provide a definition for a word concept 'W' from a set of positive and negative examples. The computational model we are trying to explore uses data mining techniques of 'closed set' mining and 'association rule' mining over data streams. Our model should be able to provide a DNF formula for a 'word' concept given a set of positive and negative examples. Learning of a 3-Term DNF from a set of positive and negative examples has been shown to be intractable. But our model tries to narrow down the search space by focusing only on the 'Frequent Closed Sets' encountered in the boolean world. The initial results we obtained was encouraging [1] and was submitted to CogSci'12 for the reviewers' comments. We are currently trying to improve our algorithm for learning N-Term DNF from a set of postive and negative examples.

Bibliography

[1] Sujith Thomas. Word learning using closed frequent sets of boolean features. Unpublished.

Sumit Kalra

Supervisor : not assigned

Course Work

I joined PhD program in December 2011 after my B.Tech. In academic year 2011-12, I have completed my first semester with following courses.
CS646 PARALLEL ALGORITHMS
CS687 ALGORITHMIC INFORMATION THEORY
CS618 INDEXING AND SEARCHING TECHNIQUES IN LARGE DATABASES
CS785 MULTIAGENT SYSTEMS : GAMES, ALGORITHMS, AND MULTIGENT SYSTEM

Project Work

1. Individual vs Society: Riding Bicycle to Work
   In this project, I analyzed the problem of how the scoiety and people in surrounding affect individual's decision making in daily life where those decision are not directly related to society itself. I have used some survey done by US State Govt. on Bicycle usage and its effects.

2. Splitting Free Disk Based Octree Structure for Indexing of Large Uniform Database
   In this project, we propsed a new algorithm for indexing in very large scale databases, which doesn't require any splittiing and other costly mechanism, and performs equally well for insertion and point search operation. In theoritical analysis of algorithm, we found that it should work better than existing state of the art algorithm for indexing in very large scale databaes.

Surya Prakash

Supervisor : Prof. Phalguni Gupta

Empirical study on 10,000 human ear images has observed the evidence of the uniqueness of the ear [3] which has motivated researchers to use ear for human recognition. Ear based system is one of the few biometric systems which can manage some of the critical issues generally occurred in any trait. Ear provides stable characteristics over the age. It is found to be invariant to different facial expressions and is unaffected by cosmetics. Background of the ear is predictable as it always exists at the middle of the profile face. Ear has an option to be used in a stand alone system or it can be easily combined with the face to design an efficient recognition system.

Three-dimensional (3D) images are being used in biometrics as they offer resilience to problems such as sensitivity towards pose, illumination and scale variations, common in two-dimensional (2D) images. Further, cost of 3D scanners has been drastically reduced. As a result, efforts are being made to design recognition systems based on 3D images.

In spite of ear having numerous rich characteristics as compared to other rival biometric traits, poor accuracy of a 2D ear recognition system has kept it away from practical applicability. However, the use of 3D ear images has helped in enhancing the recognition accuracy and recently there exist many techniques based on 3D ear images [2, 9, 7, 4, 5, 6]. Most of these techniques directly make use of ICP algorithm to match the 3D data of the ears. Since performance of ICP based matching algorithm highly depends on the initial state of the two data sets to be matched, these recognition techniques do not give good alignment in most of the cases.

In this work we propose a complete human recognition system based on 3D ear. The system is capable of automatically locating the ear in the profile face 3D range image and to use it further for recognition. The ear localization technique is invariant to rotation and scale. It makes use of graph connected components constructed using the edges of the depth map image of the range data for ear localization. It can detect left and right ears at the same time without imposing any additional cost or specific training.

The recognition technique uses 3D ear data along with their respective 2D images for performing the task. The technique first coarsely aligns the 3D ear data using local features computed from registered 2D ear image and then uses ICP based matching technique for final alignment. Coarse alignment of the data before applying ICP helps in providing a good initial starting point for ICP based matching algorithm. The technique integrates Generalized Procrustes Analysis (GPA)[8] with ICP for robust 3D ear matching. The recognition system has been tested on University of Notre Dame 3D profile face database, Collection J2 (UND-J2) [1] having range images with scale and pose variations.

Related Publication

An Efficient Technique for Ear Detection in 3D: Invariant to Rotation and Scale, **Surya Prakash** and Phalguni Gupta,, IAPR/IEEE International Conference on Biometrics (ICB 2012), New Delhi, India, March-April 2012.

Travel Support by Research-I foundation

I received travel support to attend International Conference on Biometrics (ICB 2012) at New Delhi, India from March 29 to April 1, 2012.

References

[1] University of Notre Dame Profile Face Database, Collection J2. http://www.nd.edu/~cvrl/CVRL/DataSets.html

[2] Chen, H., Bhanu, B.: Human ear recognition in 3D. IEEE Transactions on Pattern Analysis and Machine Intelligence **29**(4), 718–737 (2007)

[3] Iannarelli, A.: Ear Identification. Paramount Publishing Company, Freemont California (1989)

[4] Islam, S.M., Davies, R., Mian, A.S., Bennamoun, M.: A fast and fully automatic ear recognition approach based on 3D local surface features. In: Proc of 10th International Conference on Advanced Concepts for Intelligent Vision Systems (ACIVS '08), pp. 1081–1092 (2008)

[5] Islam, S.M.S., Bennamoun, M., Mian, A.S., Davies, R.: A fully automatic approach for human recognition from profile images using 2D and 3D ear data. In: Proceedings of 4th International Symposium on 3D Data Processing, Visualization and Transmission (3DPVT '08), pp. 131–135 (2008)

[6] Islam, S.M.S., Davies, R., Bennamoun, M., S. Mian, A.: Efficient detection and recognition of 3D ears. International Journal of Computer Vision **95**(1), 52–73 (2011)

[7] Passalis, G., Kakadiaris, I., Theoharis, T., Toderici, G., Papaioannou, T.: Towards fast 3D ear recognition for real-life biometric applications. In: Proceedings of IEEE Conference on Advanced Video and Signal Based Surveillance (AVSS 2007), vol. 3, pp. 39–44 (2007)

[8] Schonemann, P., Carroll, R.: Fitting one matrix to another under choice of a central dilation and a rigid motion. Psychometrika **35**(2), 245255 (1970)

[9] Yan, P., Bowyer, K.: Biometric recognition using 3D ear shape. IEEE Transactions on Pattern Analysis and Machine Intelligence **29**(8), 1297–1308 (2007)

Tejas Gandhi

Supervisor : Prof. Piyush Kurur

Thesis

I had joined the Phd programme in July 2011 and finished my course work requirements in the first semester. Currently, I am working with Prof Piyush Kurur in the area of Quantum error correcting codes. We are working over a specific class of codes know as quantum stabilizer codes. We aim to construct an efficient quantum stablizier codes from algebraic geometric codes.

Umarani Jayaraman

Supervisor : Prof. Phalguni Gupta

Fingerprint recognition system is used to identify a subject (human) from a large biometric database. One can do this task by searching a query image (henceforth termed as *query fingerprint*) against all images in the database (henceforth termed as *model fingerprints*) of subjects. Generally, the process to retrieve each model fingerprint from the database and to compare it against the query fingerprint for a match is computationally inefficient. A fingerprint image has the following characteristics:

- Number of minutiae extracted from a fingerprint of any subject at any two time instants may not be same.
- There may be too many minutiae in a fingerprint; Some of them may be false. Also, there is a possibility of missing some true minutiae.
- There may be partial occlusion in a fingerprint of a subject and it may overlap with some other subjects that are not present in the database.
- A query fingerprint may be rotated and translated with respect to the corresponding model fingerprints in the database.

Existing fingerprint indexing techniques can be classified on the basis of the methods of extracting features such as singular points [11], directional field [4], local ridge-line orientations [2, 5, 7, 8, 14], orientation image [9, 13], minutiae [10, 15, 16], minutiae descriptor [6], multiple features [3], matching scores [1, 12] and SIFT features [17]. But most of the matching algorithms are based on minutiae; so use of minutiae to index the fingerprint database is beneficial in many respects.

An efficient indexing technique for fingerprint database using minutiae based geometric hashing has been proposed. This technique consists of two stages, known as *indexing* and *searching*. For an accurate match at the time of searching, it has proposed a fixed length feature vector built from each minutia, known as Minutia Binary Pattern. Unlike any existing geometric based indexing technique, the proposed technique inserts each minutia along with the feature vector exactly once into a hash table. As a result, it reduces both computational and memory costs. Since minutiae of all fingerprint images in the database are found to be well distributed into the hash table, no rehashing is required. Experiments over FVC 2004 datasets prove the superiority of the proposed indexing technique against well known geometric based indexing techniques.

Related Publication

Minutiae Based Geometric Hashing for Fingerprint Database, **Umarani Jayaraman**, J. Viswanathan, Aman K. Gupta, Phalguni Gupta, International Conference on Intelligent Computing (ICIC 2012), China July 2012.

References

[1] Aglika, G., Arun, R.: A Novel Coding Scheme for Indexing Fingerprint Patterns. In: Proceedings of the Joint IAPR International Workshop on Structural, Syntactic, and Statistical Pattern Recognition, pp. 755–764 (2008)

[2] Alessandra, L., Dario, M., Davide, M.: Continuous Versus Exclusive Classification for Fingerprint Retrieval. Pattern Recognition Letters **18**(10), 1027–1034 (1997)

[3] de Boer, J., Bazen, A.M., Gerez, S.H.: Indexing Fingerprint Databases Based on Multiple Features. In: Proceedings of the 12th Annual Workshop on Circuits, Systems and Signal Processing, pp. 300–306 (2001)

[4] Cappelli, R., Maio, D., Maltoni, D.: Indexing Fingerprint Databases for Efficient 1:N Matching. In: Proceedings of International Conference on Control, Automation, Robotics and Vision, Invited Paper (2000)

[5] Cappelli, R., Maio, D., Maltoni, D.: A Multi-Classifier Approach to Fingerprint Classification. Pattern Analysis and Application. **5**(2), 136–144 (2002)

[6] Feng, J.: Combining Minutiae Descriptors for Fingerprint Matching. Pattern Recognition. **41**(1), 342–352 (2008)

[7] Jiang, X., Liu, M., Kot, A.C.: Fingerprint Retrieval for Identification. IEEE Transactions on Information Forensics and Security **1**(4), 532–542 (2006)

[8] Lee, S.O., Kim, Y.G., Park, G.T.: A Feature Map Consisting of Orientation and Interridge Spacing for Fingerprint Retrieval. In: Proceedings of International Conference on Audio and Video Based Biometric Person Authentication, pp. 184–190 (2005)

[9] Li, J., Yau, W.Y., Wang, H.: Fingerprint Indexing Based On Symmetrical Measurement. In: Proceedings of International Conference on Pattern Recognition, pp. 1038–1041 (2006)

[10] Liang, X., Asano, T., Bishnu, A.: Distorted Fingerprint Indexing Using Minutiae Detail and Delaunay Triangle. In: Proceedings of the 3rd International Symposium on Voronoi Diagrams in Science and Engineering, pp. 217–223 (2006)

[11] Liu, T., Zhu, G., Zhang, C., Hao, P.: Fingerprint Indexing Based on Singular Points. In: Proceedings of International Conference on Image Processing, pp. 293–296 (2005)

[12] Maeda, T., Matsushita, M., Sasakawa, K.: Characteristics of the Identification Algorithm Using a Matching Score Matrix. In: Proceedings of First International Conference on Biometric Authentication, pp. 330–336 (2004)

[13] Manhua, L., Xudong, J., Chichung, K.A.: Fingerprint Retrieval by Complex Filter Responses. In: Proceedings of International Conference on Pattern Recognition, pp. 1042–1048 (2006)

[14] Raffaele, C., Alessandra, L., Dario, M., Davide, M.: Fingerprint Classification by Directional Image Partitioning. IEEE Transactions on Pattern Analysis and Machine Intelligence **21**(5), 402–421 (1999)

[15] Raffaele, C., Matteo, F., Davide, M.: Fingerprint Indexing Based on Minutia Cylinder-Code. IEEE Transactions on Pattern Analysis and Machine Intelligence **33**(5), 1051–1057 (2011)

[16] Ratha, N.K., Karu, K., Chen, S., Jain, A.K.: A Real-time Matching System for Large Fingerprint Databases. IEEE Transactions on Pattern Analysis and Machine Intelligence **18**(8), 799–813 (1996)

[17] Shuai, X., Zhang, C., Hao, P.: Fingerprint Indexing Based on Composite Set of Reduced SIFT Features. In: Proceedings of International Conference on Pattern Recognition, pp. 1–4 (2008)

J. Vishwesh

Supervisor : Prof. Sanjeev K Aggarwal and Prof. Amey Karkare

I have joined as a PhD student in CSE department in July 2011. During the last one year I have pursued 8 courses. And I am still in the process of completing my course requirement. My thesis supervisor is Prof.Sanjeev K Aggarwal and Co-supervisor is Dr.Amey Karkare. My current research work is on Parallelizing LINQ for heterogeneous architecture.

These are the following courses and projects that I have completed during the last year.

1. CS601: Fundamentals of CSE: Maths for CS and Algorithms.

   In this course the concepts of logic,linear algebra and probability are covered in fundamental of Maths and fundamental of algorithm paradigms are revised.

2. CS641: Modern Cryptology

   We studied about two types of encryption and decryption algorithms i.e, Public key and Private key algorithms along with their cryptanalysis techniques. And also studied advanced topics on security.

3. CS727: Topics in Internet Technologies.

   This course gave an exposure to several topics in the current internet technologies.
   In this course I have done term project on 'Semantic Bug Search' here the objective is to semantically search for similar bug in a bug database. In this project I have restricted myself to hierarchical relationship among the objects.

4. CS645: Topics in Design and Analysis of Algorithms.

   In this course we have studied various topics on Computational geometry algorithms, advanced data structures and algorithms for optimizing minimum spanning tree etc.

5. CS738: Advanced Compiler Optimizations.

   This course has discussed advanced topics on compilers such as data flow analysis, SSA form, monotonicity of data flow analysis and also on Pointer analysis, Inter procedural and Heap area analysis. Also I have done a project in this course on 'Implementing Higher Level Loop Optimizations' here I have implemented loop optimizations such as loop interchange, skewing,reversal and selection algorithms and extracted the parallelism from the loop..

6. CS618: Indexing and Searching Techniques in Databases.

   Here we studied about indexing and searching techniques for single and multi dimensional database and also discussed on several data structures that improve the performance of database queries.
   The term project was done on 'Temporal RDF' - the objective here is to introduce time component into a RDF graph using R* tree and I have developed a semantic issue tracking system based on this concept.

7. CS797: Special Topics in Computer Science

   Here I have chosen the topic of self study on 'Cloud computing'. I have focused on the concept of cloud computing and techniques for virtualization of various resources such as CPU ,memory ,I/O and

network.
And also carried out a term project on 'Benchmarking for Cloud Infrastructure' here the I have evaluated difference between the performance results observed in various resources of host machine and virtual machine by using the benchmarks.

8. CS629: Parallel Execution of Programs.

   The objective of this course is to extract the parallelism from a sequential program using dependency analysis. And also discussed on several transformation techniques to extract the parallelism from the loop bodies.

I have started my thesis work during summer.Currently I am focusing on the transformation techniques to extract parallelism from a program so that it can be efficiently executed on heterogeneous architecture.