Research I Foundation

Annual Report

2009-2010

Contents

Expenditure Sheet for 2009-2010

Expenditure Sheet for April 2009- March 2010

Expenditure incurred towards Travel Allowance	Rs. 4235114
Expenditure incurred towards IITK Workshop	Rs. 3000000
Expenditure incurred towards Salary	Rs. 2075333
Expenditure incurred towards Honorarium	Rs. 1340000
Expenditure incurred towards Equipments	Rs. 412750
Expenditure incurred towards Contingency	Rs. 417382
Expenditure incurred towards consumables	Rs. 25760
Total Expenditure incurred	**Rs. 11506339**

Research Activities of Faculty Members

Prof. Anil Seth

Temporal Logics for Interacting Process Classes

P. S. Thiagarajan (Nat. Univ. of Singapore) and Anil Seth (IIT Kanpur)

In this project the IIT Kanpur author of the project worked on model checking of extensions of pushdown systems such as higher order pushdown systems (*hpds*) and multi-stack pushdown systems (*mpds*). This work is summarized below.

Higher order pushdown systems (*hpds*) generalize order-1 pds in that they can have nested stacks. An order-2 pds has a finite set of control states and a stack of order-1 stacks. Order$-n$ pds have a stack of nested depth n. Operations to copy the topmost order$-i, i \leq n$, stack and to pop it are provided on order$-n$ pds. We show that for any n, the set of winning configurations of a player in a parity game over an order$-n$ pushdown system is regular. More specifically, we extend the approach of Walukiewicz and Cachat from order-1 pds, to give one step reduction from parity games over configuration graphs of order$-n$ pds to parity games over finite graphs. Our reduction is based on a careful analysis of copying of stacks and matching push and pop operations of hpds. We use higher order functions in our reduced finite game to capture popping conditions for various pop operations. The proof of correctness also gives a uniform way to execute a winning strategy by an order-n pushdown automaton in the entire winning region of a player in an order$-n$ pds parity game. This result was proved independently of the work of Carayol, Hague, Meyer, Ong and Serre (LICS 2008). Our proof approach is different from theirs. Our proof can also be extended to collapsible pushdown systems.

A multi-stack pushdown system (*mpds*) has a finite set of control states and a fixed number of stacks. The transition function of a *mpds* takes as input its control state and the topmost symbols of each stack and may (nondeterministically) do a push or a pop operation on any stack along with a possible change in control state of *mpds*. Multi-stack pushdown systems can be used to model a class of programs with recursion and threads. Each thread has its own stack for its procedures calls and communication among threads is through the common finite states of *mpds*.

Bounded phase multi-stack pushdown automata have been studied recently. In our work we show that parity games over bounded phase multi-stack push-down systems are effectively solvable and winning strategy in these games can be effectively synthesized. We show some applications of our result, including a new proof of a known result that emptiness problem for bounded phase multi-stack automata is decidable. It may also be noted that showing parity games effectively solvable implies decidability of many temporal logics on the configuration graphs of *mpds*.

We also show that bounded phase *mpds* also admit global reachability analysis. Given a set $\mathcal{C}$ of configurations of a *mpds* $\mathcal{M}$, let $pre^*_{\mathcal{M}}(\mathcal{C}, k)$ be the set of configurations of $\mathcal{M}$ from which $\mathcal{M}$ can reach an element of $\mathcal{C}$ in at most k phases. We show that for any *mpds* $\mathcal{M}$, any regular set $\mathcal{C}$ of configurations of $\mathcal{M}$ and any number k, the set $pre^*_{\mathcal{M}}(\mathcal{C}, k)$, is regular. We use saturation like method to construct a non-deterministic finite multi-automaton $\mathcal{A}_{pre,k}$ recognizing $pre^*_{\mathcal{M}}(\mathcal{C}, k)$. Size of the automaton constructed is double exponential in k which is optimal as the worst case complexity measure. We also give an algorithm which from any accepting run of $\mathcal{A}_{pre,k}$ on configuration d constructs a witnessing execution sequence of transitions of $\mathcal{M}$ starting at d and ending at some $e \in \mathcal{C}$.

We also define higher order multi-stack pushdown systems and show that parity games over bounded phase higher order multi-stack pushdown systems are effectively solvable and winning strategy in these games can be effectively synthesized. This work is non-trivial extension of the order-1 case. It also implies decidability of emptiness problem for bounded phase higher order multi-stack automata and decidability of temporal logics on configuration graphs of *hmpds* in the same way as for *mpds* case.

From these results, it seems that the technique of Walukiewicz and Cachat can be generalized to a unified framework for obtaining decidability of μ-calculus (or equivalently solving parity games) on many variants of pushdown systems.

Following publications and manuscripts resulted from this work.

Publications

[1] Anil Seth. Another Proof of Regularity of Winning Regions in Parity Games over Higher Order Pushdown Automata. Manuscript, April 2008.

[2] Anil Seth. Games on Higher Order Multi-stack Pushdown Systems. In *Third InternationalWorkshop on Reachability Problems (RP09)*, volume 5797 of *LNCS*, pages 203–216, Palaiseau, France, September 23-25, 2009.

[3] Anil Seth. Games on Multi-Stack Pushdown Systems. In Sergei N. Artëmov and Anil Nerode, editors, *Logical Foundations of Computer Science*, volume 5407 of *LNCS*, pages 395–408. Springer, Heidelberg, 2009.

[4] Anil Seth. Global Reachability in Bounded Phase Multi-Stack Pushdown Systems. In *22nd International Conference on Computer Aided Verification (CAV 2010)*, Edinburgh, UK, July 15-19, 2010.

Prof. Piyush P. Kurur

The complexity of Matrix Multiplication

Firstly I thank the Research I Foundation for their generous funding. A summary of work done as part of the project NRNM/CS/20030163.

One of the main goals of this project was understanding the complexity of matrix multiplication. Cohen and Umans [2] recently introduced a representation theoretic approach to this problem. Although, we were unable to improve the state of art of the matrix multiplication problem, we managed to prove two results where representation theoretic approach helped. We gave a modular arithmetic based algorithm for fast integer multiplication which is currently the fastest algorithm. It is a modular variant of the algorithm given earlier by Fürer [5]. A crucial step in our algorithm is the use of multivariate Fourier transform which is the representation theory connection that we exploit here.

The next result [4] of ours is to understand graph isomorphism from a representation theoretic background. We say that a group G is *representable* over a graph X if there is a *non-trivial* homomorphism form G to the automorphism subgroup $\mathrm{Aut}(X)$ of X. In this context, notice that a representation of a group G on a vector space V is a homomorphism from G to its automorphism group $\mathrm{GL}(V)$. The graph representability problem is the problem of deciding given a group G and a graph X, whether G is representable on X. We show that the graph isomorphism problem reduces to the abelian graph representability problem. On the other hand solvable representability problem is reducible to graph isomorphism problem. Thus, as long as we restrict attention to groups that are solvable, we get only the power of graph isomorphism. What exactly is the complexity of representability problem in general is open. We believe that it is harder than graph isomorphism. Specifically we show that representability problem on trees is as hard as permutation representability problem: Given a finite group G as an explicit table and an integer n (in binary) the problem of *permutation representability problem* is to check if G is homomorphic to a subgroup of S_n. This problem appears hard. On the other hand Tree isomorphism is in P.

References

[1] Peter Bürgisser, Michael Clausen, and M. Amin Shokrollahi. *Algebraic Complexity Theory*. Springer Verlag, 1997.

[2] Henry Cohn and Christopher Umans. A group-theoretic approch to fast matrix multiplication. In *44th Annual IEEE Symposium on Foundations of Computer Science*, page 438, 2003.

[3] Anindya De, Piyush P Kurur, Chandan Saha, and Ramprasad Saptharishi. Fast integer multiplication using modular arithmetic. In *40th ACM Symposium on Theory of Computing*, 2008.

[4] Sagarmoy Dutta and Piyush P Kurur. Representing Groups on Graphs. In *34th International Symposium on Mathematical Foundations of Computer Science (MFCS)*, pages 295–306, 2007.

[5] Martin Fürer. Faster integer multiplication. In 39^{th} *ACM Symposium on Theory of Computing*, pages 57–66, 2007.

Prof. R. K. Ghosh

Report on Research-I project

1. Some breakthrough has been achieved in the ongoing work on modeling energy aware routing in wireless sensor networks (WSN). Two students were worked this problem during 2009-10. We were able to establish some preliminary lower bound which matched with simulation experiments. One M. Tech student is continuing this work and I think we should be able to write a paper soon on the work done so far on this problem. We are also developing energy aware routing algorithms on the basis of the energy lower bound result.

2. As indicated in 2008-09 progress report, I spent a semester during sept-dec, 2008 at University of Texas at Arlington as a visiting faculty. During this time we worked on sensor localization. As a follow-up to this I received a research funding amounting to 40000 Euros from European Aeronautics Defence and Space Company (EADS).We started some initial work. Already one publication has accrued from this work in the form of chapter in a forthcoming book on Theoretical Aspects of Distributed Computing in Sensor Networks (Springer Verlag).

3. I also pursued the other thread of investigation related to the problem of tracking intruders. As a follow up work of the earlier paper on tracking targets, recently we were able to get some good results. Two different methods for tracking target using deployment knowledge of sensors have been proposed, and extensive simulations were done using NS-2 and MATLAB. These results showed significant improvements over existing tracking methods. We have just submitted a paper in COMSNETS based on this work.

4. Another thread of investigation mentioned in 2009-10 report, relates to use of mobile agent for data dissemination in a WSN. We first built light-weight protocol mobile agent based computing with intention to use for routing in WSN. But I decided to discontinue this line of research, as I think agent infrastructure is unsuitable for WSN environment. But as substantial work was done, I oriented the work on application of mobile agents in mobile services, and wrote it in a tutorial format. The work has been accepted to appear as chapter in the forthcoming book on Application and services for Mobile Systems (Taylor and Francis).

5. I started a new thread of investigation for downstream routing on WSNs. I idea was to replace dissemination based downstream routing by a scheme which would allow unicast routing to specific sensor from any IP node which is connected to base-station. The first idea we were able to develop is the one in which an IP node can send message to individual sensor nodes via its base-station using a DSR type source routing scheme. We also designed a toy application where a sensor node is instructed to raise alarm when its monitoring data exceeds a threshold. The first phase of the protocol has been completed and the student who worked on this problem has completed his thesis. However, as we have developed it as source based scheme, the message can reach sensors upto 8-9 hops from base-station. The problem is now being re-looked by developing a single hop broadcast based routing scheme. The protocol is now being simulated.

6. Another new thread of investigation started recently relates to monitoring structural parameters of bridges by placement of a wireles distributed network consisting of a combination of WSN and WiFi nodes. The idea came up while working on the downstream routing for sending data from IP node to any targeted sensor node. We have formulated a scheme whereby structural health of unattended railway bridges can be monitored through WSN, and the monitored data can be gathered by passing trains fitted with WiFi nodes and also through GPRS network when needed. Currently an M. Tech student is working on the problem.

7. The other problem of fringe interest, which I have started working on, is to ensure secure communication in WSN. A scheme have been developed for one hop key management using deployment knowledge. It will not require any master key. The keys are pre-stored before deployment, but if some pairs of nodes are unable to find a pair of shared key, shared keys can be established through a handshaking protocol which require only few message exchanges. The scheme has been analyzed theoretically and also simulated over TOSSIM. A preliminary paper based on the idea has been submitted to HiPC workshop.

8. Finally, we are also working on analysis of energy efficient MAC protocols for WSN. A comprehensive survey on existing protocols is nearing completion. We have been able to simulate all CSMA based MAC protocols over TOSSIM (no such study exists at present). The results are now being assembled together for a realistic comparison. Preliminary results appears to suggest PMAC scores over other protocols. We have started a theoretical analysis to back the experimental findings.

Prof. Shashank K. Mehta

Representation for Cyclotomic Fields and Their Subfields

Let G be a finite group and $n \in \mathbb{Z}^+$. A matrix representation of G is any homomorphism from G into $GL_n(\mathbb{F})$, where $GL_n(\mathbb{F})$ is the group of invertible $n \times n$ matrices with entries from the characteristic zero field $\mathbb{F}$. The representation is called faithful if the image of the homomorphism is ismorphic to G. Similarly question arises whether an extension field of a field $\mathbb{F}$ has a representation in $M_n(\mathbb{F})$, the ring of $n \times n$ matrices over $\mathbb{F}$. The simple extension of a field $\mathbb{F}$ obtained by adjoining α, algebraic over $\mathbb{F}$, is isomorphic to $\mathbb{F}[A]$ where A is the companion-matrix of the minimal polynomial of α, and $\mathbb{F}[A]$ is the set of all polynomials in A over $\mathbb{F}$. Consequently every element in $\mathbb{F}(\alpha)$ is represented by a matrix, which is a polynomial in A.

Now for a given finite simple extention $\mathbb{F}(\alpha)$ over $\mathbb{F}$ the questions which arise naturally are

1. does there exists a matrix A in $M_n(\mathbb{F})$ with some specified properties such that $\mathbb{F}[A] \cong \mathbb{F}(\alpha)$?
2. if it exists, what is the smallest possible n ?

For example, does there exists a matrix A which is either a circulant over $\mathbb{Q}$ or a 0-1 companion-matrix such that $\mathbb{Q}[A] \cong \mathbb{Q}(\zeta_n)$? (where ζ_n is the primitive n-th root of unity). Clearly this is true only when $n = 1$ *or* 2. But for $n > 2$, consider the matrix

$$W_n = \begin{bmatrix} 0 & 0 & \dots & 0 & 1 \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \end{bmatrix},$$

which is a circulant-matrix as well as the companion-matrix of $x^n - 1$ with 0-1 entries. Further, $\Phi_n(x)$ the n-th cyclotomic polynomial, which is the minimal polynomial of ζ_n devides the minimal polynomial of W_n. We will see that $\mathbb{Q}[W_n]/\langle \Phi_n(W_n) \rangle \cong \mathbb{Q}(\zeta_n)$. In this case we say that the pair $(W_n, \Phi_n(W_n))$ represents the field $\mathbb{Q}(\zeta_n)$.

In general, we say that the pair of $n \times n$ matrices (A, B) over $\mathbb{F}$ represents an extension field $\mathbb{K}$ if $\mathbb{K} \cong \mathbb{F}[A]/\langle B \rangle$ where $\mathbb{F}[A]$ denotes the smallest subalgebra of $M_n(\mathbb{F})$ containing A and $\langle B \rangle$ is an ideal in $\mathbb{F}[A]$ generated by B. It can be easily shown that if $\mathbb{K} = \mathbb{F}(\alpha)$ is a simple extension and $A \in M_n(\mathbb{F})$ is a matrix with α as an eigenvalue, then $\mathbb{K} \cong \mathbb{F}[A]/\langle q(A) \rangle$, where $q(x)$ is the minimal polynomial of α in $\mathbb{F}[x]$. Hence finiding a representation of $\mathbb{F}(\alpha)$ is equivalent to finiding a matrix A which has α as an eigenvalue and hence we also say that A represents $\mathbb{F}(\alpha)$. Here it needs to be pointed out that if $\mathbb{K} = \mathbb{F}(\alpha) = \mathbb{F}(\beta)$, then the matrices having α as an eigenvalue as well as the matrices having β as an eigenvalue give representation for $\mathbb{K}$. But these matrices are distinct in general.

In his paper Godsil proved that every algebraic integer occurs as an eigenvalue of the adjacency matrix of some Cayley digraph. Equivalently every finite extention over $\mathbb{Q}$ can be represented by the adjacency matrix of some Cayley digraph. We end this section by stating a few relavant facts from field theory.

Elementary facts from field theory

In this paper, we will be concerned with fields that have characteristic 0. For any zero characteristic field $\mathbb{F}$, $\mathbb{Q} \subseteq \mathbb{F} \subseteq \mathbb{C}$, where $\mathbb{Q}$ is the field of rationals and $\mathbb{C}$ is the field of complex numbers. An element $\alpha \in \mathbb{C}$ is said to be algebraic over a subfield $\mathbb{F}$, if α is a root of a polynomial $f(x) \in \mathbb{F}[x]$. $f(x)$ is said to be the minimal polynomial of α over $\mathbb{F}$, if α is a root of $f(x)$ and $f(x)$ is monic and is irreducible in $\mathbb{F}[x]$.

Lemma 1 (Dummit & Foote, Page 497)**.** *Let $f(x) \in \mathbb{F}[x]$. Then $\mathbb{F}[x]/\langle f(x)\rangle$ is a field if and only if $f(x)$ is irreducible in $\mathbb{F}[x]$. Furthermore, if α is a root of the irreducible polynomial $f(x) \in \mathbb{F}[x]$ then the field $\mathbb{F}(\alpha)$ coincides with the ring $\mathbb{F}[\alpha]$ which is isomorphic to $\mathbb{F}[x]/\langle f(x)\rangle$.*

Therefore, by Lemma 1, for any $\alpha \in \mathbb{C}$ with minimal polynomial $q(x)$ over $\mathbb{F}$, one has $\mathbb{F}(\alpha) \cong \mathbb{F}[x]/\langle q(x)\rangle$. Let $\mathbb{F}$ be a subfield of $\mathbb{K}$. Then $\mathbb{K}$ is a vector space over $\mathbb{F}$. If the dimension of this vector space is finite then $\mathbb{K}$ is said to be a finite extension of $\mathbb{F}$ and its dimension is denoted by $[\mathbb{K} : \mathbb{F}]$. In case $\mathbb{K} = \mathbb{F}(\alpha)$ then $[\mathbb{F}(\alpha) : \mathbb{F}]$ is the degree of the minimal polynomial of α over $\mathbb{F}$. Two algebraic elements α and β over $\mathbb{F}$ are said to be conjugates over $\mathbb{F}$ if they have the same minimal polynomial over $\mathbb{F}$. A polynomial $f(x) \in \mathbb{F}[x]$ is said to be separable if all its roots are simple (multiplicity one). It is well known that all the irreducible polynomials over a field of characteristic 0 are separable. Following is another important result.

Lemma 2 (Dummit & Foote, Page 493)**.** *Let α be algebraic over a characteristic zero field $\mathbb{F}$ with minimal polynomial $q(x) \in \mathbb{F}[x]$ of degree n.*

1. *Then each element of $\beta \in \mathbb{F}(\alpha)$ has a unique expression as $\beta = a_0 + a_1\alpha + \cdots + a_{n-1}\alpha^{n-1}$ with all $a_i \in \mathbb{F}$. Observe that β is a polynomial of degree at most $n-1$ in α.*
2. *Then for any $\gamma \in \mathbb{F}(\alpha), \gamma \neq 0$, $\mathbb{F}[\gamma]$ is a field. Moreover, the degree of the extension, $\big[\mathbb{F}[\gamma] : \mathbb{F}\big]$ divides $n = \big[\mathbb{F}(\alpha) : \mathbb{F}\big]$. In particular, $\mathbb{F}(\alpha) = \mathbb{F}(a\alpha + b)$ for any $a(\neq 0), b \in \mathbb{F}$.*
3. *Let α and β be conjugates over $\mathbb{F}$. Then there exists a field isomorphism $\phi : \mathbb{F}(\alpha) \longrightarrow \mathbb{F}(\beta)$ that fixes $\mathbb{F}$ and sends α to β. Hence, for any $g(x) \in \mathbb{F}[x]$, $g(\alpha)$ and $g(\beta)$ are also conjugates over $\mathbb{F}$. That is, if $\alpha_1, \alpha_2, \ldots, \alpha_n$ are the conjugates over $\mathbb{F}$, then for any $g(x) \in \mathbb{F}[x]$, $g(\alpha_1), g(\alpha_2), \ldots, g(\alpha_n)$ are also conjugates, not necessarily distinct, over $\mathbb{F}$. Also, the degree of the minimal polynomial of $g(\alpha)$ over $\mathbb{F}$ divides* $\deg(q(x))$.

Let $\mathbb{F}$ be a field. An extension $\mathbb{K}$ of $\mathbb{F}$ is said to be simple if there exists an element $\alpha \in \mathbb{C}$ such that $\mathbb{K} = \mathbb{F}(\alpha)$. Let ζ_n denote a primitive n-th root of unity. Field $\mathbb{Q}(\zeta_n)$ is called the n-th cyclotomic field. Since $\mathbb{Q}(\zeta_n)$ has characteristic zero, any of its subfields is of the form $\mathbb{Q}\big(f(\zeta_n)\big)$ for some $f(x) \in \mathbb{Q}[x]$.

Main Accomplishments of this Project

1. We showed that any subfield of a cyclotomic field can be represented by some circulant-matrix over $\mathbb{Q}$ and conversely every circulant-matrix over $\mathbb{Q}$ represents a subfield of a cyclotomic field. Further we also characterize the smallest such representations.
2. We showed that some real subfields of $\mathbb{Q}(\zeta_n)$ can be represented by a polynomial in adjacency matrix of cyclic graph. Consequently those real subfields of $\mathbb{Q}(\zeta_n)$ has integer symmetric circulant-matrix representation.
3. If p is any prime and $\mathbb{F}$ is a subfield of a cyclotomic extension $\mathbb{Q}(\zeta_p)$ then we obtain a zero-one circulant-matrix A of size $p \times p$ such that $(A, \mathbf{J})$ represents $\mathbb{F}$, where $\mathbf{J}$ is the matrix with all entries 1.
4. Fix a positive integer n and let $\mathcal{A}_n$ be as defined earlier. Then $\mathcal{A}_n$ is empty set whenever $n = p^k$ for some prime p. For the above choice of n, $\Phi_n(x)$ divides $x^n - 1$ and it is the polynomial of least degree whose companion-matrix is a 0-1 matrix.

5. If n has exactly two distinct prime factors p_1 and p_2 then the degree of the polynomial of least degree in $\mathcal{A}_n$ is shown to be $\frac{n}{p_1 p_2}(p_1 p_2 - 1)$. If n is even and has at least three prime factors then a bound is obtained for the degree of the polynomial with least degree in $\mathcal{A}_n$. For the odd case the best upper bound we could find is $\frac{n}{p_1 p_2}(p_1 p_2 - 1)$. It is also established that if $n = p_1^{a_1} p_2^{a_2} \dots p_k^{a_k}$ is a prime factorization of n into distinct primes then $\min\{\deg(f(x)) : f(x) \in \mathcal{A}_n\} = \frac{n}{n_0} \cdot \min\{\deg(f(x)) : f(x) \in \mathcal{A}_{n_0}\}$ where $n_0 = p_1 p_2 \dots p_k$.

A paper based on this work has been submitted to Journal of Linear Algebra and Applications.

Prof. Sumit Ganguly

Data Stream Computation: Algorithms and Lower Bounds
Proposal funded by Research I Foundation 2007-2010

The data streaming model presents a computational model for a number of applications where data arrives rapidly and continuously and has to be processed in an online fashion. The rate and volume of data arrival makes it impossible to store the input in its entirety in the main memory, or to maintain index structures in secondary storage. Further, many applications of streaming data do not require exact answers to detail queries, rather, it often suffices to obtain a small amount of useful information or aggregate statistics of the stream. A data stream $\mathcal{S}$ is viewed as a sequence of updates the form $(index, i, v)$, where, $i \in [n] = \{1, \ldots, n\}$ and v depicts the change in the *frequency* of i and $v \in \{-M, -M+1, \ldots, M-1, M\}$. The frequency of an item i in the stream $\mathcal{S}$ is defined as $f_i(\mathcal{S}) = \sum_{(index,i,v) \text{ appears in } \mathcal{S}} v$. Let m denote the size of the stream, that is, the number of records. In this model, an algorithm is typically given a small amount of memory to summarize a large and often rapidly arriving dataset and is usually allowed a single pass (or at most a few passes) over the data. Algorithms must answer global queries over the data-set using only the data summary.

The project is concerned with the development of algorithms for data stream processing. The following works and activities were done as part of the project.

1. We consider the problem of estimating $F_p = \sum_{i \in [n]} |f_i|^p$ for $p \in (0, 2)$ to within approximation factor of $1 \pm \epsilon$ and with high constant probability (7/8). The work in [6] presented the first algorithm that had $O(\text{polylog}(1/\epsilon, n, m))$ update time while maintaining a space requirement of $O(\epsilon^{-2-p} \log^{O(1)}(n, m, M))$. All prior algorithms required update time that were $\Omega(1/\epsilon^2)$.

2. In [3], we improve the above solution by presenting an algorithm that estimates F_1 in time $O(\text{polylog}(1/\epsilon, n, m))$ but requires space that is $O(\epsilon^{-2} \log^{O(1)}(n, m, M))$. A similar algorithm was discovered independently by Kane, Nelson and Woodruff.

3. We develop a theory of stream automaton for characterizing deterministic stream computations over integer frequency vectors. We use it to study the problem of deterministically estimating frequencies (i.e., return $\hat{f}_i$ such that $|\hat{f}_i - f_i| \leq \epsilon f_i$). This is used to present stronger lower bounds $\Omega(\epsilon^{-2})$ as compared to the known lower bound $\Omega(\epsilon^{-1})$ for the problem of deterministically estimating frequencies to within an accuracy of ϵ [2].

4. We consider the issue of flexible massive parallelization of frequency dependent tasks, inspired by the popular MAP-REDUCE paradigm. Using the theory of stream automaton, we show that if there is a parallel computation tree with vectors at the leaf and the tree computes a total function of the sum of these vectors, then, every computation tree can be programmed to do the same [3]. This shows that the parallel computation of total function of distributed sum of vectors is fully flexible.

5. We present a novel algorithm for finding frequent items (i.e., return a set of items i such that $f_i \geq \epsilon L_p$ and do not return any item i such that $f_i < (\epsilon - \phi) L_p$, where, $L_p = (\sum_{i=1}^{n} |f_i|^p)^{1/p}$ and $p \in \{1, 2\}$.

We consider update data streams, that is, streams that allow arbitrary insertion and deletion of items. Our algorithm [8] improves upon the space requirement of $O(\phi^{-2})$ of the only existing algorithm of Cormode and Muthukrishnan to $O(\phi^{-1})$ for $p = 2$ and is experimentally shown to have better precision and recall.

6. We demonstrate a novel and simple paradigm for designing algorithms for data stream processing using expander graphs [1]. We pose the problem of $k - l$ separator over a data stream that decides whether the number of items i with $f_i \neq 0$ is at most k or is at least l. The $k-2k$-separator problem for streams with non-negative integer frequency vector is solved using an application of expander graphs. This is then used to present the first space-optimal solution for the k-sparsity problem (i.e., a $k - (k + 1)$ separator) for non-negative integer frequency vectors.

7. We consider the problem of hybrid frequency moments over two-dimensional streams. In a two dimensional stream, each stream record is of the form (i, j, v) signifying that $A_{i,j}$ is updated to $A_{i,j} + v$, where, $1 \leq i, j \leq n$. The p, q hybrid moment is defined as $F_{p,q} = \sum_{j=1}^{n} (\sum_{i=1}^{n} |A_{i,j}|^p)^q$. Estimation of hybrid moments finds applications in network monitoring. We present the first algorithm with space requirement that is poly-logarithmically dependent on the size of the matrix for the range $q \in [0, 1]$ and $p \in [0, 1]$ [4]. We also present an $O(n^{1-1/q})$ space algorithm for $q > 1$ and $p \in [0, 2]$. A fuller version of this work appears in the journal [5].

8. In [9], we study the d-dimensional knapsack problem in the data streaming model. The knapsack is modelled as a d-dimensional integer vector of capacities, whose capacities (in each dimension) is scaled to 1. There is an input stream of n items, each item is modelled as a d-dimensional integer column of non-negative integer weights and a scalar profit. The input instance has to be processed in an online fashion using sub-linear space. After the items have arrived, an approximation for the cost of an optimal solution as well as a template for an approximate solution is output. Our algorithm achieves an approximation ratio $(2(\frac{1}{2} + \sqrt{2d + \frac{1}{4}}))^{-1}$ using space $O(2^{O(d)} \cdot \log^{d+1} d \cdot \log^{d+1} \Delta \cdot \log n)$ where $\{\frac{1}{\Delta}, \frac{2}{\Delta}, \ldots, 1\}, \Delta \geq 2$ is the set of possible profits and weights in any dimension, and P is the ratio between the minimum and maximum prot. We also show that any data streaming algorithm for the $t(t-1)$-dimensional knapsack problem that uses space $o(\sqrt{\Delta}/t^2)$ cannot achieve an approximation ratio that is better than $1/t$. Thus, even using space Δ^γ, for $\gamma < 1/2$, i.e. space polynomial in Δ, will not help to break the $1/t \approx 1/\sqrt{d}$ barrier in the approximation ratio.

Other Activities. The author was invited to present a talk and participate in Dagsthul Seminar series on Sub-linear Algorithms (Seminar No. 8341) and to present a seminar on stream automaton at the University of Frankfurt, Germany. The conference papers listed below were presented at the respective conferences and the travel was funded by the Research I foundation.

References

[1] Sumit Ganguly. Data Stream Algorithms via Expander Graphs. In *International Symposium on Algorithms, Automata and Computation (ISAAC)*, page 5263, 2008.

[2] Sumit Ganguly. Lower bounds for frequency estimation over data streams. In *Computer Science Symposium of Russia (CSR)*, volume 5010, pages 204–215. Springer, 2008.

[3] Sumit Ganguly. Distributing Frequency-Dependent Data Stream Computations. In *Computing: the Australasian Symposium (CATS)*, Wellington, New Zealand, January, 2009.

[4] Sumit Ganguly, Mohit Bansal, and Shruti Dube. Estimating hybrid frequency moments of data streams. In *International Frontiers of Algorithms Workshop*, Changsha, Hunan, China, June, 2008.

[5] Sumit Ganguly, Mohit Bansal, and Shruti Dube. Estimating hybrid frequency moments of data streams. *J. Combinatorial Optimization*, August 2010. DOI 10.1007/s10878-010-9339-1.

[6] Sumit Ganguly and Graham Cormode. On Estimating Frequency Moments of Data Streams. In *International Workshop on Randomization and Computation (RANDOM)*, 2007.

[7] Sumit Ganguly and Purushottam Kar. Estimating F_1 of data stream in nearly optimal space and time. In *International Conference of Italian Computer Science*, Cameron, Italy, August, 2010.

[8] Sumit Ganguly, Abhayendra N. Singh, and Satyam D. Shankar. Finding frequent items over general update streams. In *International Conference on Scientific and Statistical Database Management (SSDBM)*, July, 2008.

[9] Sumit Ganguly and Christian Sohler. d-Dimensional Knapsack in the Streaming Model. In *European Symposium on Algorithms (ESA)*, pages 468–479, 2009.

Prof. Surender Baswana

Efficient algorithms to solve problems approximately and/or in dynamic scenario

Duration of the project : three years (July 2007 to June 2010)
Principal Investigator : Surender Baswana, Assistant Professor, CSE, IIT Kanpur.

The project involved doing research in two areas of algorithms. The first research area is that of dynamic graphs algorithms. The second research area deals with designing algorithms which achieve better running time at the expense of solving the problem approximately instead of exactly. We addressed two fundamental problems in these areas. The first problem deals with graph spanners which are well known concepts in graphs. The second problem is the shortest paths problem. We summarize the work done in the last three years along with the publications (total four). We have used n and m to denote respectively the number of vertices and edges of a given graph.

Dynamic Algorithms for Graph Spanners

Spanner of an undirected graph $G = (V, E)$ is a sub graph which is sparse and yet preserves all-pairs distances approximately. More precisely, a spanner with *stretch* $t \in \mathbb{P}$ is a subgraph $(V, E_S), E_S \subseteq E$ such that the distance between any two vertices in the subgraph is at most t times their distance in G. Spanners have many applications in computing approximate distances in a graph, distributed computing, compact routing, and computational biology.

We designed two fully dynamic algorithms [3] for maintaining a sparse t-spanner of an unweighted graph. Our algorithms significantly improve the existing fully dynamic algorithms for graph spanners. The expected size of the t-spanner maintained at each stage by our algorithms matches the worst case optimal size of a t-spanner up to poly-logarithmic factor, and the expected amortized time required to process an update (insertion/deletion of an edge) is *close* to optimal.

The journal version of the above paper has also been accepted for publication (after revision) in *ACM Transaction on Algorithms*.

Streaming Algorithms for spanners

A streaming model has the following two characteristics: firstly the input data can be accessed only sequentially in the form of a stream; secondly the working memory is considerably smaller than the size of the entire input stream. An algorithm in this model is allowed to make one or more passes over the input stream to solve a given computational problem. The number of passes, the size of working memory, and the processing time per data item are the parameters which one aims to optimize in a streaming algorithm. We address the problem of computing spanner of an undirected unweighted graph in streaming environment. We present a one pass streaming algorithm that spends just constant average time per edge and computes a

$(2k-1)$-spanner of expected size $O(kn^{1+1/k})$. Note that the size of the spanner (and the working memory) is away from the conjectured optimal bound just by a factor of k, and so is essentially optimal for any constant k. Moreover, one pass and $O(m)$ time to process the stream is the best one can hope for in the streaming model. This result has appeared in [1].

Distance Oracles in sub-quadratic time

Thorup and Zwick, in the seminal paper [Journal of ACM, 52(1), 2005, pp 1-24], showed that a weighted undirected graph on n vertices can be preprocessed in subcubic time to design a data structure which occupies only subquadratic space, and yet, for any pair of vertices, can answer distance query approximately in constant time. The data structure is termed as approximate distance oracle. Subsequently, there has been improvement in their preprocessing time, and presently the best known algorithms achieve expected $O(n^2)$ preprocessing time for these oracles. For a class of graphs, these algorithms indeed run in $\Theta(n^2)$ time. In this paper, we are able to break this quadratic barrier at the expense of introducing a (small) constant additive error for unweighted graphs. In achieving this goal, we have been able to preserve the optimal size-stretch trade offs of the oracles. This result appeared in [2].

Approximate shortest paths avoiding a failed vertex : optimal size data structures for unweighted graphs

A path $\mathcal{P}$ between any two vertices $u, v \in V$ is said to be t-approximate shortest path if its length is at most t times the length of the shortest path between u and v. We consider the problem of building a compact data structure for a given graph G which is capable of answering the following query for any $u, v, z \in V$ and $t > 1$. This result appeared in [4].

REPORT t-APPROXIMATE SHORTEST PATH BETWEEN u AND v WHEN VERTEX z FAILS.

We present data structures for the single source as well all-pairs versions of this problem. Our data structures guarantee optimal query time. Most impressive feature of our data structures is that their size *nearly* match the size of their best static counterparts.

The journal version of the above result has also been accepted for publication (after revision) in *Algorithmica*.

References

[1] Surender Baswana. Streaming algorithm for graph spanners - single pass and constant processing time per edge. *Information Processing Letters*, 106(3):110–114, 2008.

[2] Surender Baswana, Akshay Gaur, Sandeep Sen, and Jayant Upadhyay. Distance Oracles for Unweighted Graphs: Breaking the Quadratic Barrier with Constant Additive Error. In *35th International Colloquium on Automata, Languages, and Programming (ICALP)*, pages 609–621, 2008.

[3] Surender Baswana and Somojit Sarkar. Fully Dynamic Poly-logarithmic Algorithms for Graph Spanners. In *19th Symposium on Discrete Algorithms (SODA),*, pages 672–681. ACM and SIAM, 2008.

[4] Neleesh Khanna and Surender Baswana. Approximate Shortest Paths Avoiding a Failed Vertex: Optimal Size Data Structures for Unweighted Graphs. In *27th International Symposium on Theoretical Aspects of Computer Science (STACS)*, pages 513–524, 2010.

Research Activities of PhD Students

Aditya Nigam

I have joined the department as a PhD student in the even semester(January) 2010 and have completed my coursework. Currently I am preparing for comprehensive examination. My adviser is Prof. Phalguni Gupta. My current research work is on face recognition. Also, I am working on automatic segmentation of tractography of brain data.

Courses studied and projects done:

1. Computational Complexity : *Term paper* : Interactive and Zero knowledge proofs for QNR, GNI problems.
2. Digital Watermarking Techniques : *Project* : Developed an algorithm for hiding data in a compressed video.
3. Software Architecture : *Project* : Extracted the software architecture of ray tracing software (Tachayon) using various architecture mining tool
4. Special Topics in Computer Science : *Project* : Edge weighted dissimilarity measure for face recognition.

Publications

[1] Aditya Nigam and Phalguni Gupta. A New Measure for Face Recognition System. In *5th International Conference on Image and Graphics (ICIG2009)*. IEEE CS, Xi'an, China, September, 2009.

[2] Aditya Nigam and Phalguni Gupta. Comparing Human Faces using Edge Weighted Dissimilarity Measure. In *11th International Conference on Control, Automation, Robotics and Vision (ICARCV 2010)*. IEEE CS, Singapore, December, 2010.

Ajitha Shenoy K. B.

Randomized search strategies work well in practice for many problems, although their worst case behaviour, in general, is exponential in time. Therefore, it is of interest to show that an algorithm based on such a strategy for a specific problem satisfies certain performance guarantees, although, in most cases, this is not an easy task.

I have learnt different kind of randomized search heuristics: Randomized local Search (RLS), Metropolis Algorithm (MA), Simulated Annealing (SA), Evolutionary Algorithm (EA), Go With Winner Algorithm (GWW) and Ant Colony Optimization (ACO).

I have completed my SOTA in August 2010. We have designed Metropolis Algorithm to solve Shortest Lattice Vector Problem (SVP). Based on some preliminary experimental data, we have compared MA performance with that of the LLL Algorithm and noticed that our algorithm gives better and quicker result than that of the LLL algorithm.

Now we have to analyze our algorithm and also have to see whether some other randomized search heuristics (like EA) will give better and quicker results for shortest Lattice Vector Problem (SVP).

Amrita Chaturvedi

I was selected under EURECA (European Research and Educational Collaboration with Asia) project 2009 to do research at Vrije University, Amsterdam, The Netherlands for 6 months. There, I did research under Dr. Hans Van Vliet, Professor, Department of Information Management and Software Engineering (IMSE), Vrije University, Amsterdam, The Netherlands. I attended Seventh joint meeting of the European Software Engineering Conference (ESEC'09) and the ACM SIGSOFT symposium on the Foundations of Software Engineering (FSE'09) held at Vrije University, Amsterdam, The Netherlands on August 24-28 2009. I also attended the PhD defense of Rik Farenhorst and Remco de Boer who were completing their PhD under the supervision of Dr. Hans Van Vliet and Dr. Patricia Lago in 2009 from Vrije University, Amsterdam, The Netherlands where Dr. Paul Clements, Professor, Software Engineering Institute, Carnegie Mellon University, USA and Dr. Philip Kruchten, Professor, University of British Columbia, Canada were called as examiners. I attended the semantic web meetings and WAI (weekly artificial intelligence) meetings held under the supervision of Dr. Frank Van Harmelen, Professor, CS Department, Vrije University, Amsterdam, The Netherlands and also presented my work in one of those meetings. I also presented my work in Software Architecture meeting held under the supervision of Dr. Hans Van Vliet and Dr. Patricia Lago, Professors, Department of Information Management and Software Engineering (IMSE), Vrije University, Amsterdam, The Netherlands.

After coming back to India, I successfully took my 'State Of The Art' exam and then extended my work done at Vrije University. I wrote a paper which analyses the role and effect of semantic components in software engineering and software systems. It begins by stating briefly what we mean by the term semantic component and enlists the various forms in which semantic components can be represented. It explores the relationship between semantic components and quality attribute response of an architectural design. It targets the problem of impact of semantic technologies on software architecture. For this analysis, a wine production case study and its domain model is used, whose business goals and domain assumptions are known a priori. The paper proposes two architectural designs for the wine production case study: one with and the other without semantic components. These two proposed designs are then analyzed and compared based on the quality attribute scenarios and their responses to different scenarios. Their comparative analysis reveals important changes in quality attributes that semantic components can bring in an architectural design.

Arpita Korwar

Having joined the PhD program in December, 2009, I pursued my course work for the next one semester. I took the following courses:

1. "Computational Complexity", where we studied about time and space complexity classes, randomized complexity classes. In the end of the course, I presented a paper "Applications of Universal Hashing in Complexity Theory" by V.Arvind and M.Mahajan. They present a result by Goldwasser and Sipser. They use universal hashing techniques to prove that the verifier does not gain much by keeping the result of the coin tosses private. That is, private and public coin protocols are equivalent.

2. A technical presentation course, where I read and presented four papers in the area of randomized hashing algorithms. The following topics were studied:

   (a) Universal hashing a common technique to reduce the number of random bits required in a hash function.

   (b) Perfect hashing a method for storing static data, that takes only constant number of look-up operations.

   (c) Cuckoo hashing a hashing algorithm that takes $O(1)$ time for look-up, expected amortized $O(1)$ time for insertion and $O(n)$ space. Though this algorithm was introduced by Pagh and Rodler fairly recently (2002), the randomized algorithms community has already performed lot of work around cuckoo hashing. One such paper is "On Risks of Using Cuckoo Hashing with Simple Universal Hash Classes" by Dietzfelbinger and Schellbach. Here, it was proved that if a multiplicative hash family is used and the set to be hashed is fairly dense in the universe then the probability that cuckoo hashing cannot hash this set if very high.

   (d) Bloom filters use k hash functions to record an element in this set. The probability of a false positive should be low. In 2006, Kirsch and Mitzenmacher proved in the paper "Less Hashing, Same Performance: Building a Better Bloom Filter" that the same performance can be achieved by using only two hash functions.

   (e) A hashing algorithm that takes two memory accesses for look-up, has 83.75% space utilization and inserts in $\leq O(\log n)$ time with high probability, constant in expectation. (reference: "Efficient Hashing with Lookups in two Memory Accesses" by Rina Panigrahy).

3. "Finite Automata on Infinite Input", where we studied about machines which don't have an 'END' state.

4. "Quantum Computing", where we study the basic gates used in the Quantum computer and how they can be combined to form circuits.

Ashish Agrawal

I have joined the PhD program of CSE department in January 2010. My areas of interest are Process Modeling, Software Architecture and Semantic Web. I am working under the guidance of Prof. T. V. Prabhakar. During this time I have completed my course work. Apart from course work, I got the opportunity to work on following idea:

Non-functional Requirements in Business Process Modeling (BPM and SOA perspective)

Non-functional requirements are either ignored or not represented explicitly in the process models. However, from SOA perspective, in order to provide agility, these requirements are necessary for generation of system model. In this work, we analyzed the BPMN specification and its extensions to identify constructs to represent NFRs. Our analysis showed that very few NFRs can directly be represented by BPMN and its extensions. Further, we analyzed the relationships between design decisions of process modeling and NFRs of the system. As a result we have given a list of relations between design choices and NFRs. This list of relations will help in identifying NFRs in the process model and thus a process model can be evaluated at the design time itself.

I have also finished my comprehensive examination. The oral part of comprehensive examination was presented on Security Protocols for Wireless Sensor Networks. Presently I am reading papers to find an interesting and relevant thesis topic in the area of software architecture.

Badrinath G. S.

During my last academic year from April 2009-March 2010, I have proposed some features extraction techniques to improve the recognition rate, speed and robustness of IRIS and Palmprint based biometric systems individually. The proposed features for IRIS based system include

1. Point pattern features using Harris corners and entropy,
2. SIFT features, and
3. Indexing using DCT features.

It is been observed that the propsed system performs with identification accuracy more than 99%. It is also observed that the Indexing system based on DCT features uses less than 10% of dataset for identification.

In case of Palmprint based biometric systems, some feature extraction techniques are proposed using

1. Zernike moments of Local region,
2. Stockewell Transform, and
3. SURF features.

It is observed that the proposed Palmprint systems perform with identification accuracy of 100%. The systems are tested with test images syntheticaly rotated various angales, synthetically scaled and occluded with various sizes. It is observed that the developed system performs with accuracy more than 95% for various modifications (Scale, Rotation and Occlusions). Hence the system is considered as robust to modifications.

Following are the publications obtained from the work done in the last academic year :

Publications

[1] Hunny Mehrotra, Badrinath G. S., Banshidhar Majhi, and Phalguni Gupta. A Efficient Iris Recognition using Local Feature Descriptor. In *16th IEEE International Conference on Image Processing (ICIP 2009)*, Cairo, Egypt, November, 2009.

[2] Hunny Mehrotra, Badrinath G. S., Banshidhar Majhi, and Phalguni Gupta. An Efficient Dual Stage Approach for IRIS Feature Extraction using Interest Point Pairing. In *IEEE Workshop on Computational Intelligence in Biometrics (CIB 2009)*, Nashville, TN, USA, April, 2009.

[3] Hunny Mehrotra, Badrinath G. S., Banshidhar Majhi, and Phalguni Gupta. Indexing Iris Biometric Database using Energy Histogram of DCT Subbands. In *International Conference on Contemporary Computing (IC3-2009)*, 2009.

[4] Badrinath G. S. and Phalguni Gupta. Palmprint based Verification System Robust to Rotation, Scale and Occlusion. In *12th International Conference on Computer and Information Technology (ICCIT-09)*. IEEE Computer Society Press, Dhaka, Bangladesh, December, 2009.

[5] Badrinath G. S. and Phalguni Gupta. Palmprint based Verification System Using SURF Features. In *International Conference on Contemporary Computing (IC3-2009)*, Noida, August, 2009.

[6] Badrinath G. S. and Phalguni Gupta. Robust Biometric System using Palmprint for personal Verification. In *IAPR/IEEE International Conference on Biometrics (ICB 2009)*, number 5558 in LNCS, pages 554–565, Sassari, Italy, June, 2009.

[7] Badrinath G. S. and Phalguni Gupta. A Novel Representation of Palm-print for Recognition. In *Asian Conference on Computer Vision (ACCV-2010)*, November, 2010.

[8] Badrinath G. S. and Phalguni Gupta. Stockwell Transform based Palm-print Recognition. *Applied Soft Computing, Elsevier*, 2011.

[9] Badrinath G. S., Naresh Kachi, and Phalguni Gupta. Palmprint based verification System Robust to Occlusion using Low-order Zernike Moments of Sub-images. In *British Machine Vision Conference (BMVC 2009)*, London, UK, September, 2009.

[10] Badrinath G. S., Naresh K. Kachi, and Phalguni Gupta. Palmprint based Verification System Robust to Occlusion using Low-order Zernike Moments of Sub-images. *Special Issue of Biometrics Systems and Applications in the Journal of Telecommunication Systems, Springer Verlag*, 2010.

Balwinder Sodhi

I joined the CSE/IITK department in July 2009 with a master's degree in Computer Science. My academic load in the 2009-I semester mainly consisted of course-work which I completed with CPI 9.5. 2009-II semester focussed on comprehensive exam and SOTA requirements of the PhD program. As part of the above mentioned program related activities I have worked on the following research problems :

A Lightweight Hybrid Approach for Developing High Performance Web Applications

This work was done as part of the CS697 special topics course. Here we analysed existing web application development framework for thier suitability for special deployment environments such as a cloud or a resource (CPS, memory and network bandwidth etc.) constrainted platform. We proposed a design technique for allowing better resource utilization under such deployment scenarios. This work was accepted at W2T'2010 (Web2Touch) workshop of NOTERE'10 (but was withdrawn due to travel date constraints). We have enhanced this work and collected more diverse experimentation results, and it is pending review (as of 20-Sep-2010) in Web Technologies track of SAC'2011.

Application Architecture Considerations for Cloud Platforms

This work was done as part of the SOTA literature study. The focus of this work is to examine the key differences in cloud-centric and traditional approaches of application architecture and design from various architectural quality-attributes aspects. Developing applications targeted for cloud deployment is not quite same as a traditional application deployment scenario. Appearance of many different vendor specific application development platforms and frameworks has made the task of designing applications for cloud difficult. In this work we analysed the background of relevant cloud computing aspects. We then introduce two primary application design approaches for the cloud: cloud-aware application design and cloud-agnostic application design. These approaches are then evaluated in context of few application scenarios that are common across most business domains. In the evaluation we determine how various non-functional quality attributes of the applications are impacted by the choice of design approach. This work is pending review (as of 20-Sep-2010) in COMSNETS'11.

Presently I am trying to gain a deeper understanding of virtualization and infrastructural level aspects of cloud computing and how it impacts the application software design and architecture and am in the process of finding problems to work on in the coming months.

Chandan Saha

In continuation to my research on the problem of Polynomial Identity Testing (PIT) the previous year (2008-09), this year (2009-10) I have studied the complexity of two very natural problems on identity testing. One is a generalization of a problem considered earlier by Neeraj Kayal and Nitin Saxena in 2006. Kayal and Saxena gave a deterministic polynomial time algorithm to test if a given depth-3 circuit with bounded top fanin computes an identically zero polynomial. We consider a generalization of this problem: Test if the output of a given depth-3 circuit with bounded top fanin equals a given sparse polynomial.

The second problem we study is the following: Check if a given sparse multivariate polynomial equals the product of a given set of other sparse polynomials. This problem was noted as an open question by Joachim von zur Gathen in 1983 in one of his papers. So far we are not aware of any progress made towards solving even any interesting special case of this problem.

Our contributions to these problems are as follows. We give a deterministic polynomial time algorithm for the first problem thereby extending the Kayal-Saxena result. As for the second problem, we give a deterministic polynomial time solution for the natural special case where every polynomial in the input set of factors is a sum of univariates. In particular, there is an efficient way of checking if a given sparse polynomial equals a given product of linear functions. The solutions to both these problems use a technique called dual representation of polynomials that was introduced earlier by Nitin Saxena in 2008 in one of his work.

This is a joint work with Nitin Saxena and Ramprasad Saptharishi.

Kamlesh Tiwari

Research Activities (Dec-2009 to July 2010)

I joined the Ph.D programme of the department in December-2009. This was the first year of my Ph.D. programme. I took four courses in the semester

1. *Digital watermarking and stenography*,
2. *Parallel computing*,
3. *Software Architecture*,
4. *Computational Complexity*.

As part of the research project for the digital watermarking and stenography course we have proposed a method for information hiding in MPEG compressed video. The method had used motion vectors of P and B frames to embed the information. In parallel computing course we have proposed the design of a processor for document classification application of multi-processor environment. We have also developed specialized instruction set and the SIMD based system architecture. The new architecture is tested with respect to a proposed algorithm and showed satisfiable results. In Software architecture course we have learned a lot of things specially the deep effect of non-functional attributes on the system. We have exercised architecture discovery tools like Lattix, Graphviz, Doxygen etc. to reveal and study the architecture of an open source ray tracing engine Tachyon.

In the past few year, my work was in the area of systems security and applied cryptography. I had worked for the development of e-cash system for secure financial transactions. My other areas of interest are biometrics, image processing, data structures and algorithms, computer architecture, operating systems and computer networks.

Kiran Kumar Reddy

I was working on codification the design patterns knowledge [1, 2]. The transformation of architectural requirements to multiple design views can be understood as three-stage process: Design alternative analysis, Best alternative selection, and Architecture documentation. During the design alternative analysis stage, for each requirement the set of design alternatives which achieve them are analyzed. After analyzing the alternatives, the best design alternative is selected as design decision based on various tradeoffs and constraints of the system.

Software design patterns document the most recommended solutions to recurring design problems. Selection of the best design pattern in a given context involves analysis of available alternatives, which is a knowledge-intensive task. Pattern knowledge overload hardens this analysis. Sometimes designers choose recently used design decisions when a thorough alternative analysis is not possible. Under these circumstances, designers can benefit by a competent knowledge base to generate available alternatives. The tools which integrate such knowledge base as one of their components are termed Design Assistants.

We focus on codifying an important part of patterns knowledge which includes essential design concepts such as: *Pattern to Tactic relationship, Pattern to Pattern relationship, Pattern to Quality-attribute relationship* and *Pattern to Application-type relationship*. We perform analysis of these relationships for patterns in the two popular pattern catalogues viz GoF and POSA1.

One basic difference of our analysis and others' is that we analyze patterns from a bottom-up perspective; our analysis is based on an underlying tactics based formal model of patterns.

We also proposed a graph based model called *Design Decision Topology Model* (DDTM) to deal with the relationship analysis problem. The objective of this model is to reduce pattern semantics to syntax - a graph which delivers the pattern functionality (quality) through elementary functionality (quality) - nodes of the graph are elementary functional functionality and edges are dependencies. Conceptually, the DDTM technique is analogous to the Decision view in the architecture domain. This model treats each pattern as a micro-architecture and defines the pattern as a topology of a set of design decisions. Using this model, different relationships are analyzed using graph properties. For example,

Patterns A and B are duplicates if Graph(A) $\subset$ graph(B).

Pattern A comprises-of patterns B and C if Graph(B) $\subset$ Graph(A) AND Graph(C) $\subset$ Graph(A).

References

[1] Kiran Kumar and Prabhakar T. V. Design Decision Topology Model for Pattern Relationship Analysis. In *Asian Conference on Pattern Languages of Programs*, Tokyo, Japan, March 15-17, 2010.

[2] Kiran Kumar and Prabhakar T. V. Pattern-oriented Knowledge Model for Architecture Design. In *Pattern-oriented Knowledge Model for Architecture Design*, Reno/Tahoe Nevada, October 15-18, 2010.

Pawan Kumar Aurora

Some of the claims stated in my previous report turned out to be false. We could not make any significant progress along that direction. Since early this year, we have been looking at the graph coloring problem. In particular, we have been looking at the problem of coloring 3-colorable graphs. It is an important problem in the sense that there exists a huge gap between the known lower and upper bounds. Making any headway in either direction would be useful. So far we have been looking at the significant contributions made by various researchers towards the upper bound. At this point we more or less understand the state-of-the-art in this area.

In parallel we have been learning various techniques in approximation algorithms using the text by Vazirani as the main reference.

I attended the DIMAP Summer School on Approximation and Randomized Algorithms held at the University of Warwick during July 12-16, 2010.

Purushottam Kar

In the year 2009-10, my academic load was a mix of course-work and research. I did courses on Computational Number Theory and Algebra, Algorithmic Game Theory, Data Streaming Algorithms (audit) and Differential Geometry (audit). This year I continued with my investigation of efficient algorithms for massive, high dimensional datasets that we had begun with our work on low distortion embeddings of statistical distance measures in 2008-09 [1]. I collaborated with several people on projects that aimed at tackling this problem in different ways.

In collaboration with Prof. Sumit Ganguly, I worked on developing sketching schemes for data streams with fast update times. Problems in data streaming require one to work under severe space constraints since the dimensionality of the domain is typically too high to allow one to work explicitly in the input spaces. The problem we worked on was that of estimating the L_1 norm of a vector which is presented to us as a series of updates (on a stream). This problem has numerous applications in database query optimization and network monitoring. Existing algorithms, while being space optimal, took $O(1/\epsilon^2)$ (along with certain logarithmic factors) time to process each stream update. We designed an algorithm which although is a logarithmic factor away from achieving space optimality, was able to provide an exponential speedup in the update time by taking only $O(\log(1/\epsilon))$ (ignoring some logarithmic factors) time to process each update. Our work was submitted and subsequently accepted for publication at the 12th Italian Conference on Theoretical Computer Science [3].

The other major project I undertook was in collaboration with Aman Dhesi (then a UG student with the EE department and now a graduate student at Princeton University). The problem was to tackle high dimensional point sets by observing that these point sets typically lie on low dimensional manifolds. The aim here was to develop algorithms whose performance depends upon the dimensionality of the underlying manifold (which is typically low) and not on the ambient dimensionality (which is usually high). We started working toward algorithms for efficient nearest neighbor search for manifold datasets. This problem is ubiquitous in the domains of information retrieval and database technology.

The work is still in progress - however, in pursuit of our goals, we started an investigation of the Random Projection Tree data structures proposed by Dasgupta and Freund in STOC 2008. We could show guarantees for this data structure which were near equivalents of those given by the k-d Tree data structure which is widely used in data retrieval applications. Our results were submitted and subsequently accepted for publication at the 24th Annual Conference on Neural Information Processing Systems [2]. We are now working on an extension to the Random Projection Tree data structure that allows it to be used in fast nearest neighbor routines.

I am also currently working on several other projects in the area of classification and kernel-based learning. We also completed two successful seasons of our research group "Special Interest Group on Machine Learning" (`http://www.cse.iitk.ac.in/users/sigml/`) with a total of 13 seminars being given in the two semesters (Fall 2009 and Spring 2010).

References

[1] Arnab Bhattacharya, Purushottam Kar, and Manjish Pal. On Low Distortion Embeddings of Statistical Distance Measures into Low Dimensional Spaces. In *20th International Conference on Database and*

Expert Systems Applications (DEXA), 2009.

[2] Aman Dhesi and Purushottam Kar. Random Projection Trees Revisited. In *24th Annual Conference on Neural Information Processing Systems (NIPS)*, 2010.

[3] Sumit Ganguly and Purushottam Kar. Estimating F_1 of data stream in nearly optimal space and time. In *International Conference of Italian Computer Science*, Cameron, Italy, August, 2010.

Sagarmoy Dutta

Introduction

I am Sagarmoy Dutta, a Ph.D student in Computer Science and Engineering department of Indian Institute of Technology Kanpur. My research interests span computational algebra, complexity theory and quantum computing. Currently, under the supervision of my thesis guide Dr. Piyush P. Kurur, I am working on generalising the idea of classic cyclic code to quantum error correcting code (QECC). In the following sections I will briey discuss the background and then explain what we have done so far and how we intend to proceed.

Cyclic code and QECC

In classical coding theory data of length k are elements of $\mathbb{F}^k$ where $\mathbb{F}$ is some field which acts as alphabet here. Data is coded by applying an injective map from $\mathbb{F}^k$ to $\mathbb{F}^n$ $(n > k)$. It is designed in such a way that even if some of the symbols of a codeword is changed it is still possible to decode it back to the original data. Classic codes has an integer parameter d, called 'distance', such that upto d errors in a codeword can be detected and $\lfloor \frac{d-1}{2} \rfloor$ errors can be corrected.

Cyclic codes are linear subspaces of $\mathbb{F}^n$ such that if $a_1 \dots a_n$ is a codeword, its cyclic shift $a_2 \dots a_n a_1$ is also a codeword. Cyclic codes occupy an immensely important place in classic coding theory. Staring from the very basic repetition code to well known BCH or Reed-Solomon codes are special cases of cyclic code. The power of cyclic codes comes from a nice algebraic characterisation that these are ideals of the ring $\mathbb{F}[x]/(x^n - 1)$. Choosing the field and the generating polynomial judiciously one can ensure good distance and apply efficient algorithms for decoding.

Quantum data of length k are elements of $\mathbb{C}(\mathbb{F}^k)$ which is a $\mathcal{L}_2$ normed vector space of dimension $|\mathbb{F}|^k$ over complex numbers with orthonormal basis labelled by the elements of $\mathbb{F}^k$. Coding is done via an injective map from $\mathbb{C}(\mathbb{F}^k)$ to $\mathbb{C}(\mathbb{F}^n)$ and distance is also defined analogously. Except for a few examples, all the QECC studied so far belongs to a class called stabiliser code. Some work has been done by other researchers to use techniques similar to BCH codes to construct quantum codes. However these codes fall into a subclass of stabiliser code called CSS code. This is a severe restriction because there are optimal codes like Lafflamme code, which meets the sphere packing bound for QECC, that are not CSS.

Future direction

Our goal is to generalise the idea of classical cyclic codes to quantum error correcting codes which can encompass non-CSS and perhaps even non-stabiliser codes. Let U be an unitary operator which maps $|a_1 \dots a_n\rangle$ to $|a2 \dots a_n a_1\rangle$. We observed that many well known codes including non CSS optimal code like Laflamme code satisfy the property that if $|\phi\rangle$ is a codeword so is $U|\phi\rangle$. We take this property as the definition of quantum cyclic code (QCC). Interestingly, according to this definition certain non-stabiliser codes are also cyclic. Our targets are the following.

1. Give a succinct algebraic characterisation of QCC

2. Use this characterisation to construct new families of QECC

3. Device efficient algorithm for decoding

We have already got some success in the first two. We showed that any QCC can be characterised by a set of polynomial pairs which has very similar structure to an ideal. CSS codes can be seen as special cases where this set satisfies one extra restriction. In another special case where the base field is $\mathbb{F}_2$ this set is indeed an ideal of $\mathbb{F}_4[x]/(x^n - 1)$. This provides a method to construct a family of binary non-CSS stabiliser code where n is of the form $4^t + 1$. In case of $n = 1$ it is in fact the Lafflamme code. For $n = 2$ we get two codes of dimension 1 and 9 respectively which can correct error upto what is possible by quantum sphere packing bound. In classical coding theory there is a lower bound on distance of the code called BCH bound. If weight of an error is less than half of BCH bound then it can be efficiently corrected using Berlekamp algorithm. It is not clear how to find a similar bound for general QCC. However, we are able to do that for $4^t + 1$ qubit codes we have constructed and found efficient decoding algorithm for errors upto that bound. Also we want to generalise our construction to find codes where n is not restricted to be of the form $4^t + 1$. Other possible direction of generalisation is towards QCC over over fields other than $\mathbb{F}_2$ or even non-stabiliser codes.

Satyam Sharma

I am a second year Ph.D. student working in the area of cryptology and security. My other areas of interest are computational complexity, concurrent data structures and algorithms, operating systems and computer networks. My thesis advisor is Prof. Rajat Moona.

In the past year, I have finished my comprehensive examination and given my state-of-the-art seminar after reading relevant literature and some initial research to finalize the topic of my thesis work. The broad area of my research is in the field of secure multiparty computation. I will now give a brief overview of this exciting field.

General distributed computation deals with the problem of efficiently computing an n-ary randomized functionality that maps n inputs to n outputs in a setting consisting of n different processors that communicate by passing messages over an interconnection network. The motivation behind the idea of distributed computation remains the improvement of efficiency, reliability and the difficulty in amassing all the computing and storage requirements at a centralized processor. Within this broad context, secure multi-party computation protocols specify a kind of distributed computation that preserves certain security properties such as independence of inputs, privacy, output delivery, correctness and fairness. Such a definition makes the notion of secure multi-party computation general and powerful enough to encompass a wide variety of applications from e-voting to e-cash and from secure auctions to privacy-preserving data mining. Also, unlike the previous setting where we wish to avoid centralized processing due to logistical and efficiency reasons, secure computation deals with the possibility of malicious counterparties and necessitates the need for such decentralization.

Historical research in this area has dealt with feasibility and impossibility results that make assumptions and characterize the kind of adversaries that can safely be dealt with in the construction of generic protocols that preserve the requisite security properties. However, apart from being of independent theoretical interest in itself, the applications and possibilities opened up by the idea of secure computation have fuelled additional research into making the theoretical constructions more efficient and practically useful. Some questions and problems that I have identified and intend to explore during the course of my thesis work are:

1. Can we improve (minimize the local computation and interaction requirements) the aforementioned general protocols ?

2. Can we obtain new feasibility or impossibility results under new models by generalizing or specializing the assumptions behind them ?

3. Can we construct specialized protocols (simpler and more efficient than the general ones) that deal with specific practically useful functionalities ?

Saurabh Joshi

Academic activities during April - 2009 to March - 2010

I visited my co-supervisor Prof R K Shyamasundar at TIFR , Mumbai during summer (May 2009 - July 2009). During this period we explored a few works on mobile barriers and dynamic barrier synchronization.

We have worked on improving the existing works on May-happen-in-parallel (MHP) analysis. We have come up with phase interval analysis (PIA) which can be carried out in the same time complexity as earlier work on MHP analysis of X10 programs but provide opportunities for more precise MHP information when the loops are regular. In addition, to MHP information, PIA provides information on the ordering of the two statement blocks as well, which can be used in other concurrent program optimizations like copy propagation etc. PIA essentially, computes the phase interval with respect to X10 clocks (dynamic barriers) in which a given statement block can execute. These interval can also provide condition functions which are linear functions over loop induction variables when loops are regular. These condition functions, provide conditions under which two statement blocks may execute in parallel. The work is a definite improvement over existing work on MHP with X10 clocks and we are trying to submit it to some reputed international conference.

With the partial support from the department (Around Rs. 6100/-), I attended international conference on Principles and Practices of Parallel Programming (PPoPP ' 2010) held in January at IISc, Bangalore. The rest of the financial support and registration waiver was received from Conference Organizers.

I also worked on protocol on 'Distributed generalized dynamic barrier synchronization' with a Researcher from IBM India Research Lab and my co-supervisor. I rectified a few errors in the work already done and suggested some improvements. This work is under submission to ICDCN' 2011.

Seetha Ramaiah

In an effort to study the possibilities of autopoietic behaviour in software systems, several things including the following were considered:

1. Extending IBM's model of Autonomic Computing with self-replication
2. Autopoiesis as a quality attribute

Extending IBM's model of Autonomic Computing with self-replication

Autonomic computing refers to the self-managing capabilities of distributed computing resources so that they can adapt to unpredictable changes, while hiding the intrinsic complexity from the end users. It is inspired by the autonomic nervous system which controls many of the organs of the human body. The autonomic nervous system functions in such a way that we won't have to worry about what goes on behind the scenes. Moreover, it always keeps working. Inspired by these characteristics, autonomic computing strives to achieve high availability and involuntary and reflexive ways of functioning so that the end users will not have to worry about how the system functions. The self-manageability of autonomic computing systems is a collection of four attributes: self-configuration, self-healing, self-optimization and self-protection. These four attributes together are referred to as self-CHOP and are shown in the figure below. Self-configuration refers to the ability of the system to adapt to new environments automatically by deploying new components and removing existing components as needed. Self-configuration presumes a set of policies provided by the IT professionals. Self-healing refers to the ability to detect the malfunctions of the system and initiate the respective corrective actions. Self-optimization is the ability to automatically monitor and tune the resources and reallocate them in response to dynamically changing workloads so as to maximize resource utilization. Finally, self-protection is the ability to anticipate, detect, identify external attacks on the systems and protect itself from such attacks. These external threats could arise from sources such as unauthorised access and virus infection. In addition to these four attributes, a fifth attribute namely self-replication was proposed. Self-replication is the ability of the system to replicate parts or whole of itself as per necessity. The system may have to replicate itself when it detects a potential virus threat on it so that when one copy is known to be attacked by a virus, the other copy could kill the first one and resume the operations.

Autopoiesis as a quality attribute

From a software-architectural perspective, autopoiesis can be seen as a quality attribute just like performance, security, availability, scalability etc. With this perspective, we tried to see whether autopoiesis is an intrinsic property (that is, a property innate to a particular component of the system) or an emergent property (that is, a property that emerges as a result of combining certain components in a certain manner). For this, examples from human immunology were considered to see how the human immune system as a whole achieves autopoietic behaviour.

Sujith Thomas

Fall Semester 2009-10

During the fall semester 2009-10 did the following courses:

Course No.	Course Name
CS601	FUNDAMENTALS OF CSE; MATHEMATICS FOR CS, ALGORITHMS
CS684	INTRODUCTION TO ALGORITHMS AND LOGICS IN GAME THEORY
CS697	SPECIAL TOPICS IN COMPUTER SCIENCE

Project

I did the following project during first semester 2009-10.

Learning sequencing rules for the Game of Life
Game of Life is a Cellular Automata devised by John Conway. Game of Life has simple rules which define the state transition of a cell between discrete time-steps. Complex patterns emerge from these simple rules which we call Emergent patterns. In this project we try to detect Oscil- lators using a Recurrent Neural Network. We also use a 4-layer Neural Network to detect gliders.

Spring Semester 2009-10

During the spring semester 2009-10 did the following course :

Course No.	Course Name
EE658	FUZZY SET, LOGIC SYSTEMS AND APPLICATIONS

Thesis

I did the following project/thesis during second semester 2009-10.

Computing with words : fuzzy queries and linguistic data summaries from IPL T20 statistics
In this project we construct results for fuzzy queries run on a database containing IPL T20 data. The results for fuzzy queries are constructed using the fuzzy logic principles. Then we use Zadeh's proto-forms and calculus of linguistically quantified propositions to generate summaries from the database with each statement having a truth-value associated with it.

Knowledge representation

Formal Concept Analysis (FCA) is a lattice based knowledge representation technique. It gives a hierarchical structure to the concepts based on subset relationship between concepts. I read books and other literature on FCA.

Conceptual Graphs (CG) is a graphical knowledge representation technique. It employs first order logic to provide the semantics for the various graphical constructs. I read several literature on CGs.

Surya Prakash

An Efficient Ear Recognition Technique Invariant to Illumination and Pose[1]

Introduction

A biometric based security system is expected to fulfill user's demand such as low error rates, high security levels, testing for liveliness of the subject, possibility of fake detection etc. Even though the recognition performance of biometric systems has been significantly improved in recent past, there is a need of further improvement of existing techniques. Most of the existing ear recognition techniques have failed to perform satisfactorily in presence of varying illumination, occlusion and poor image registration. This research work proposes an efficient ear based recognition technique which can handle some of these issues. In this proposed technique, an ear image is enhanced using three image enhancement techniques applied in parallel. SURF feature extractor is used on each enhanced image to extract local features. A multi-matcher system is trained to combine the information extracted from each enhanced image. The technique is found to be robust to illumination changes and works well even when ear images are not properly registered.

The use of multiple image enhancement techniques has made it possible to counteract the effect of illumination and poor contrast while SURF [2] based local feature helps in matching the images which are not properly registered and suffer from pose variations. For a given ear image, three enhanced images are obtained which are used by the SURF feature extractor to generate three sets of SURF features for an ear image. Three nearest neighbor classifiers are respectively trained on these three sets of features and finally the output of all the classifiers are fused to get the final result.

Proposed Technique

The proposed ear recognition technique follows three major steps: Image Enhancement, Feature Extraction and Classification and Fusion. Overview of the proposed system is shown in Figure 1.

Image Enhancement involves three enhancement techniques: Adaptive Histogram Equalization (ADHist) [7], Non-Local Means (NLM) algorithm [6] and Steerable Filter (SF) [3]. It is intended to enhance the contrast of the ear image and to normalize the effect of illumination and shadow. The purpose of enhancement is to get the accurate SURF feature descriptor vectors for a feature points and to help in establishing the correct point correspondence between the feature points in two images. The enhancement algorithms have been used in parallel on each input ear image to get enhanced images which are later used for feature extraction.

Feature Extraction uses SURF for feature extraction which provides representation of an image in terms of a set of salient feature points, each point associated with a descriptor vector of 128 feature elements.

[1] **Publication based on this work:**
Surya Prakash and Phalguni Gupta, *An Efficient Ear Recognition Technique Invariant to Illumination and Pose*, Telecommunication Systems Journal, special issue on Signal Processing Applications in Human Computer Interaction, Springer, 2011, *(to appear)*.

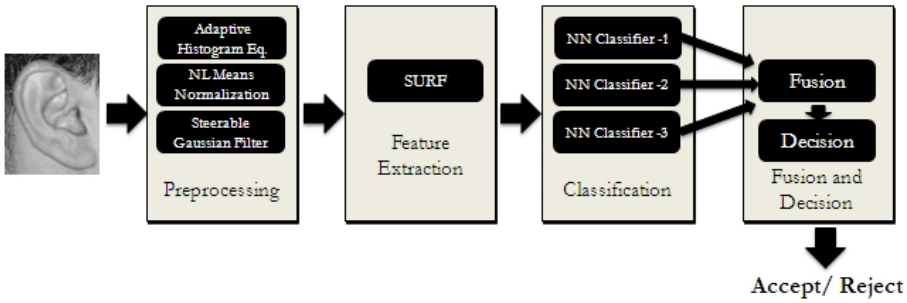

Figure 1: Block diagram of the proposed ear recognition technique

Table 1: Comparison of performance of the proposed technique with the latest reported results for UND-E database

Technique	Accuracy (*FAR*,*FRR*)	*EER*	*EUC*
Proposed in [4]	-	4.20	3.00[a]
Proposed in [5]	-	-	1.50
Proposed Technique	**96.75 (2.58,3.92)**	**3.80**	**1.13**

[a]reported in [5] for the technique proposed in [4]

A technique for feature level fusion is proposed to obtain a fused representative template for a subject by combining the features of multiple training samples of the subject.

At *Classification and Fusion* step, features obtained from each enhanced image are used to train nearest neighbor classifiers. Matching scores produced by classifiers are normalized using min-max normalization technique and are then fused using weighted sum rule. Final classification decision is taken by using the fused score.

Experimental Findings

Experiments are conducted on two databases, namely IIT Kanpur database and University of Notre Dame database (Collections E) [1] and encouraging results are achieved. Comparative performance of the proposed technique with the best known results for UND-E database is summarized in Table 1. Results obtained by the proposed technique are averaged over 30 experiments; hence it shows more stable performance compared to the results reported in [4] and [5] where they are averaged only for 10 and 20 experiments respectively. *ROC* curves for UND-E database are shown in Figure 2 where the *ROC* curve employing all three image enhancement techniques is found to be superior to others.

Conclusions

The technique proposed in this work uses three different image enhancement techniques in parallel to overcome the effect of illumination and contrast and extracts local features from the enhanced images using SURF. Fusion at score level is carried out to combine the scores generated from three classifiers and decision is taken based on the fused score. The technique has been evaluated on two ear databases, namely IIT Kanpur ear database and University of Notre Dame ear database (Collection E). Experimental results show that the proposed technique provides a considerable improvement in terms of performance over existing techniques.

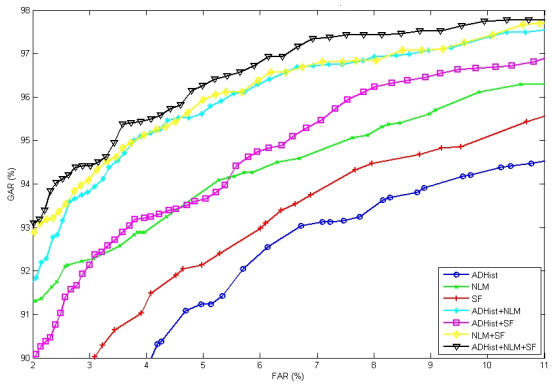

Figure 2: *ROC* curves for UND-E database for combinations of various enhancement techniques

References

[1] University of Notre Dame Ear Database, Collection E. http://www.nd.edu/ cvrl/CVRL/DataSets.html

[2] Bay, H., Ess, A., Tuytelaars, T., Van Gool, L.: Speeded-up robust features (SURF). CVIU **110**(3), 346–359 (2008)

[3] Freeman, W.T., Adelson, E.H.: The design and use of steerable filters. IEEE PAMI **13**(9), 891–906 (1991)

[4] Nanni, L., Lumini, A.: A multi-matcher for ear authentication. PRL **28**(16), 2219–2226 (2007)

[5] Nanni, L., Lumini, A.: Fusion of color spaces for ear authentication. PR **42**(9), 1906–1913 (2009)

[6] Struc, V., Pavesić, N.: Illumination invariant face recognition by non-local smoothing. In: Proc. of Joint COST 2101 and 2102 Int'l Conference on Biometric ID Management and Multimodal Communication (BioID MultiComm' 09), LNCS 5707, pp. 1–8 (2009)

[7] Zuiderveld, K.: Graphics Gems IV, chap. Contrast limited adaptive histogram equalization, pp. 474–485. Academic Press Professional, Inc. (1994)

Umarani Jayaraman

Efficient search and retrieval techniques in multi-modal biometric database

Biometric system provides an automated method to verify or to identify an individual based on unique behavioral or physiological characteristics. The task of the authentication module in a biometric system is to recognize a subject either by identification of one person among many, or verification that a person's biometric matches a claimed identity. In case of identification, for given query image it has to compare the whole database to declare its identity.

In case of security checks at airports and border crossing, the biometric database is very large. Performing an exhaustive search in a database involving billion of comparisons will be computationally expensive. If an effective search method is designed for a biometric database, it will reduce the number of comparisons on biometric database. Thus, the computational complexity in searching a given query image in a very large database will be reduced significantly. In other words, the indexing method selects a small subset of images in the database from which the feature matching algorithm determines the correct match.

Published Papers

[1] Umarani J, Surya Prakash, and Phalguni Gupta. An Efficient Technique for Indexing Multimodal Biometric Databases. *International Journal of Biometrics*, 2009.