Annex 2

List of Completion Research Reports In Multimedia Technologies

(2010 - 2012)	* FADE: Secure Cloud Storage with File Assured Deletion * Security and Detection Protocols for P2P-Live Streaming Systems
(2009 - 2011)	* An Opportunistic Approach to Capacity Enhancement in Wireless Multimedia Networks * Computer-Aided Second Language Learning through Speech-based Human-Computer Interaction
(2008 - 2010)	* Pattern Computation for Compression and Performance Garment
(2007 - 2009)	* Real-time Transmission of High Definition (HD) 3D Video and HD Audio in Gigabit-LAN * High Dynamic Range Image Compression and Display * Multimedia Content Distribution over Hybrid Satellite-Terrestrial Communication Networks
(2006 - 2008)	* Automatic Video Segmentation and Tracking for Real Time Multimedia Services * Information Retrieval from Mixed-Language Spoken Documents * Wireless Networks and Its Potential for Multimedia Applications
(2005 - 2007)	* Mobile Wireless Multimedia Communication * An Automatic Multi-layer Video Content Classification Framework * Automatic Multimedia Fission, Categorization and Fusion for Personalized Visualization in Multimedia Information Retrieval

FADE: Secure Cloud Storage with File Assured Deletion

Principal Investigator: Patrick Pak-Ching Lee (1)
Co-Investigator (if any): Nil.

Research Team Members: Arthur Rahumed, Chung-Pan Tang (1)

(1) Dept. of Computer Science and Engineering, The Chinese University of Hong Kong

Project Start Date: 1st July 2010
Completion Date: 30th June 2012

ABSTRACT

We can now outsource data backup to third-party cloud storage services so as to reduce data management costs. However, security concerns arise in terms of ensuring the privacy and integrity of outsourced data. We design and implement FADE, a cloud storage system that enforces access control of active data and protects deleted data with policy-based file assured deletion. FADE is built upon a set of cryptographic key operations that are maintained by a quorum of key management entities, and encrypts outsourced data files to guarantee their privacy and integrity. It uses file access policies to provide a fine-grained control of how active files are accessible, and assuredly deletes files to make them unrecoverable to anyone upon revocations of file access policies. In particular, FADE acts as an overlay system that works seamlessly atop today's cloud storage services. To demonstrate this objective, we implement a working prototype of FADE atop Amazon S3, one of today's cloud storage services, and empirically show that FADE provides security protection for outsourced data with a minimal trade-off of performance overhead. Our work provides insights on how to incorporate value-added security features into current data outsourcing applications, such as backup applications for large digital multimedia files.

1. OBJECTIVES AND SIGNIFICANCE

This project achieves the following objectives:

- Design and implement a practical, deployable cloud storage tool called FADE that supports file encryption, assured deletion, and access control for today's cloud storage services.
- Design a lightweight policy-based deletion scheme that dynamically associates files with different file access policies and assuredly deletes files that are tied to obsolete file access policies.
- Design a practical access control mechanism based on attribute-based encryption and a practical key management framework based on threshold secret sharing
- Design a deduplication scheme that minimizes the use of cloud storage space.
- Design a robust version control system for virtual machine storage on a cloud computing platform.
- Release our system as open-source software that enables individuals and enterprises to use and extend.

2. RESEARCH METHODOLOGY

We now present the design details of FADE. FADE aims to achieve the following properties:

- It provides security guarantees for outsourced data on the cloud, including data confidentiality, file access control, and file assured deletion. In particular, one novel security feature of FADE is policy-based file assured deletion.
- It provides fault-tolerant key management.
- It works seamlessly with today's cloud storage services.

Our main design intuition is as follows. We decouple the management of encrypted data and encryption keys, such that the huge amount of encrypted data remains on the third-party (untrusted) cloud provider, while a smaller size of encryption keys are managed by a key manager service, which can be self-maintained and fully controlled by the data owner. We can then add security and resilience features to key management and hence the outsourced data. We leverage several novel techniques in the field of cryptography to achieve a robust design.

Overview of security properties

We first summarize the security properties that we want to achieve in FADE.

Data confidentiality. All outsourced data files are encrypted, so that they are inaccessible to unauthorized parties including cloud storage providers.

Access control. We leverage a cryptographic technique called *attribute-based encryption (ABE)*, whose main idea is to provide a fine-grained access control based on different file access policies. For example, we may specify that a file is accessible only if "the user is Alice and the access time is before 2001-01-01". This technique ensures the authentication guarantees of file upload and download operations.

Assured deletion. We give a formal definition of assured deletion as follows: if the file access policy associated with a data file is revoked, then the data file will become inaccessible by anybody (including the owner of data file). Here, we leverage the cryptographic property of encryption: *we store encrypted copies of data files, and delete the encryption keys that are independently maintained*. As long as the keys are removed, all persisting copies remain encrypted and it is computationally infeasible to recover the original data.

System architecture

Figure 1 illustrates the architecture of FADE. It is composed of two elements: (i) FADE interface and (ii) FADE key manager.

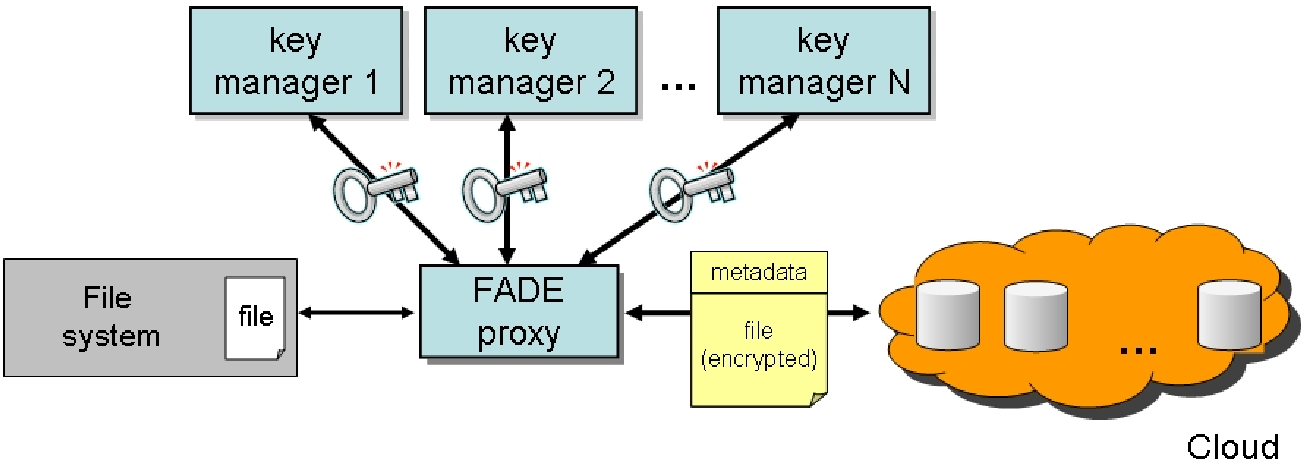


Figure 1. The FADE architecture.

FADE interface: It bridges the cloud client applications that generate data files as well as the cloud storage provider that hosts data files. It is implemented as a client-side driver that intercepts any data files sent to or received from the cloud, and applies cryptographic encryption or decryption. It also interacts with the key manager (see below) to obtain the appropriate keys for cryptographic operations. We design the FADE interface as a driver program, so as to enable it to work seamlessly with different variants of cloud-client applications without amending their implementation.

FADE key manager: It is a trusted service that is deployed outside the cloud, and manages a set of cryptographic keys that are associated with pre-defined file access policies (e.g., read/write permissions of authorized users, time expiration). Note that the FADE key manager is implemented as a *quorum scheme*, in which we create n key shares for a key K, such that we need at least $k < n$ of the key shares to recover key K. The key shares are then stored in separate secure storage devices (e.g., trusted platform module (TPM), http://www.trustedcomputinggroup.org/developers/).

It is important to note that FADE does not require any structural changes to the cloud provider. Therefore, we call FADE an *overlay* system, as it is deployed atop the existing cloud storage services.

Secure file upload/download operations with access control

We now describe the main ideas of the secure file upload/download operations, with the objectives of providing access control and assured deletion guarantees. We consider a file F to be hosted on the cloud storage provider. Let (P, Q) be the public-private key pair associated with a file access policy for F (note that (P, Q) is maintained by the key manager). Also, let $\{m\}k$ denote the operation of encrypting a data unit m with key k.

Before uploading F to the cloud, the FADE interface first generates a random key K and encrypts F with K. It also encrypts K with the public key P published by the FADE key manager. The encrypted content $(\{K\}P, \{F\}K)$, together with the metadata, will be hosted on the cloud provider. Note that the cloud storage provider cannot recover F without knowing the private key Q, which is maintained by the key manager.

To restore F, the FADE interface downloads $(\{K\}P, \{F\}K)$ from the cloud, and requests the decryption of $\{K\}P$ from the key manager. Before that, the FADE interface establishes an authenticated channel with the key manager using attribute-based encryption (ABE), such that an entity can access the channel if and only if the entity holds the credentials associated with the file access policy of F. After decrypting K, the file F can be recovered.

Secure file removal with assured deletion

To revoke a file access policy, the key manager simply destroys the associated key pair (P, Q). Files that are tied to the file access policy will then be assuredly deleted, as they cannot be recovered without the key associated with the file access policy. Even a file copy is retained, this file copy is an encrypted copy and remains unrecoverable.

Design considerations

In the design of FADE, we must take into account several considerations so that it can be practically deployed.

- **Performance**. In general, providing security guarantees for a system will degrade the performance of the system. For example, applying encryption/decryption to data can introduce computational overhead. It is important that the security mechanisms incur minimal performance overhead.

- **Economic cost**. When storing data on the cloud, cloud users need to pay for (1) the storage capacity, (2) the amount of data transfer, and (3) the number of requests made to the clouds. It is important that our developed methodologies can make data backups on the cloud secure while economically viable.

- **Version control**. It is important to have *version control* for outsourced data so that cloud clients can roll-back to extract data from earlier versions. Version control is generally coupled with the *incremental backup* technique, such that each version is incrementally built from the changes of the previous version. Our goal is to provide assured deletion for aged version whose storage time expires. However, combining version control and assured deletion is technically nontrivial, as there are data dependencies across different versions. Deleting an aged version may make the future versions unrecoverable.

3. RESULTS ACHIEVED

The project achieves the following results:

- **FADE design.** We published a conference paper on the basic design of FADE in SecureComm 2010 [1] and a journal paper on the extended design of FADE in IEEE Transactions on Dependable and Secure Computing (TDSC) [5]. In [1], we conduct empirical performance evaluation of FADE atop Amazon S3 and show that the performance overhead of FADE is minimal for large outsourced files (e.g., at least 1MB). In [5], we add two novel features: access control with attribute-based encryption (ABE) and fault-tolerant key management. We point out that the acceptance rate of TDSC is 10-12%, as reported by the TDSC editorial board in 2009. This shows the competitiveness of our work. See reference: (http://www.computer.org/portal/web/csdl/doi?doc=abs/html/trans/tq/2009/01/ttq2009010001.htm).

- **Version control**. We integrate FADE into a version-controlled cloud backup system, which exploits deduplication to reduce storage space on the cloud. We combine version control and assured deletion into a single design. We show that the FADE overhead is minimal compared to the original backup design. The shows the applicability of putting FADE atop a cloud backup system. The work is published in the ICPP workshop 2011 [3]

- **VM image storage**. We explore the robustness of cloud storage in the context of storing VM images. We extend [3] and shows the effectiveness of applying version control in reducing the storage overhead of VM images. We also show the applicability of our system via prototype experiments on the open-source cloud platform Eucalyptus. The work is published in NOMS 2012 [4]. We are currently extending our system to include assured deletion and access control.

- **Source code and demo released**. We release the source code of FADE (see http://ansrlab.cse.cuhk.edu.hk/software/fade). We also export library APIs from FADE, so that developers can integrate FADE into different data outsourcing applications. A video demo is also posted on the project website.

- **Publicity**. We have showcased a demo of FADE that is built on top of a mobile platform to the public. The press release is on http://www.cpr.cuhk.edu.hk/en/press_detail.php?id=1291.

4. PUBLICATION AND AWARDS

[1] Yang Tang, Patrick P. C. Lee, John C. S. Lui, Radia Perlman. "*FADE: Secure Overlay Cloud Storage with File Assured Deletion*." Proceedings of SecureComm 2010, Singapore, September 2010. (acceptance rate = 28/112 = 25%). Also published in Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, 2010, Volume 50, Part 7, 380-397, DOI: 10.1007/978-3-642-16161-2_22.

[2] Patrick P. C. Lee. "*Assured Deletion of Digital Files on Cloud*". IEEE COMSOC MMTC E-Letter (invited paper), 5(6), pp. 43-44, November 2010. http://committees.comsoc.org/mmc/e-news/E-Letter-November10.pdf

[3] Arthur Rahumed, Henry C. H. Chen, Yang Tang, Patrick P. C. Lee, and John C. S. Lui. *"A Secure Cloud Backup System with Assured Deletion and Version Control"***.** 3rd International Workshop on Security in Cloud Computing (CloudSec) (in conjunction with ICPP'11), Taipei, Taiwan, September 2011.

[4] Chung-Pan Tang, Tsz-Yeung Wong, and Patrick P. C. Lee. "*CloudVS: Enabling Version Control for Virtual Machines in an Open-Source Cloud under Commodity Settings*". Proceedings of the 13th IEEE/IFIP Network Operations and Management Symposium (NOMS 2012), April 2012. (acceptance rate: 55/210 = 26.2%).

[5] Yang Tang, Patrick P. C. Lee, John C. S. Lui, Radia Perlman. Secure Overlay Cloud Storage with Access Control and Assured Deletion. In IEEE Transactions on Dependable and Secure Computing, 9(6), pp. 903-916, November 2012. (acceptance rate = 10-12%, as reported by the TDSC editorial board in 2009). DOI link to the paper: http://doi.ieeecomputersociety.org/10.1109/TDSC.2012.49

SECURITY AND DETECTION PROTOCOLS FOR P2P-LIVE STREAMING SYSTEMS

Principal Investigator: John C.S. Lui
Co-Investigator (if any): Nil

Research Team Members: Yongkun Li

[1] Dept. of Computer Science & Engineering

Reporting Period: Project start date – July 31, 2012

ABSTRACT

The aim of this project is to design and implement a set of defense and detection protocols to counter content pollution attacks in P2P live streaming systems. Many people are using the P2P streaming software to watch live video content on the Internet. Recent study shows that P2P applications consume about 60-70% of Internet traffic. Pollution attack is one of the most urgent and critical security problems in P2P networks and it is known to have a disastrous effect on existing infrastructures: it can reduce the number of legitimate P2P users by as much as 85%, and at the same time, generates abundant wasteful data exchange which may deplete the communication bandwidth. In this research project, we propose a set of defense and detection method, which is a set of ``randomized'' and ``fully distributed'' protocols that can be executed by any legitimate user in a P2P system. In addition, we also aim to provide an analytical framework which can precisely quantify (a) the probability of false negative, (b) the probability of false positive, as well as (c) the distribution of time needed to detect all malicious peers. Using these models, we can dynamically adjust the system parameters so as to quickly discover malicious peers and counter the pollution attack.

1. OBJECTIVES AND SIGNIFICANCE

(a). Address a common and urgent security problem in P2P live streaming system, which is the content pollution attack.

(b). Design and implement a set of defense and detection algorithms to robustly identify malicious nodes whose intention is to pollute the streaming content.

(c). Provide mathematical models to quantify the performance measures of the distributed detection algorithms.

(d). The set of algorithms will be implemented tested. We also plan to apply for ITF to realize a set of software modules that can be used as a plug-in to many existing P2P streaming software.

2. RESEARCH METHODOLOGY

In this research, we plan to use the following methodologies to address the stated research problem:

- distributed algorithm via randomization

- coding theory for checking, verification and detection

- analytical performance evaluation to examine the detection accuracy

- system prototype and implementation to test the validity and deployment issues of our algorithms

3. RESULTS ACHIEVED

In this project, we have achieved the following:

a) Take an existing P2P live streaming protocol, evaluate and understand its implementation and communication protocols, and validate that indeed it is easy to launch a pollution attack.
b) Carry out experiments to show the feasibility and severity of pollution attack in P2P live streaming systems, both for PPLive, UUSee and PPStream systems.
c) Propose a *randomized distributed algorithm* to detect pollution attack
d) Have completed the analysis on the detection capability, as well as its performance measures like probability of false positive, probability of false negative, and time for convergence.
e) Design and implement a simulator to test our randomized algorithm
f) We collaborated with PPLive and UUSee, and helped their engineers to realize this feature into their P2P streaming software

4. PUBLICATION AND AWARDS

Please list out and number all the publications arising from the funded project. All these publications must be directly acknowledged the SHIAE funding support. The list can be numbered in alphabetic order. When referring to them for the submission in CD, name the file with corresponding reference number in square brackets as “[1].pdf”.

[1] Yongkun Li, John C.S. Lui. *``Stochastic Analysis of A Randomized Detection Algorithm for Pollution Attack in P2P Live Streaming Systems''.* IFIP Performance Conference, 2010.

[2] Yongkun Li, John C.S. Lui. *``Stochastic Analysis of A Randomized Detection Algorithm for Pollution Attack in P2P Live Streaming Systems''.* Elsevier Journal on Performance Evaluation (under fast track submission).

[3] Yongkun Li and John C. S. Lui. *``Friends or Foes: Detecting Dishonest Recommenders in Online Social Networks''.* IEEE 20th Int. Conference on Computer Communciation Networks (ICCCN) 2011.

[4] Yongkun Li, John C. S. Lui. ``On Detecting Malicious Behaviors in Interactive Networks: Algorithms and Analysis''. IEEE COMSNETS, 2012.

[5] Yongkun Li, John C.S. Lui. ``Epidemic Attacks in Network-Coding Enabled Wireless Mesh Networks: Detection, Identification and Evaluation''. Accepted for publication in the IEEE Transactions on Mobile Computing.

5. Current Work

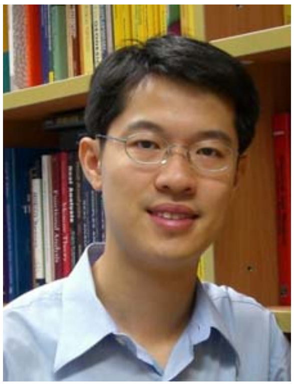
Currently, we are working on:

- Explore potential collaboration with Huawei about technology transfer
- Enhance collaboration with Tencent to explore collaboration in other areas such as malicious mobile application in Android-based systems.

AN OPPORTUNISTIC APPROACH TO CAPACITY ENHANCEMENT IN WIRELESS MULTIMEDIA NETWORKS
Completion Report

Principal Investigator: SO, Man Cho Anthony [1]
Co-Investigator (if any): ZHANG, Yingjun Angela [2]

[1] Department of Systems Engineering and Engineering Management
[2] Department of Information Engineering



Project Start Date: 1st July 2009
Completion Date: 30th June 2011

ABSTRACT

It has long been realized that many communication and networking problems can be formulated as optimization problems, for which deep theories and efficient algorithms are available. As it turns out, most if not all of those formulations either assume that the data defining the optimization problem – such as channel state information, users' preferences, etc. – are exactly known, or simply ignore the stochastic nature of the data. However, such an approach would often lead to sub-optimal or even infeasible solutions for practical systems. In this project, we propose to use chance constrained programming techniques to deal with uncertain data in communication and network optimization problems, particularly those that arise in wireless multimedia communication systems. Our approach is motivated by the observation that in many communications applications, the constraints need not be satisfied all the time. By adopting the chance constrained programming approach, we are able to utilize the distributional characteristics of the data (which are well studied in the communications literature), thus turning the underlying uncertainties to our advantage and opening new avenues for improving the performance of existing systems. In particular, using tools from probability theory and optimization theory, we have developed a fairly general framework for obtaining efficiently solvable formulations of various chance constrained communication and network optimization problems. To demonstrate the practical value of our approach, we have tested it on several resource allocation problems in wireless channels. Simulation results indicate that our approach can offer substantial gain in performance when compared to existing approaches. As the use of chance constrained programming techniques in the area of communications is still relatively new, we believe that our proposed study will have a significant impact in the area.

1. OBJECTIVES AND SIGNIFICANCE

The goal of this project is twofold:

1. To develop algorithmic machineries to handle data uncertainties and to exploit their distributional properties in optimization problems that arise from wireless multimedia communication applications, and
2. to demonstrate the benefit of incorporating distributional information in optimization by comparing the performance of existing approaches – most of which do not exploit such information – to the performance of our proposed algorithms.

Results from the proposed study will facilitate the incorporation of distributional information of the data in the optimization process, thus allowing one to design high performance communication systems.

2. RESEARCH METHODOLOGY

One of the central problems in communications is to obtain an allocation decision (such as transmission power, bandwidth, etc.) that optimizes certain performance metric while satisfying various service constraints. Oftentimes, the data defining such optimization problems involve channel state information, which is not deterministic but rather follows a certain probability distribution. One way to tackle this is to require the allocation decision to be robust against *every* possible realization of the data. In other words, the service constraints must be satisfied under any circumstances. However, such an approach could be highly conservative. In fact, in many applications, such as multimedia applications, it is not necessarily to satisfy the constraints all the time, as users can tolerate an occasional dip in the quality of service. Thus, one can potentially obtain a much better performance by exploiting the distributional properties of the data. This motivates us to consider chance constrained formulations of the aforementioned optimization problems. Informally, a chance constrained program is an optimization problem of the form:

$$(*) \quad \min_{x \in X} f(x) \text{ subject to } \Pr_{\xi}\left[F(x,\xi) \le 0\right] \ge 1-\alpha$$

where $\xi \in R^d$ is a random vector, $X \subset R^n$ is a non-empty feasible set, and $\alpha \in (0,1)$ is a tolerance parameter. Note that the real-valued function F depends on both the decision x and the realization of the random parameter ξ. In particular, for each fixed x, $F(x,\cdot)$ is a random variable and hence it makes sense to talk about the probability $\Pr_{\xi}\left[F(x,\xi) \le 0\right]$. By adjusting the tolerance parameter, we have a tradeoff between the conservatism and optimality of the solution.

Of course, there is no free lunch, and the catch with (*) is that it is generally a computationally difficult problem (even when the set X and the objective function f are convex). To circumvent this, one can develop a so-called *safe tractable approximation* of the probabilistic constraint in (*) – that is, a system of constraints H such that (i) x is feasible for (*) whenever it is feasible for H, and (ii) the constraints in H are efficiently computable. Currently, there are two major approaches for obtaining safe tractable approximations of the generic chance constrained problem (*). The first is the *moment generating function* approach, where the goal is to derive an efficiently computable bound on the moment generating function of the random variable $F(x,\cdot)$. The second is the *concentration inequality* approach, in which suitable concentration inequalities are used to obtain an efficiently computable upper bound on $\Pr_{\xi}\left[F(x,\xi) > 0\right]$. Both of these approaches are particularly useful when there is relatively precise information on the distribution of the uncertain data (such as support, moments, etc.), and the function $F(x,\cdot)$ is affine in the random parameter ξ for each x and the components of ξ are independent. In this project, we advance the chance constrained programming approach on both the practical and theoretical fronts. On the practical front, we focus on two important application scenarios, namely Slow Adaptive Orthogonal Frequency Division Multiple Access and Optimal Spectrum Sharing in Multiple-Input Multiple-Output Cognitive Radio Networks. Through the design of high powered optimization algorithms, we demonstrate how the chance constrained programming approach can significantly improve the performance of the systems in question. On the theoretical front, we employ sophisticated tools from probability theory to vastly extend the power and applicability of the aforementioned moment generating function approach and concentration inequality approach. In particular, we obtain safe tractable approximations of chance constrained optimization problems in which F is a *vector-valued* function, and the components of ξ are not necessarily independent but have "nice" dependence structure. Besides their theoretical value, these results have interesting implications in many communications problems, such as robust beamforming design.

3. RESULTS ACHIEVED

3.1 Slow Adaptive Orthogonal Frequency Division Multiple Access (OFDMA) (Ref. [1] in Section 4)

Adaptive OFDMA, where the subcarrier and power allocation to mobile users is dynamically adjusted according to wireless channel conditions, has recently proven to be a powerful technique that can significantly enhance the capacity of wireless systems. Unfortunately, considering the fact that wireless channel fading varies in the order of milliseconds, the on-the-fly adaptation of resource allocation quickly becomes computationally infeasible. To address this issue, we design a slow adaptive OFDMA system based on chance constrained programming techniques. Our formulation guarantees the short-term data rate requirements of individual users except in rare occasions. We then exploit the special structure of the probabilistic constraints in our problem to construct safe tractable constraints (STC) based on recent advances in the chance constrained programming literature. Finally, we design an interior-point algorithm that is tailored for the slow adaptive OFDMA problem, since the formulation with STC, although convex, cannot be trivially solved using off-the-shelf optimization software. Our algorithm can efficiently compute an optimal solution to the problem with STC in polynomial time.

Simulation results indicate that our slow adaptive OFDMA system is indeed competitive with typical fast adaptive OFDMA schemes, which are much more computationally intensive. In Figure 1, we compare the spectral efficiency of slow adaptive OFDMA (in which the outage probability is at most 0.1 for each user) with that of fast adaptive OFDMA (in which zero outage of short-term data rate requirement is ensured for each user). Here, we assume that the control signaling overhead consumes a bandwidth equivalent to 10% of a slot length every time subcarrier allocation is updated. Note that within each window that contains 1000 slots, the control signaling has to be transmitted 1000 times in the fast adaptation scheme, but once in the slow adaptation scheme. The figure shows that although slow adaptive OFDMA updates subcarrier allocation 1000 times less frequently than fast adaptive OFDMA, it can achieve on average 71.88% of the spectral efficiency. Considering the substantially lower computational complexity and signaling overhead, slow adaptive OFDMA holds significant promise for deployment in real-world systems.

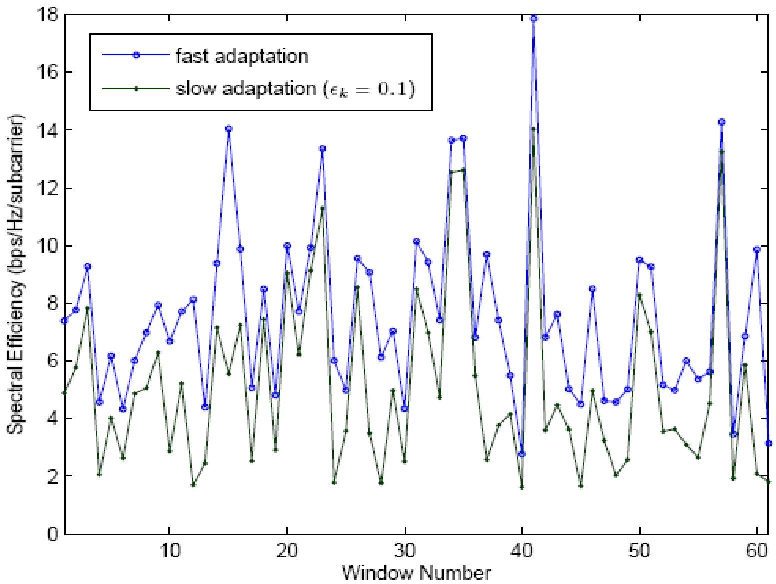


Figure 1: Comparison of System Spectral Efficiency between Fast and Slow Adaptive OFDMA Schemes

3.2 Optimal Spectrum Sharing in Multiple-Input Multiple-Output (MIMO) Cognitive Radio Networks (Ref. [2] in Section 4)

In cognitive radio (CR) networks with multiple-input multiple-output (MIMO) links, secondary users (SUs) can exploit "spectrum holes" in the space domain to access the spectrum allocated to a primary system. However, they need to suppress the interference caused to primary users (PUs), as the secondary system should be transparent to the primary system. Thus, the SU faces an optimization problem in which it is trying to maximize its throughput while keeping the interference temperature at the primary receivers below a certain threshold.

Unlike traditional MIMO systems, SUs may not have the luxury of knowing the channel state information (CSI) on the links to PUs. This presents a key challenge for a secondary transmitter to steer interference away from primary receivers. In this study, we focus on three scenarios, namely when the SU transmitter has complete, partial, or no knowledge about the channels to the PU receivers. In particular, when complete CSI is not available, the interference-temperature constraints are to be satisfied with high probability, thus giving rise to chance constraints in the respective optimization problems. Our contribution in this study is fourfold. First, by analyzing the distributional characteristics of MIMO channels, we propose a unified homogeneous quadratically constrained quadratic program (QCQP) formulation that can be applied to all three scenarios with different levels of CSI knowledge and with either deterministic or probabilistic interference-temperature constraints. The homogeneous QCQP formulation, though non-convex, is amenable to semidefinite programming (SDP) relaxation methods. Secondly, we show that the SDP relaxation admits no gap when the number of primary links is no larger than 2. A polynomial-time algorithm is presented to compute the optimal solution to the QCQP problem efficiently. Thirdly, we propose a randomized polynomial-time algorithm for constructing a near-optimal solution to the QCQP problem when there are more than 2 primary links. Finally, we show that when the secondary transmitter has no CSI on the links to primary receivers, the QCQP problem reduces to a matrix eigenvalue-eigenvector decomposition problem. In this case, the optimal beamforming solution can be obtained very efficiently without the need to solve the QCQP.

3.3 Outage Constrained Robust Transmit Optimization for Multiuser MISO Downlinks (Ref. [3] in Section 4)

Recently, robust transmit beamforming has drawn considerable attention because it can provide guaranteed receiver performance in the presence of channel state information (CSI) errors. Assuming complex Gaussian distributed CSI errors, we investigate the robust beamforming design problem that minimizes the transmission power subject to probabilistic signal-to-interference-plus-noise ratio (SINR) constraints. The probabilistic SINR constraints in general have no closed-form expression and are difficult to handle. Based on a Bernstein-type inequality for quadratic forms of complex Gaussian random variables, we propose a conservative formulation to the robust single-cell beamforming design problem. The semidefinite relaxation technique can be applied to efficiently handle the proposed conservative formulation. Simulation results show that, in comparison with existing methods, the proposed method is more power efficient and is able to support higher target SINR values for receivers.

3.4 Safe Tractable Approximations for Joint and Nonlinearly Perturbed Chance Constraints (Ref. [4] in Section 4)

In this study, we aim at extending the current theory of chance-constrained optimization, so that more applications can benefit from the model. Indeed, most existing safe tractable approximations of the chance constrained optimization problem (*) assume that (i) F is a real-valued function (and hence there is only a single chance constraint), (ii) $F(x,\cdot)$ is affine in the random parameter ξ for each x, and (iii) the components of ξ are independent. However, there are applications in which we need to deal with a vector-valued F (the so-called *joint chance constraints*), or with the case where F is not affinely dependent on ξ. For instance, in certain robust beamforming design problems, the function F has a quadratic dependence on ξ, i.e., $F(x,\xi)=\xi^T A(x)\xi$. We developed safe tractable approximations for the joint chance constrained programming problem, as well as for a large class of nonlinearly perturbed chance constrained programming problems (which include the quadratically perturbed chance constraints mentioned above). Moreover, the approximations can be formulated as semidefinte programs, which imply that they can be solved using off-the-shelf solvers. Our results greatly enrich the arsenal of tools available for modeling and solving chance-constrained optimization problems.

4. PUBLICATION AND AWARDS

[1] William Wei-Liang Li, Ying Jun (Angela) Zhang, Anthony Man-Cho So, and Moe Z. Win, “Slow Adaptive OFDMA Systems through Chance Constrained Programming,” *IEEE Transactions on Signal Processing* 58(7): 3858-3869, 2010.

[2] Ying Jun (Angela) Zhang, and Anthony Man-Cho So, “Optimal Spectrum Sharing in MIMO Cognitive Radio Networks via Semidefinite Programming,” *IEEE Journal on Selected Areas in Communications, Special Issue on Advances in Cognitive Radio Networking and Communications* 29(2): 362-373, 2010.

[3] Kun-Yu Wang, Tsung-Hui Chang, Wing-Kin Ma, Anthony Man-Cho So, and Chong-Yung Chi, “Probabilistic SINR Constrained Robust Transmit Beamforming: A Bernstein-Type Inequality Based Conservative Approach,” Proceedings of the 2011 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2011), pp. 3080-3083, 2011.

[4] Sin-Shuen Cheung, Anthony Man-Cho So, and Kuncheng Wang, “Linear Matrix Inequalities with Stochastically Dependent Perturbations and Applications to Chance-Constrained Semidefinite Optimization,” accepted for publication in *SIAM Journal on Optimization*, 2012.

COMPUTER-AIDED SECOND LANGUAGE LEARNING THROUGH SPEECH-BASED HUMAN-COMPUTER INTERACTIONS

Principal Investigator: Helen Meng [(1)]
Co-Investigator (if any): Virginia Yip [(2)], Pauline Lee [(3)]

Research Team Members:
Wai-Kit Lo, Research Associate [(1)]
Kun Li, Research Assistant[(1)] (now a PhD student at CUHK)
Peng-Fei Liu, Research Assistant [(1)] (now a PhD student at CUHK)
Ka-Ho Wong, Research Assistant [(1)] (now a PhD student at CUHK)
Ming-Xing Liu, Research Assistant[(1)] (now a Research Assistant at Baptist University)

[(1)] Department of Systems Engineering & Engineering Management
[(2)] Department of Linguistics and Modern Languages
[(3)] Independent Learning Center

Project Start Date: 1st July 2009
Completion Date: 30th June 2011

ABSTRACT

This project aims to develop speech recognition and synthesis technologies that assist second language (L1) acquisition of English by adult learners whose primary language (L1) may be Putonghua or Cantonese. English is the lingua franca of our world. As world events such as the Beijing Olympics 2008 and the Shanghai World Expo 2010 continue to be hosted in different major locations across China, it is of prime importance that our people acquire communicative competence in English. However, the process of second language acquisition (L2) is affected by well-established perceptions of sounds and articulations in the primary language (L1). Putonghua is the official dialect of China and Cantonese is a major dialect predominant in Hong Kong and South China. These Chinese dialects have stark linguistic contrasts in comparison with English. Hence, we often observe notable L1 (i.e. Chinese) interference with L2 (i.e. English) speech. The interference can become ingrained with age and hamper acquisition of proficiency, especially for adult L2 learners. Improvements require persistent and individualized training. Recent advancements in multimedia technologies and Web technologies have opened up new possibilities in computer-aided language learning. Automatic speech recognition technologies offer a new means of productive training. Multimedia technologies such as computer graphics, animation and automatic speech synthesis technologies offer a new means of perceptual training. The Web offers universal access to such computer-aided, productive and perceptual language training through a round-the-clock, self-paced and personalized platform. The long-term objective of this project is to develop speech technologies on a Web-based learning platform that can potentially assist the vast population of Chinese learners in achieving communicative competence and overall proficiency in spoken English.

1. OBJECTIVES AND SIGNIFICANCE

(i) Design and develop a corpus-based study to identify spoken language inaccuracies in L2 English based on learners whose L1 is Putonghua, as well as learners whose L1 is Cantonese[1]

Significance: There is a lack of speech corpora for L2 English in general, not to mention ones that are based on speakers whose L1 is Chinese (either Putonghua or Cantonese). We have developed the recording software and designed the text prompts for corpus collection.

(ii) Perform acoustic analysis to describe the spoken language inaccuracies

Significance: This part of the study enables us to understand where acoustic deviations occur in L2 speech and forms an important basis for subsequent work in mispronunciation detection and diagnosis.

(iii) Develop error detection techniques for spoken language inaccuracies, based on approaches to automatic speech recognition

Significance: This is an important technological development that can support productive training in CALL (computer-aided language learning) applications. At present, most of the COTS (commercial off-the-shelf) software supports only perceptual training. While perceptual training targets listening in order to discriminate between different sounds, productive training targets articulation in order to enunciate different sounds.

(iv) Design human-computer speech-based interactions that target salient, erroneous speech productions for a learner, with automatic response generation techniques that highlight detected errors and provide corrective feedback

Significance: Most CALL applications output a score as feedback. However, differences in scores do not necessarily reflect differences in proficiency, because scoring is affected by a variety of factors, including speaking style, ambient conditions, equipment, gender, etc. In this project, we strive to generate feedback that is more conducive towards pronunciation training, i.e. not only to detect the presence of mispronunciations, but also to diagnose how or why the speech segment is incorrect.

(v) Develop a personalized platform that integrates spoken language inaccuracy detection and corrective feedback generation into an interactive, computer-aided language learning system.

Significance: We believe it is very important to demonstrate the efficacy of our technologies in a research prototype. Systems development involves significant engineering effort.

2. RESEARCH METHODOLOGY

We elaborate on the research methodology used in each research objective laid out in Section 1.

Objective 1 – speech corpus design and collection

We have previously collected the CU-CHLOE (CU Chinese Learners of English Corpus) which includes L2 English recordings from 100 native Cantonese and 100 native Putonghua speakers. We have also designed a set of recording prompts that cover all phonemes in English, as well as common phonetic confusions. CU-CHLOE primarily covers L2 phenomena in segmental phonology. .

In addition to the above, we have developed a set of new recording prompts that cover L2 phenomenon in suprasegmental phonology. The scope covers lexical stress, reduced or unreduced function words, narrow focus, intonation, prosodic disambiguation in semantic interpretation and short story reading.

[1] L1 denotes primary language, L2 secondary language.

Objective 2 – acoustic analysis of L2 speech

We have previously developed a set of handwritten phonological rules to capture phonetic confusions in L2 speech. The rules are written by a linguist with reference to previous publications as well as our speech corpus. We have developed a data-driven approach that enables us to automatically derive such phonological rules from the speech corpus. This approach offers the advantage of rapid rule development, where the rule set can achieve a broad coverage of observed mispronunciations. We have also performed analysis on the pitch trajectories of L2 speech, in relation to prosodic patterns.

Objective 3 – develop automatic speech recognition techniques for mispronunciation detection

The phonological rules obtained above are used to generate pronunciation variants based on the canonical pronunciations of English words. In this way, we can create an extended pronunciation dictionary (EPD) that contains, for each word, its canonical pronunciation as well as possible mispronunciations. The EPD is incorporated into our automatic speech recognition engine which can then explicitly model the mispronunciations acoustic-phonetically.

Objective 4 – design human-computer speech-based interactions

We have designed a set of prompts that are organized as three assignments. These have been incorporated into a system prototype (named Enunciate) with which we have and continue to conduct pilot runs in English classes offered by the English Language Teaching Unit (ELTU) and also language laboratories under the Independent Learning Center of CUHK. Through this exercise, we were able to gauge the pedagogical efficacy of our research prototype.

Objective 5 – research and development of a personalized language learning platform

We believe that it is important for each student to maintain his/her usage history of the system, which include speech recordings, detected errors and an overall impression of his/her performance. It is also important to provide each student with an analysis of his/her frequent mispronunciations to pinpoint areas in need of improvement.

3. RESULTS ACHIEVED

We elaborate on the results achieved in relation to each research objectives laid out in Section 1.

Objective 1 – speech corpus design and collection

We have recruited a linguist to phonetically transcribe the CU-CHLOE corpus, including the sections of the AESOP fable, minimal word pairs and phonemic sentences. The section on TIMIT utterances is not transcribed as it will be too costly due to the large volume of speech. As for the suprasegmental corpus, we have recorded from 80 native Cantonese and 80 native Putonghua speakers.

The recording tool that we have developed to support this effort has been licensed by Academia Sinica, Taiwan and Waseda University Japan. These institutions are collecting similar speech corpora for Taiwanese-accented English and Japanese-accented English, respectively. They have also invited us to be co-investigators of their research grants. We have allowed other research institutions to freely download the software, which is currently in use by groups in India, Thailand, Vietnam, etc.

Objective 2 – acoustic analysis of L2 speech

We have developed a data-driven approach that automatically generates context-sensitive rules that capture L2 phonetic confusions, based on mispronunciations observed in L2 data. The context-sensitive rules are in the format of $\varphi \rightarrow \psi / \lambda \_ \rho$, which denotes that phone φ may be substituted by the phone ψ when it is preceded by the phone λ and followed by the phone ρ. We have also developed a rule selection method that is statistically-driven to optimize an F-measure based on precision and recall of mispronunciations. Running the data-driven rule derivation procedure on the CU-CHLOE corpus generates over 200 phonological rules.

Objective 3 – develop automatic speech recognition techniques for mispronunciation detection
We have realized the approach to mispronunciation detection described above. We have also demonstrated the use of finite-state transducers that can represent the pronunciation variants in the form of an extended recognition network (ERN). The ERN offers a compact representation that incorporates the EPD into recognition.

Objective 4 – design human-computer speech-based interactions
Our system has been used in the Faculty of Education's Professional Development Course entitled "Proficiency in English Phonetics". It has also been used in the pronunciation training classes offered by the English Language Teaching Unit (ELTU). In addition to in-class use, our system is accessible across campus, and especially in the language laboratories managed by the Independent Learning Center. We have asked users to volunteer their time to fill our a questionnaire that gives us a user evaluation of our system. We have collected over 270 questionnaires and very encouraging feedback has been received. 87% of the respondents indicated that the system was easy to use; 75% indicated that the speed of responses was acceptable; and 73% indicated that the system's feedback was helpful. The system is hosted in our laboratory's research server and continues to be used by classes offered by ELTU and is accessible across CUHK campus.

Objective 5 – R&D for a personalized language learning platform
As mentioned earlier, the Enunciate system has been made available across CUHK campus for students to use. We have implemented an automatic scoring system that can provide a score to convey an overall impression of a student's performance in each speech recording or in an entire exercise. Additionally, we have put in a student accounts system where each account is password-protected so that user data could be maintained securely across usage sessions. This also allowed us to implement user histories so that the user can access their previous responses to a given exercise. Additionally, an assignment submission function for students to submit their data (including speech data) to their course teacher was implemented. This feature in combination with a new teacher administrative interface enabled teachers to view their student data online.

4. PUBLICATION AND AWARDS

Please list out and number all the publications produced under the funded project. All these publications must be directly acknowledged the SHIAE funding support and stated the affiliation with the Institute. The list can be numbered in alphabetic order. When referring to them for the submission in CD, name the file with corresponding reference number in square brackets as "81150xx-Pub[1].pdf".

Publications:

[1] Shuang Zhang, Kun Li, Wai-Kit Lo and Helen Meng, "Perception of English Suprasegmental Features by non-native Chinese Learners", Proceedings of the Fifth International Conference on Speech Prosody, 2010.

[2] Wai-Kit Lo, Shuang Zhang and Helen Meng, "Automatic Derivation of Phonological Rules for Mispronunciation Detection in a Computer-Assisted Pronunciation Training System", INTERSPEECH2010, 2010.

[3] Helen Meng, Wai-Kit Lo, Alissa M. Harrison, Pauline Lee, Ka-Ho Wong, Wai-Kim Leung and Fanbo Meng, "Development of Automatic Speech Recognition and Synthesis Technologies to Support Chinese Learners of English: The CUHK Experience," in the Proceedings of the 2nd Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA2010), December 2010. (Best Oral Paper Award)

[4] Ka-Wa Yuen, Wai-Kim Leung, Peng-fei Liu, Ka-Ho Wong, Xiao-jun Qian,Wai-Kit Lo and Helen Meng, "Enunciate: An Internet-Accessible Computer-Aided Pronunciation Training System and Related User Evaluations ", in the Proceedings of Oriental COCOSDA 2011, Hsinchu, Taiwan, 26-28 October, 2011.

[5] Pengfei Liu, Ka Wa Yuen, Wai Kim Leung, Helen Meng, "mEnunciate: Development of a Computer-Adied Pronunciation Training System on a Cross-Platform Framework for Mobile, Speech-Enabled Application Development, in the Proceedings of International Symposium on Chinese Spoken Language Processing, 2012.
[6] Kun Li and Helen Meng, "Perceptually-motivated Assessment of Automatically Detected Lexical Stress in L2 Learners' Speech," Proceedings of International Symposium on Chinese Spoken Language Processing, 2012.
[7] Jia Jia, Wai Kim Leung, Ye Tian, Lianhong Cai and Helen Meng, "Analysis on Mispronunciations in CAPT based on Computational Speech Perception," Proceedings of International Symposium on Chinese Spoken Language Processing, 2012.

Invited Talks:

- Meng, H., "Developing Automatic Speech Recognition to support Computer-Aided Pronunciation Training for Chinese Learners of English," Speech Workhop 2010, Taiwan. http://office.csie.ncyu.edu.tw/sws2010/program.html
- Meng, H. Lee, P., Lo, W.K., and V. Yip. "Computer-aided Language Learning: applications for early childhood education", CUHK Workshop on Bilingualism and Language Acquisition, Distinguished Speakers Lecture Series, 17 March 2010 http://www.cuhk.edu.hk/lin/cbrc/workshop/#program
- Meng, H., "Developing Speech Recognition and Synthesis Technologies to Support Computer-Aided Pronunciation Training for Chinese Learners of English" The 23rd Pacific Asia Conference on Language, Information and Computation, December 2009. http://paclic23.ctl.cityu.edu.hk/PACLIC23_invitedSpeakers.html
- Meng, H. Keynote Speech, "Developing Speech Recognition and Synthesis Technologies to Support Computer-Aided Pronunciation Training for Chinese Learners of English", National Conference on Man-Machine Speech Communication NCMMSC, November 2011

Awards

- Best Oral Paper Award, Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA) 2010 – acceptance rate of oral papers was 23.4%
- Tier3 Funding from the Innovation and Technology Fund, "An Internet-based Platform with Automatic Speech Recognition Technologies to Support Online Computer-Aided Pronunciation Training for Chinese Learners of English (2009-2011). PI: Helen Meng. Funding: HKD998,200.
- Tier3 Funding from the Innovation and Technology Fund, "Research and Development of Text-to-Audiovisual Speech Synthesis Technologies on a Mobile E-learning Platform for Computer-Aided Pronunciation Training (2012-2014) PI: Helen Meng. Funding: HKD998,200.

PATTERN COMPUTATION FOR COMPRESSION AND PERFORMANCE GARMENT

Principal Investigator: Charlie C.L. Wang [(1)]
Co-Investigators: Kin-Chuen Hui [(1)], Penelope Watkins [(2)]

Research Team Members: Yunbo Zhang [(1)], Siu-Ping Mok [(1)], Juncong Lin [(1)], Shengjun Liu [(1)]

[(1)] Dept. of Mechanical and Automation Engineering, The Chinese University of Hong Kong
[(2)] Center of Fashion Science, London College of Fashion

Project Start Date: 1st July 2008
Completion Date: 30th June 2010

ABSTRACT

In the clothing industry, compression garments are increasingly being used to mould and confine the malleable shape of the human body. The garment design may require variation in pressure ranging from normal through increased strain in specific areas for the particular application. Therefore compression garments need to be customized because body shapes differ and require different strain distribution. The 3D body shape can be obtained by any popular 3D data-acquisition means (e.g., a human body laser scanner). However, it is the 2D pattern – fabricated into the 3D shape combined with the fabric parameters, which brings about the desired strain distribution – that has to be determined. At present, the 2D pattern design and garment-to-body fit is accomplished through trial-and-error. This subjective procedure is inefficient, inaccurate and costly, as many prototypes have to be produced. Therefore, a computer program is desirable that can automatically generate 2D patterns from an input of the 3D shape that, when fabricated into the final 3D garment, the 2D pattern profiles give the designed strain and pressure on the human body. In this research we propose to develop techniques to compute 2D meshes that can generate a user-defined strain distribution through proper distortion, when folded onto the 3D body.

1. OBJECTIVES AND SIGNIFICANCE

There are three major objectives to reach in this research project:

- Develop a new physical/geometric approach that is able to model complicated elastic behaviors of fabrics, and determine the 2D patterns of a given 3D mesh surface satisfying the given strain and/or normal pressure distribution.
- Investigate numerical schemes to compute optimal 2D patterns efficiently.
- Develop a test-bed to evaluate the material-related coefficients in the developed physical/geometric model and verify the simulation results according to experimental tests.

2. RESEARCH METHODOLOGY

In order to reach the objectives of this project, we conduct several research works that are presented below.

2.1. A new physical/geometric approach for 2D pattern computation

We develop a new physical/geometric approach. Every triangle edge is simulated by a tensile truss bar, and the equilibrium equation will be established at every vertex on the given mesh surface. Let f_e be the force on truss bar $e \in E(v)$ with $E(v)$ denoting the collection of edges linking to v, we have the following equilibrium equation in the tangent plane P^v of v

$$\sum_{e \in E(v)} P^v(f_e) \equiv 0$$

where $P^v(\cdots)$ stands for the projection of a vector onto the plane P^v (see Figure 1).

Motivated by the strain-stress theories in solid mechanics, we stipulate that the relationship between the tensile strain and the normal pressure can be modeled as

$$p_v = s\int_0^{2\pi} \kappa_n(\theta)\sigma(\theta)d\theta$$

where $\kappa_n(\theta)$ is the normal curvature in direction θ on the tangent plane at the surface point, $\sigma(\theta)$ denotes the normal stress in θ, and *s* is a material-related coefficients to be determined.

The behavior of elastic fabric pieces on a compression garment should not only be governed by the equilibrium equation, but also have correct geometric constraints to ensure that the 3D shape can really be fabricated from 2D patterns. This leads to the following developability constraints imposed on every interior mesh vertex. For a vertex v on M, using $\alpha_f(v)$ to symbolize the vertex angle of triangular face f at v (see Figure 2), the condition

$$\sum_{f \in F(v)} \alpha_f(v) \equiv 2\pi$$

should be satisfied if all the faces *f* around *v* are to be flattened without any distortion (i.e., cracks and/or overlapping), where *F*(*v*) is the set of triangles incident at *v*. Note that, in our problem setting, the triangles in 3D have already been purposely distorted in order to generate compression. Therefore, we must use the angle of face *f* in its original relaxed state for $\alpha_f(v)$.

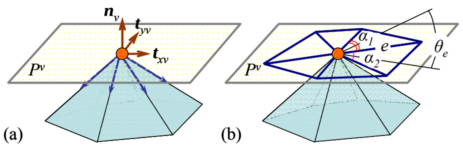


Figure 1. Illustration for the physical model at an interior vertex: (a) a local frame is constructed at every internal vertex on its tangent plane, and (b) the strain equivalent equation can then be derived on each tangent plane.

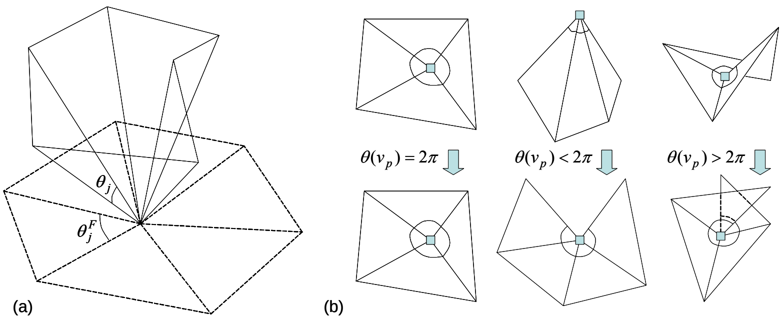


Figure 2. Illustration for flattenability: (a) the inner angles before and after flattening the triangles around a vertex, and (b) Euclidean vertex (left), spherical vertex (middle) and hyperbolic vertex (right).

2.2. Numerical simulation

In order to efficiently determine the optimal value of strain on truss bars according to the above physical/geometry mode, we linearize the governing equation for normal pressure and strains, the strain equivalent equation, and the geometric constraints. Then, the optimal strain on each truss bar, whose computation is originally non-linear, can be computed by iteratively solving a sequence of linear equation systems. Moreover, during the numerical computation, the requirement of keeping truss bars in tensile must be safeguarded for generating compression.

2.3. Test-bed development

To evaluate the material-related coefficients and verify the simulation results, a test-bed needs to be built. The test-bed consists of several cylindrical, conical and spherical models in various dimensions. Compression pressure sensor will be integrated on the surface of the models. When warping difference fabrics onto these models, the value of compression pressure will be evaluated. Meanwhile, several markers will be printed on the fabrics which can be identified by an imaging system to report the corresponding strains. The pressure and strain values will be used to calibrate the material-related coefficients in our physical/geometric model. This setup will also be employed to verify the simulation results generated by the numerical computation.

3. RESULTS ACHIEVED

3.1. Physical/geometry approach for 2D pattern computation and numerical simulation

We have implemented the proposed method in a prototype program written in C++. Several experimental results have been obtained. We tested the developed model on a medical elastic brace example with four different configurations of normal pressure assignment. In Figure 3, (a)-(d) display the results of strain distribution and planar patterns under different normal pressure assignments. It is worth mentioning that the configuration in Figure 3(d) cannot be simulated by our previous research of the woven-based method. For comparison, the 2D patterns generated by the least-squares conformal map (LSCM) and the length-preserved free-boundary (LPFB) are also given in Figure 3(e) and (f). Figure 4 shows the investigation about the speed of convergence in numerical computation. It is seen clearly from the figure that both norms drop quickly after a few iterations.

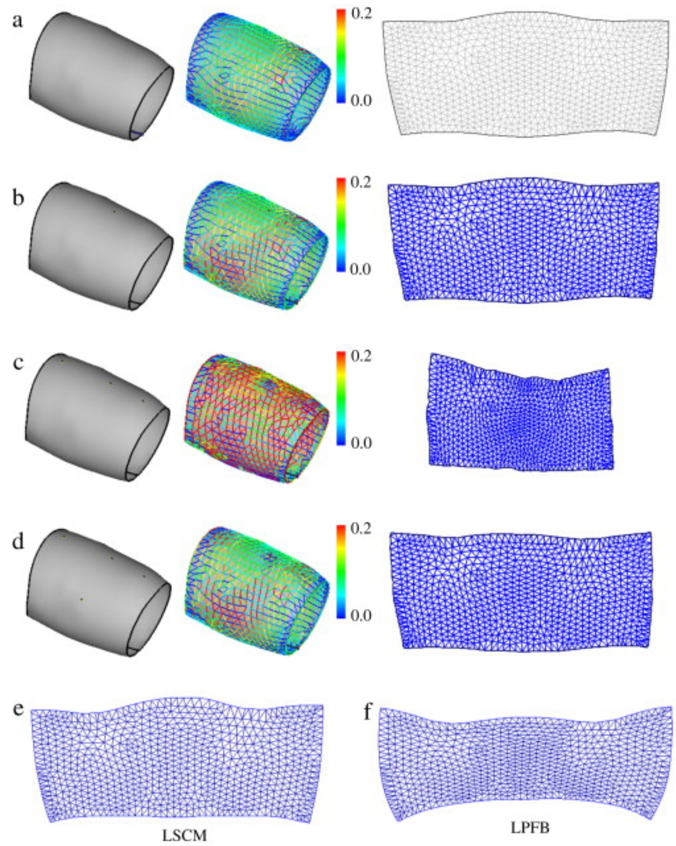


Figure 3. The medical brace example with different normal pressures specified at the anchor points (colored black): (a) the brace is free—no specific normal pressure is assigned, (b) brace A—with one anchor point, (c) brace B—with three anchor points, and (d) brace C—with four anchor points. The color maps show different strain distributions corresponding to the four different normal pressure configurations. The final 2D patterns are also given (in red): they are superimposed with the one from the free brace in (a), for comparison. (e) and (f) give the planar meshes generated by the least-squares conformal map (LSCM) and the length-preserved free-boundary (LPFB) respectively, which are very different from the patterns generated by this approach..

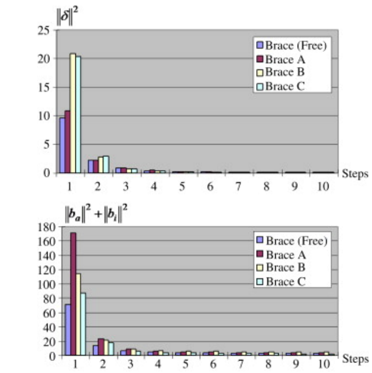


Figure 4. Convergence of the numerical iterations for the example of Figure 3.

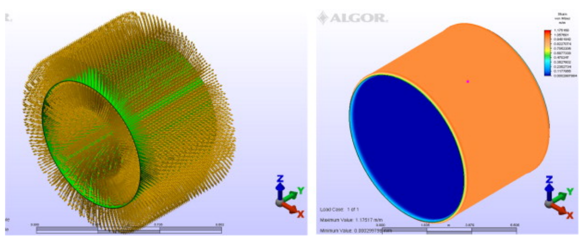


Figure 5. Testing the pressure–strain relationship by FEA software.

Table 1 FEA tests on neoprene cylinders

Radius R (mm)	Pressure (N/m^2)	Strain
64.75	150	0.9675
64.75	100	0.6450
64.75	50	0.3225
54.75	150	0.8176
54.75	100	0.5450
54.75	50	0.2725

The neoprene has Young's Modulus 1.85 GPa and density 1210 kg/m^3.

In order to verify the proposed physical/geometric approach, we compare the results computed by our scheme with the simulation results from finite element analysis (FEA). Two cylinder models under different compressions are tested by the commercial FEA software Algor. As shown in Figure 5, the uniform normal pressures are added onto the cylinder, and the strain distribution can be generated by the system. We measure the strain at the red point shown in the right figure and obtain the results listed in Table 1. It is not difficult to find that the value of strain is approximately proportional to the normal pressure loadings and the radius of the cylinder (i.e., inverted to the curvature).

Moreover, the method developed in this research leads to a novel flexible and accurate method for computing the flattened planar cloth patterns from a three-dimensional design. The new method is named as *WireWarping++*, which is a flexible and robust extension of *WireWarping* by introducing a new type of feature curves named *elastic feature*, which brings flexibility to shape control of the resultant 2D patterns. On these new feature curves, instead of strictly preserving the exact lengths, only the ranges of their lengths are controlled. To achieve this function, a multi-loop shape control optimization framework is proposed to find the optimized 2D shape among all possible flattening results with different length variations on those elastic feature curves, while the lengths of other feature curves are kept unchanged. Besides, we also present a topology processing algorithm on the network of feature curves to eliminate cases that lead to numerical singularity. Experimental results show that the WireWarping++ can successfully flatten surface patches into 2D patterns with more flexible shape control and more robust numerical performance (see Figure 6 for an example).

Figure 6. Photos of the jeans pants fabricated from the patterns generated by WireWarping versus those generated by WireWarping++. The one generated by WireWarping++ fits the back waist band and the back yoke much better – this is the comment made by a fashion specialist.

3.2. Test-bed Development

As aforementioned, several cylindrical, conical, spherical and freeform models with various dimensions are used in our test. Three sets of models with cylindrical, conical (which actually is a cone without head) and ellipsoidal shapes are made using *Rapid Prototyping* (RP) to imitate different parts of human body such as upper arm, forearm, lower leg and thigh (see Figure 7). Force sensors are used in our test to measure force applied onto the geometric RP models at various location during stretching. Since the force sensor will be placed onto the curved surface of the models, it is preferably thin and flexible. In view of these, the Force Sensing Resistors (FSR - Part No. 402) produced by Interlink Electronics are rather suitable for our test, which are therefore adopted in our test-bed development. The testing machine mainly comprises 5 shafts, 2 locked shafts, 12 bearings with holders, 2 handles, a fixed holder and an aluminum frame as shown in Figure 8. During the test, a sample fabric is inserted zigzag over each bearing with holder and each of its ends is locked on the locked shaft. In the middle of the machine is a shaft on a fixed holder. There is no bearing on this holder as it is used to hold the RP models for observation of fabric extension while the remaining shafts and locked shafts are on holders with bearings so that they can be easily rotated. The aluminum frame is used to hold all these components tightly. After the fabric is locked, we rotate the handles. The fabric is then stretched harder and harder. 5 sensors are stuck onto the surface of the RP geometric models using adhesive paper as shown in Figure 8(b). The data of the force applied on these 5 sensors are then transferred to the computer and saved for later data analysis. About 300 grids with dimension 1 cm x 1 cm are drawn onto the surface of the sample fabric to measure the extension under stretching.

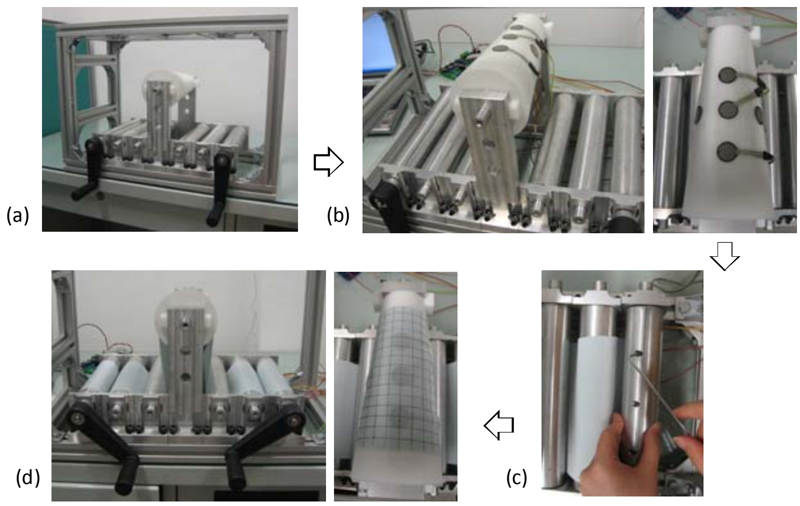


Figure 8. Testing: (a) machine with model inserted, (b) 5 sensors are inserted onto the surface of RP models, (c) sample fabric placed zigzag onto the shafts, and (d) rolling the handles for stretching fabrics.

Figure 7. Machine with camera mounted.

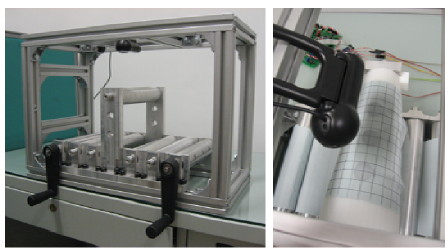


Figure 9. Machine with camera mounted.

We have developed a program for data-analysis, which allows us to use a camera mounted on top of the aluminum frame as displayed in Figure 9 to observe fabric extensions. By looking down at the surface of the shaft on the fixed holder, a 2D line can be observed which can then be calculated into a 3D one using our software with the 3D geometry of RP models input. The strain of the fabric is then resulted. Figure 10 shows the interface of our software, and Figure 11 gives some examples for the resultant stress-strain relationship on different testing fabrics.

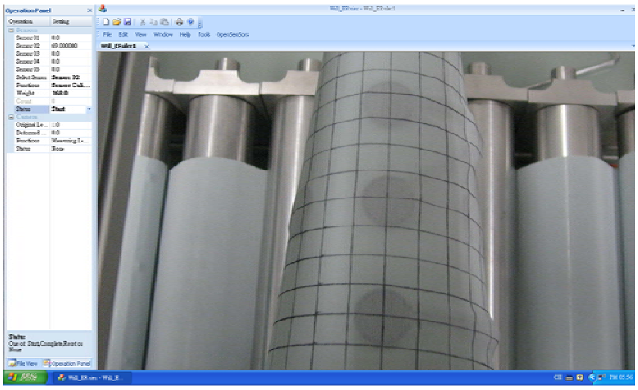


Figure 10. Software interface

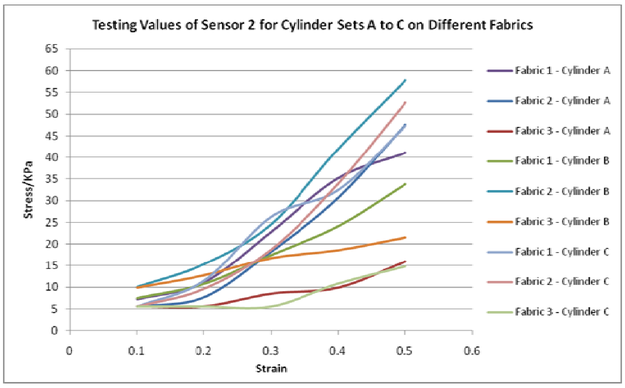


Figure 11. Example results with strain below 0.5 for cylinder using different fabrics.

Figure 12. A new setup to capture the full human body shape by cameras.

During our preliminary testing phase we tried to use the RP models with different shapes to simulate different body parts. However, it is clear that if we really want to test the strain and stress values of different fabrics acting on a human body, using RP models is not good enough. Human limbs are composed of bone, muscle, skin etc. The pressure acted on human limbs is somehow different from the one we obtained from the RP models. In view of this, we have set up a body scanning device with a number of cameras mounted on it to collect 3D data of the human body and later generate a 3D human body by some reconstruction algorithms. The set up is shown in Figure 12. It is a rather simple setup with one aluminum profile fixed over the top, and two aluminum profiles acting as the left and right legs. Several cameras are mounted onto it to capture 2D pictures of the human body from different angles. We have developed a new algorithm to assign orientation consistent normal vectors to the scattered point cloud data that can be captured from the 3D data acquisition devices. As the consistently oriented normal vectors are very important for the quality of finally reconstructed surface models, our conference paper (in IEEE International Conference on Shape Modeling and Applications – SMI 2010) about this technique was selected as one of the best papers to be published in the Computers & Graphics journal as a full-length research paper. Some results and comparisons are shown in Figure 13.

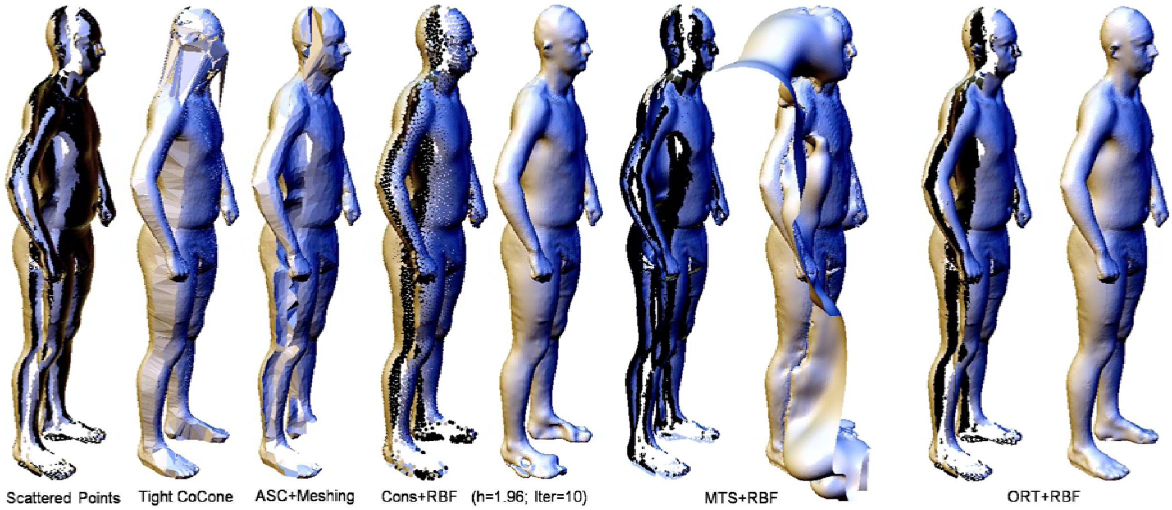


Figure 13. Examples and comparisons of our normal orientation method (ORT) to other methods on the surface reconstruction of a human body from scanned point clouds.

4. PUBLICATION AND AWARDS

[1] C.C.L. Wang, and K. Tang, “Pattern computation for compression garment by a physical/geometric approach,” *Computer-Aided Design*, Elsevier, NL, vol.42, no.2, pp.78-86, February 2010.

[2] S. Liu, and C.C.L. Wang, “Orienting unorganized points for surface reconstruction,” *Computers & Graphics*, Elsevier, NL, Special Issue of IEEE International Conference on Shape Modeling and Applications (SMI 2010), vol.34, no.3, pp.209-218, Arts et Metiers ParisTech, Aix-en-Provence, France, June 21-23, 2010.

[3] C.C.L. Wang, Y. Zhang, and H. Sheung, "From designing products to fabricating them from planar materials", *IEEE Computer Graphics and Applications*, IEEE Computer Society, vol.20, no.6, pp.74-85, November 2010.

[4] Y. Zhang, and C.C.L. Wang, "WireWarping++: Robust and flexible surface flattening with length control", *IEEE Transactions on Automation Science and Engineering*, IEEE Robotics and Automation Society, in press.

[5] S. Liu, and C.C.L. Wang, "Fast intersection-free offset surface generation from freeform models with triangular meshes", *IEEE Transactions on Automation Science and Engineering*, IEEE Robotics and Automation Society, in press.

REAL-TIME TRANSMISSION OF HIGH DEFINITION (HD) 3D VIDEO AND HD AUDIO IN GIGABIT-LAN

Principal Investigator: Professor Raymond Yeung [1]
Co-Investigator: Mr. Alan Lam [1]
Mr. Ka-Kui Choy [1]

Research Team Member: Ms. Anna Shim, Research Assistant [1]

[1] Dept. of Information Engineering

Reporting Period: 1st June 2008 – 30th April 2009

ABSTRACT

In this project, an innovative compression algorithm is proposed for handling the current challenges in HD stereoscopic 3D video transmission. The algorithm makes use of the similarity between the images on the two channels that consist of essentially the same color pixels. The main advantage of the algorithm is the unlimited threads that can be generated for different color ranges, so that it can be run very efficiently on a computer with multiple logical CPUs. Data structures are specified by parameters; hence, expansion and fine-tuning can be done in the future. The system architecture, data structures, optimization and simulation results are presented in this report.

1. OBJECTIVES AND SIGNIFICANCE

Nowadays many applications offer high-definition *stereoscopic* view, including robotic surgery and tele-immersion. The display of these 3D videos is now limited to the local display. Real-time streaming for these applications can extend the usage and to enable one-to-many monitoring and remote control. Our project, collaborating with the Prince of Wales Hospital, develops a real-time HD 3D surgical video transmission system on the *Da Vinci*® Surgical System for robotic surgery. The specific goal of this project is to achieve on the system a response time of less than 1/60s (neglecting the transmission time needed through network) so as to be able deliver a HD stereoscopic video in real time over a commercial network. The quality of the video should be high enough that there is no noticeable distortion. To our knowledge, there is no existing technology that can fulfill all of the above system requirements, including the latest technologies of H.264, JPEG3D and 3D-SPITH.

2. RESEARCH METHODOLOGY

2.1 Introduction

A new data compression algorithm for the stereoscopic images is developed as follows. With a common light source and identical hardware setups for the left and right channels, the left and right images consisting of essentially the same color pixels as illustrated in Figure-2a. In the human vision system, the left and right images are more or less a horizontal shift of each other. Accordingly, a pairs of corresponding color pixels in the left and right images are likely to have the same y-coordinate. Instead of using the traditional location index method for storing data, the color index has been used in our model. This model, referred to as the color base model, compresses data pertaining to the same color. Due to the stereoscopic nature of the problem, using the color base model can achieve efficient data compression that benefits both storage and transmission.

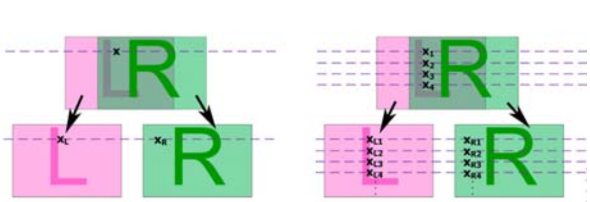


Figure-2a. General behavior on identical source to resulting input

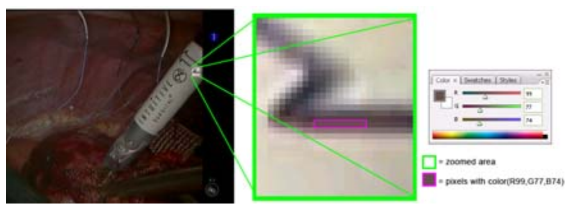


Figure-2d. Neighboring pixels with same color

The details of our compression algorithm based on the color base model are given as follows:

2.2 System Architecture

The main components of the system are Calibration, Input, Data Preparation, Flag Generation, Transmission, Decompression, and Displaying.

Calibration: To allow the best compression performance and to compensate the hardware variation (the variation of the light sensors in particular), the operator needs to spend some time to calibrate the system before the surgery starts.

Input: Two 720p 60fps videos are captured from the console of the *Da Vinci*® Surgical System which consists a synchronized pair of left and right channels. In this report, our simulation results are based on a 1400x1080@25fps training video.

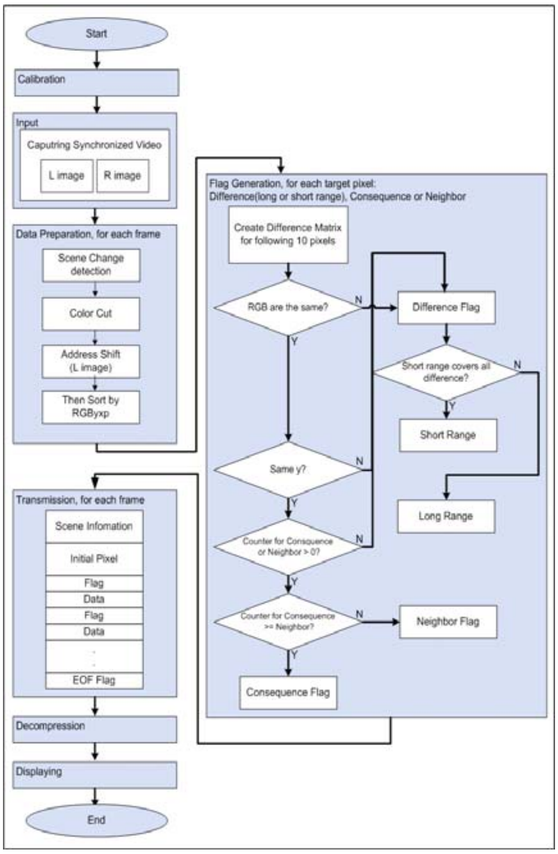


Figure-2e. Logic flow of the algorithm

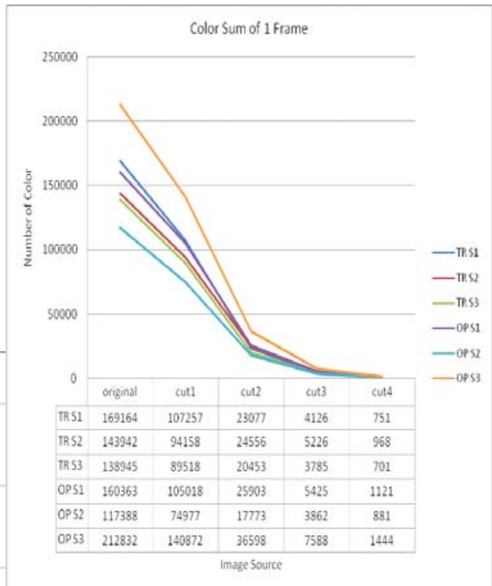
Data Preparation

<u>A. Scene Change Detection</u>

The focused object and the object distance do vary from scene to scene. Pre-settings for a number of prototypical scenes are input into the compression engine in order to reduce the time needed to re-calibrate the system upon a scene change.

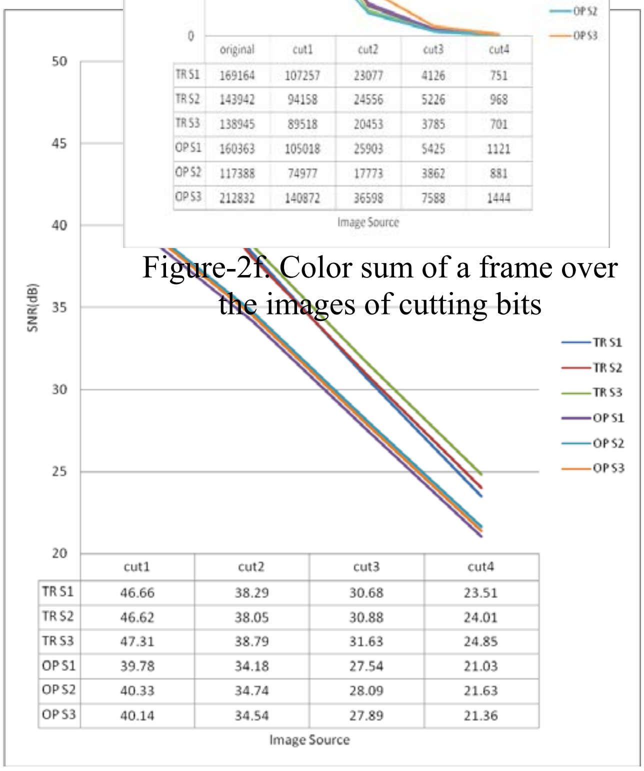
<u>B. Color Merge</u>

Due to hardware variation and degradation, it is impossible to produce the ideal left and right inputs for our system. For example, the left and right cameras are likely to suffer from color shift. In other words, the same color can produce two different colors on the left and right channels, making our color base compression algorithm inefficient. To circumvent the problem, a straightforward "bit truncation" which is found to be effective is applied. Each RGB value set specifies a unique color. By truncating the least significant bits across the left and right channels, colors with slightly different RGB values are mapped to one color, so that a minor color shift can be compensated. Figure-2f illustrates the relationship between color cut and total number of color. Here, cut1, cut2, cut3 and cut4 indicate the number of least significant bits that are cut. For example, cut2 means the 2 least significant bits in all R, G and B are cut. The images are still recognizable. However, when 3 or 4 least significant bits are cut, the image distortion becomes more obvious.



	original	cut1	cut2	cut3	cut4
TR S1	169164	107257	23077	4126	751
TR S2	143942	94158	24556	5226	968
TR S3	138945	89518	20453	3785	701
OP S1	160363	105018	25903	5425	1121
OP S2	117388	74977	17773	3862	881
OP S3	212832	140872	36598	7588	1444

Figure-2f. Color sum of a frame over the images of cutting bits



	cut1	cut2	cut3	cut4
TR S1	46.66	38.29	30.68	23.51
TR S2	46.62	38.05	30.88	24.01
TR S3	47.31	38.79	31.63	24.85
OP S1	39.78	34.18	27.54	21.03
OP S2	40.33	34.74	28.09	21.63
OP S3	40.14	34.54	27.89	21.36

Figure-2g. SNR changes among number of bits increases

From Figure-2f, the transition from original to cut1 and from cut 1 to cut 2 have shown great reduction for the total number of color. Therefore, we selectd cut2 images for our experiments. For cutting 1 to 2 bits, the image remains all basic details as original. The root mean square signal-to-noise ratio (SNR_{rms}) in dB

$$SNR_{rms} = \sqrt{\frac{\sum_{x=0}^{M-1}\sum_{y=0}^{N-1} f'(x,y)^2}{\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}\left[f'(x,y)^2 - f(x,y)^2\right]}}$$

We have also explored means other than bit cutting. According to our experimental results, merging a color with frequency below a threshold with a similar color causes serious color shift, while posterization (as in Photoshop) creates many new colors. Thus color cut becomes our final choice for compensating color shift.

C. Address Shift

The left image is shifted to the focused area to match the right image. The mapping process is done in the calibration state manually. The area outside the valid address range is mapped to the remaining part of the image. As a result, each set of yxP, where P indicates the channel of the image, can be unique.

D. Sorting

The fields to be sorted are in the R, G, B, y, x, P order. Because the shift between the left and right images are expected to be shift only in the horizontal direction, sorting y then x should have a better result than the other way around. P is sorted at the end because it consumes the least number of bits and in case P does not match with the nearby pixels, the picture array (P array) can locate all the pixels to their corresponding image. Color (RGB) and Address (yxP) are the two components that form for the *element set*. After the sorting, the neighboring pixels in a queue have the fewest different elements and also the shortest distance. The queued data will then fit to the flag generation step to be described next.

Flag Generation

The flag is the main part of the data structure which links and groups the pixel addresses. The details can be found in Table-2a. The flags and their corresponding data array can provide the necessary data for the decompression. Each Difference Flag describes the relative index of the next pixel while the Consequence and the Neighbor describe multiple pixels with an "increasing x" relationship.

Currently, Short1 was the subset of Short2 and Short3. Short2 was the subset of Short3.

Bits	Difference				Consequence	Neighbor
Bit1	0				1	
Bit2	0		1		0	1
Bit3	0: Long	1: Short1	0: Short2	1: Short3	0: without P Array, 1: with P Array	
Bit4	0: same R, 1: different R				00 to 11: the number of the same address pixels described in the flag	
Bit5	0: same G, 1: different G					
Bit6	0: same B, 1: different B				000 to 111: the number of pixel(s) following has/have the same RGBy	0: not a neighbor 1: a neighbor A neighbor is defined as pixels of same RGBy and share the nearby increasing x.
Bit7	0: same y, 1: different y					
Bit8	0: same x, 1: different x					
Bit9	0: same P, 1: different P				Followed by P Array and same address indicator array if present.	
Bit10	Followed by stating the new varied values one-by-one	0: use relative value without sign 1: use relative value with sign				
Bit11		Followed by the Difference Value using relative value to the latest transmitted pixel				
Bit12						
						Followed by P Array and same address indicator array if present.

Table-2a. Flags introduction

The process of flag generation is sequential among the colors but at the same time flag generation among the colors can be independent. The time complexity for sort row is $O(n^2)$, where n is the number of pixels to process. Through splitting the colors into ranges, the value of n for each range would decrease accordingly.

Transmission

Since flag generation and decoding can be processed in parallel for each color, a server farm of encoder, transmitter, receiver and decoder can be used to lower the coding and transmission effort. The transmission graph is shown in Figure-2h. Initially, the scene information, namely the address shift and the element set of the first pixel (RGByxP), are transmitted. Then the system continues to transmit flags and its data extension until the end of file (EOF) flag is reached.

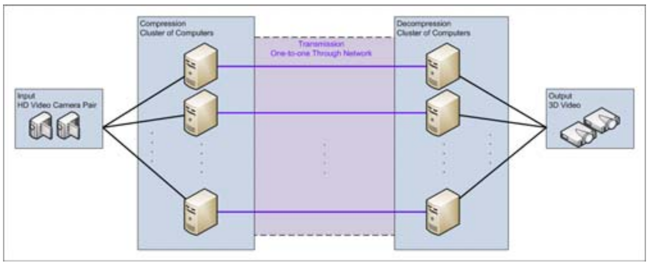


Figure-2h. Transmission Graph

Decompression

As the scene information is transmitted first, the decompressor reads the flags once they are received. Other data, such as the picture ID array for restoring the original images, then follow.

Displaying

The decoded data from the individual scenes form overlays that are ready to be displayed together. Displaying is the final output of the system. Same as for other stereoscopic HD videos, the user can watch the video by wearing polarized filter glasses.

3. RESULTS ACHIEVED

The challenge of the project is to build a system that can compress a stereoscopic HD video with very small delay while at the same time without any noticeable distortion. The resulting video stream should have data rate in the range of 500Mb/s to 2Gb/s so that it can be delivered on a commercial network. Specifically, the end-to-end processing time of the system must be less than 1/60 second in order to support real-time applications.

Our main contribution of this project is a parallel algorithm that compresses much faster than the JPEG lossless mode with no noticeable image distortion provided that the algorithm is run on a machine with a large number of logical CPUs. The speed of our algorithm is roughly proportional to the number of logical CPUs available. In the rest of this section, we describe the details of the algorithm.

3.1 Experiment

One of the assumptions for our algorithm was that the same RGB pixels in the overlapping region are to be found in both images after matching the focused regions. Thus, experiments investigating this assumption were carried out. The main focused regions of the left and right images in a frame were first matched manually and the corresponding and maximum shift of x- and y-axis were marked down. Based on the above data, the width, height and displacement of a mask were set and applied to each pixel for the right image. The exact shift and mask for some typical landscape pictures (TR S1, TR S2 and TR S3) were fed to the experiment for 3 seconds (75 frames), 1 second (25 frames) and 1 second (25 frames), respectively. The experimental results are in Table 3a and Table 3b.

Source	Original			cut2		
Scene	TR S1 000000	TR S2 002000	TR S3 004100	TR S1 000000	TR S2 002000	TR S3 004100
Successful match(%)	16.81	15.77	7.84	45.50	47.63	34.15
Scene	TR S1 000300	TR S2 002100	TR S3 004200	TR S1 000300	TR S2 002100	TR S3 004200
Successful match(%)	14.53	14.66	7.92	48.76	47.39	33.69

Table 3a. Statistics of the successful match rate for the train video

Source	Original	
Scene	OP S1	OP S2
Successful match(%)	3.23	1.93
Source	Cut2	
Scene	OP S1	OP S2
Successful match(%)	16.74	14.20

Table 3b. Statistics of the successful match rate for the operation video

It was observed that the successful matching rate in the unprocessed images is extremely low for the original image but improves significantly for cut2. The reason is that the sensors are not identical; truncating 2 bits can increase the color shift tolerance. For TR S3, the successful match rate was lower than the previous two cases. It was probably due to the fact that the objects in this scene are more vertically displaced and a vertical mask should be used. At the same time, the successful match rate for some typical surgical pictures (OP S1 and OP S2) was not significant even for cut2 due to the heavy random noise in the operation video. This will be further discussed below.

Method	our algorithm - original			our algorithm – cut2		
Scene	TR S1 000000	TR S2 002000	TR S3 004100	TR S1 000000	TR S2 002000	TR S3 004100
Compression ratio	1：1.18	1：1.20	1：1.14	1：1.56	1：1.65	1：1.57
Scene	TR S1 000300	TR S2 002100	TR S3 004200	TR S1 000300	TR S2 002100	TR S3 004200
Compression ratio	1：1.17	1：1.19	1：1.15	1：1.58	1：1.64	1：1.62

Table 3c. The simulation result of our algorithm for Train Video

Method	our algorithm - original	
Scene	OP S1	OP S2
Compression ratio(%)	1:1.02	1:1.05
Method	our algorithm - cut2	
Scene	OP S1	OP S2
Compression ratio(%)	1:1.52	1:1.48

Table 3d. The simulation result for OP Video

From Tables 3c and 3d, we see that the performances of our algorithm for S3, OP S1 and OP S2 are similar to the other scenes. This indicates that our algorithm can compress efficiently even when the focused areas of the images are not precisely matched.

The result was not similar to the masking experiment. The stability of our algorithm is probably due to the proper sorting of the pixels at the very beginning. Thus, in the subsequent flag generation, the flags designed could work well to save the number of bits needed for the description of the image.

3.2 Performance Comparison

We first compare the compression ratios achieved by our algorithm in Tables 3c and 3d (cut2) with the corresponding compression ratios achieved by the JPEG lossless mode on Matlab in Table 3e (cut2). Here we see that JPEG performs slightly better but not much better than our algorithm.

The compression delay requirement of our system as stated at the beginning of Section 3 (< 1/60 second) is so stringent that it cannot possibly be met with existing techniques such as JPEG lossless, not to say MPEG. This is clear from Table 3e.

Method	JPEG lossless mode - original					JPEG lossless mode - cut2				
Scene	TR S1	TR S2	TR S3	OP S1	OP S2	TR S1	TR S2	TR S3	OP S1	OP S2
Compression ratio	1:1.58	1:1.68	1:1.68	1:1.70	1:2.05	1:1.77	1:1.98	1:2.01	1:1.77	1:1.89
Compression time(s)	0.865	0.841	0.852	0.281	0.286	0.845	0.830	0.845	0.260	0.268

R

Table 3e. Statistics of JPEG lossless mode performance

As described in Section 2, our algorithm uses a color base model that processes pixels falling into the same color range independently. Our algorithm is programmed in a way that different threads are generated for different color ranges. On a machine with multiple logical CPUs, these threads can be run on different logical CPUs in parallel. Thus our algorithm can compress much faster than the JPEG lossless mode provided that the algorithm is run on a machine with a large number of logical CPUs. Moreover, the speed of our algorithm is roughly proportional to the number of logical CPUs available.

In addition to the potential speed that can be achieved, our algorithm is superior to many other algorithms because no special hardware is required. In our algorithm, only linear operations are involved, and so no special hardware is necessary. As a comparison, in order to run JPEG 2000 efficiently, a hardware accelerator is usually required.

4. FUTURE DEVELOPMENT

Future work for the improvement of our algorithm should focus on the following aspects:

Automation: A rapid matching algorithm is needed to replace the manual calibration step. According to our experimental results, a simple macro-block matching method is sufficient as the accuracy of matching the focused area will not affect the performance too much.

Video Quality: Normalization of left and right videos should be used to compensate the color shift.

Compression Ratio: The flag generated can be coded in a better form to reduce the bits needed.

Programming: Increasing the process parallelization ratio and developing a more efficient sorting algorithm are the keys to further improve the processing time.

5. CONCLUSION

An innovative compression algorithm over the color domain for HD stereoscopic 3D video has been designed by taking advantage of the characteristics of stereoscopic images. The images for the left and right channels share a large common part which ideally is a horizontal shift of each other. A simple bit truncation is applied to compensate the color shift between the two channels incurred by hardware variation and/or degradation. The flag generation in the algorithm is an intelligent method that connects the shared information with as few as number of bits as possible to achieve satisfactory compression.

Although the compression ratio of our algorithm falls behind that of JPEG or JPEG2000, it is not a major deficiency in light of the availability of huge bandwidth in commercial networks. Rather, the strength of our algorithm lies on the parallelization design that makes the algorithm far more superior than existing algorithms in terms of compression delay (essential for real-time applications) when it is run on a machine with a large number of logical CPUs. Such machines will be widely available before long as more and more CPUs are packed into a single processor. Besides, unlike some existing algorithms, the efficiency of our algorithm does not rely on any hardware accelerator because very simple operations are involved in the computation.

In conclusion, we have developed in this project a compression algorithm that is potentially useful for real-time transmission of HD stereoscopic 3D videos over commercial networks.

6. PUBLICATION RESULTING FROM THIS PROJECT

3D surgery videos captured in this project have been included in *eSurgical Textbook & Journal* at http://www.esurg.net/

3D surgery videos of urology captured in this project will be published in the book "泌尿肿瘤影像手术学---方法和技巧; 人民卫生出版社" in 2010.

HIGH DYNAMICC RANGE IMAGE COMPRESSION AND DISPLAY

Principal Investigator: Jiaya Jia(1)
Co-Investigator (if any):

Research Team Members: Qi Shan, Research Assistant (1)

(1) Dept of Computer Science and Engineering

Project Start Date: 1st June 2007
Completion Date: 31st May 2009

ABSTRACT

The high dynamic range (HDR) images are the new formats of the multimedia data which contain rich and broad visual information in colors and structures. Only in the recent a few years, with the development of the image and video acquisition devices, the HDR data attract much attention due to their faithful representation of the ubiquitous high contrast scenes in the real world. In this project, we plan to explore the challenging problems of efficiently and robustly rendering, compressing, and visualizing the HDR images.

1. OBJECTIVES AND SIGNIFICANCE

The research on high dynamic range (HDR) images compression has been made significant progress in recent years. However, there are still many processing and visualization problems unsolved. For instance, current LCD or CRT monitors have the global contrast no more than 1000:1. They are not capable of showing the visual details of the HDR images in which the contrast may exceed $10^5 : 1$ [1]. One example is shown in Figure 1 that the HDR images contain rich structural information. When displaying the HDR image on ordinary displays, details are lost in saturated pixels.

Figure 1: The HDR image displayed on ordinary monitor under different exposures loses details in either bright or dark regions.

Therefore, the high dynamic range should be compressed for the purpose of visualization. Another problem is related to the limitations of the image and video encoding methods, such as MPEG and JPEG, in representing the HDR format. These widely used media data compression standards mostly support 8-bit color channels. Without proper compression of the large-size HDR data, the adoption of this new format in many possible applications is hindered.

The objective of this project involves exploring the challenging problems of efficiently and robustly compressing and visualizing the HDR images. The investigation will be carried out in both the theoretical study and the empirical validation of the novel processing algorithms and visualization devices. The long-term goal of this project includes:

- To create an area of strength on the research of processing HDR data, and push the frontier of the HDR study in medical imaging and 3-D rendering.
- To build the applicable HDR data acquisition and testing environment to facilitate the algorithm benchmark and validation on hardware.
- To establish a regional platform of researching the innovative HDR applications on a variety of joint areas, such as the image enhancement.

2. RESEARCH METHODOLOGY

In this project, we plan to accomplish our objectives by conducting several tasks, specifically, in HDR data acquisition, compression, display, processing, and encoding respectively.

HDR data acquisition. The HDR images are usually taken by the high-end cameras using "raw" format or constructed by calibrating a set of low dynamic range images with fixed positions under varied exposures. However, it may not be feasible or easy for a user to setup the HDR image acquisition environment especially when the scene has very high dynamic range or contains moving objects. Thus, we will investigate the method to realistically synthesize the HDR images from ordinary low dynamic range(LDR) images [2]. The specific energy minimization problem based on the features of the LDR images can be solved to improve the visual quality of the result.

HDR image compression. Normally, the acquired HDR images cannot be easily visualized using the ordinary display or encoded using the common bmp or jpeg image format [3,4,5]. The process to reduce the high dynamic range while preserving the visual details is called range compression. In order to produce a visually satisfying compression result, on one hand, the local contrast should be maintained in a perceivable level; on the other hand, the global contrast of the HDR images should be primarily reduced. To satisfy these requirements, in this project, we shall study the window-based tone mapping method in which a global optimization problem [6] can be solved with appropriately guided local linear constraints. Specifically, our method operates on the windows containing only a few neighboring pixels. In each window, we shall use a linear function to constrain the modification of the radiance values in a monotonic order to retain the local structures. Our proposed method does not involve scale decomposition, layer separation, or image segmentation, thereby is immune from the possible artifacts such as noticeable seams between segments.

HDR data processing. Although there have been many algorithms proposed to calibrate the images, till now, there is still no many methods dedicated to processing the HDR data. It is noted that in photo-editing tools, processing the HDR images and demonstrating the desirable effect done on them is more important than just viewing them with sufficient visual details in dark and bright regions. The possible processing may include denoising, sharpening, stylization, or smoothing. So it is essential for the HDR data, before and after these operations, to not only reveal all structure details, but also faithfully illustrate the modifications performed. We thus propose to process the HDR data visually faithful to the compression.

Interactive touch-up. It is not rare that the user desires a direct image touch-up to locally modify the luminance or contrast in the processing results. We, thus, plan to provide flexible interaction tools [7] in our system using the simple brush strokes to collect pixel samples and to indicate the modifications. The process is not isolated on these samples. It fits into our global optimization framework to produce seamless image and video result.

REFERENCES

[1] Ahmet Oguz Akyuz, Roland Fleming, Bernhard E. Riecke, Erik Reinhard, and Heinrich H.

Bulthoff. “Do hdr displays support ldr content? a psychophysical evaluation” In *SIGGRAPH*, page 38, 2007.
[2] F. Banterle, P. Ledda, K. Debattista, and A. Chalmers. “Inverse tone mapping” In *GRAPHITE*, pages 349–356, 2006.
[3] K. Devlin. “A review of tone reproduction techniques” Technical Report CSTR-02-005, 2002.
[4] Durand F. and Dorsey J. “Fast bilateral filtering for the display of high-dynamic-range images” In *ACM SIGGRAPH*, pages 257–266, 2002.
[5] Raanan Fattal, Dani Lischinski, and Michael Werman. “Gradient domain high dynamic range compression” In *ACM SIGGRAPH*, pages 249–256, 2002.
[6] William W. Hager, Shu-Jen Huang, Panos M. Pardalos, and Oleg A. Prokopyev. “*Multiscale Optimization Methods and Applications (Nonconvex Optimization and Its Applications)”* Springer, 2005.
[7] Berthold K. P. Horn. “Determining lightness from an image” *Computer Graphics and Image Processing*, 3:277–299, 1974.

3. RESULTS ACHIEVED

We developed a new tone reproduction operator. We observe that HDR in an image can typically be categorized into two types – 1) regions exhibiting significant high dynamic range but with smooth radiance transition and 2) regions exhibiting sharp and significant local radiance change among neighboring pixels. Our tone reproduction operator provides a unified framework to effectively address these two types of HDR. Our approach maintains *local structures*, including sharp edges and smooth color transitions at a perceptual level while not introducing visible artifacts such as halos.

Our approach uses a window-based tone mapping method in which a *global* optimization problem is solved that satisfies *local* constraints. Specifically, our method operates on windows; in each window a linear function is used to constrain the tone reproduction in order to naturally suppress strong edges while retaining weak ones. The high dynamic range is compressed by solving an image-level optimization problem that integrates all window-based constraints.

In addition to high dynamic range compression, our method also contributes a unified framework for tone enhancement of ordinary images by solving the same optimization problem. An application of synthesizing an HDR image from a single low dynamic range (LDR) image is presented to simulate the high contrast environment in indoor and outdoor scenes containing shiny light sources or obscure shadows. The effectiveness of our HDR synthesis is evaluated by comparisons against ground truth images.

Another part of achievements is an efficient photometric stereo method to robustly estimate both surface normals and BRDF parameters. Our method does not require specularity separation in order to recover surface normals, thereby is capable of automatically and precisely reconstructing convex object surfaces with different level of roughness from a small number of images. The object surface is allowed to have complex textures and colors without influencing the computation accuracy.

In this unified framework, our method can also be applied to dealing with surfaces assembled with several different materials. The material segmentation, parameter estimation, and normal computation can be automatically achieved in our approach.

4. PUBLICATION AND AWARDS

[1] Hin Shun Chung and Jiaya Jia, “Efficient Photometric Stereo on Glossy Surfaces with Wide Specular Lobes,” IEEE Conference on Computer Vision and Pattern Recognition (***CVPR***), 2008.

[2] Qi Shan, Jiaya Jia, Michael S. Brown, “Globally Optimized Linear Windowed Tone-Mapping,”

IEEE Transactions on Visualization and Computer Graphics **(TVCG)**, accepted.

Remarks: IEEE **CVPR** is one of the most influential and competitive Computer Vision conferences with more than 1200 full submissions each year. They have reasonably low acceptance rates compared to peers. IEEE **TVCG** is the main transaction that accepts high-impact Graphics and Visualization papers.

MULTIMEDIA CONTENT DISTRIBUTION OVER HYBRID SATELLITE-TERRESTRIAL COMMUNICATION NETWORKS

Principal Investigator: Jack Y. B. Lee [1]
Co-Investigator (if any): Peter T. S. Yum [1]

Research Team Members: W. S. Wan, MPhil Graduate Student [1]
Y. Pan, PhD Graduate Student [1]
Y. Y. Lin, MPhil Graduate Student [1]
M. L. Ma, PhD Graduate Student [1]
K. H. Ngan, Engineer [1]
[1] Dept. of Information Engineering

Project Start Date: 1st June 2008
Completion Date: 31th May 2010

ABSTRACT

The Japan Aerospace Exploration Agency (JAXA) together with the Japan National Institute of Information and Communications Technology launched a new communication satellite called the Wideband Internetworking Engineering Test and Demonstration Satellite (WINDS) in Feb 2008. The WINDS satellite provides high-speed data communication for the Asia Pacific countries including Hong Kong with a goal to promote the use of satellites in such fields as Internet communications, education, medicine, disaster measures and Intelligent Transport Systems. This research proposal aims at leveraging the access to the WINDS satellite for the research and experiment of multimedia content multicasting over the WINDS satellite system. This project has two objectives. The first objective is to explore new research opportunities and to tackle the research challenges in delivering high-bandwidth multimedia contents such as audio and video to a large number of users across multiple countries and regions. The second objective is to offer a content distribution platform for JAXA WINDS members to multicast research, cultural, and educational contents to other JAXA WINDS members as well as ordinary Internet users.

1. OBJECTIVES AND SIGNIFICANCE

The Japan Aerospace Exploration Agency (JAXA) together with the Japan National Institute of Information and Communications Technology launched a new communication satellite called the Wideband Internetworking Engineering Test and Demonstration Satellite (WINDS) in Feb 2008. The WINDS satellite will provide high-speed data communication for the Asia Pacific countries including Hong Kong with a goal to promote the use of satellites in such fields as Internet communications, education, medicine, disaster measures and Intelligent Transport Systems.

Over the past several years we have been participating in the international collaboration efforts organized by JAXA and with the endorsement of JAXA we have submitted in July 2006 a formal application proposal to the Japan Ministry of Internal Affairs (MIC) to request access to and conduct experiments on the WINDS satellite. With the support from JAXA our application had been approved in May 2007.

This research project aims at leveraging the access to the WINDS satellite for the research and experiment of multimedia content multicasting over the WINDS satellite system. This project has two objectives. The first objective is to explore new research opportunities and to tackle the research challenges in delivering high-bandwidth multimedia contents such as audio and video to a large number of users across multiple countries and regions. The second objective is to offer a content distribution platform for JAXA WINDS members to multicast research, cultural, and educational contents to other JAXA WINDS members as well as ordinary Internet users.

The JAXA WINDS project is a unique and rare opportunity to conduct networking and multimedia research over a real satellite environment. The choice of using multimedia content distribution as a platform for research in this project offers two advantages. First, distributing high-bandwidth multimedia content over both wired, wireless, and satellite communication links is a unique problem that is not well studied. Therefore we expect to uncover new research problems that may open up new directions in satellite communication research. Second, the CUHK Video Multicast platform, if used or adopted by other WINDS members inside and outside Japan, will enhance the impact of this research project and potentially generate new international research collaborations in the future.

2. RESEARCH METHODOLOGY

This research tackles three related problems in the distribution of multimedia content over a hybrid terrestrial-satellite communications network, namely (a) TCP protocol performance over long-delay networks; (b) protocol performance over bandwidth-asymmetric networks; and (c) media streaming over multi-path networks.

2.1 TCP Protocol Performance

It is well-known that TCP does not perform well in modern satellite-based communications due to the channel's very large bandwidth-delay product. For example, the WINDS communications satellite launched recently by JAXA has bandwidth beginning from 24 Mbps all the way up to 155 Mbps. Even with the lowest bandwidth of 24 Mbps, a round-trip-delay (RTT) of 500 ms will lead to a bandwidth-delay product of 1.5 MB – a value far exceeding TCP's maximum window size of 64KB. In this case TCP's maximum window size of 64KB will only allow 1 Mbps of throughput – less than 5% of the channel's link-layer bandwidth.

The conventional solution to the above problem is to make use of TCP's Large Window Scale (LWS) extension as defined in RFC1323. This extension allows TCP to negotiate during connection setup a multiplier to apply to the window size so that a window size larger than 64KB can be used. However this approach relies on two assumptions. First, either the Operating System or the application must be modified to explicitly make use of TCP's LWS extension. Second, there must be a way for the application to request the use of LWS during connection setup.

While these two assumptions can be easily satisfied in the laboratory where custom network applications and operating systems can be developed to exploit TCP's LWS extension, they will likely prevent the vast amount of network applications already available in the Internet to benefit from TCP's LWS extension. We analyzed the TCP advertised window size of a number of popular network applications on Microsoft Windows platform, including web browsers (Internet Explorer, Firefox, Google Chrome), FTP clients (Windows' built-in FTP client), and Outlook Express, and are surprised to find that not only they do not make use of TCP's LWS option, their default advertised window size is merely 17KB, which will certainly further restrict the achievable throughput in satellite-based communications.

To tackle this problem we propose a novel TCP-SuperCharger (TCP-SC) to address the problem without the need to modify the network applications on both ends of the communication session. Fig. 1 below illustrates the architecture of TCP-SC.

The key component in Fig. 1 is the TCP-SC gateway – sitting between the two communicating hosts. It is implemented as a transparent gateway where all TCP packets are first received, processed, and then forwarded to the outgoing link. The TCP-SC gateway can be located anywhere as long as TCP packets can be routed through it. This can be done either via routing configuration or simply insert the gateway physically anywhere along the network path connecting the two communicating hosts.

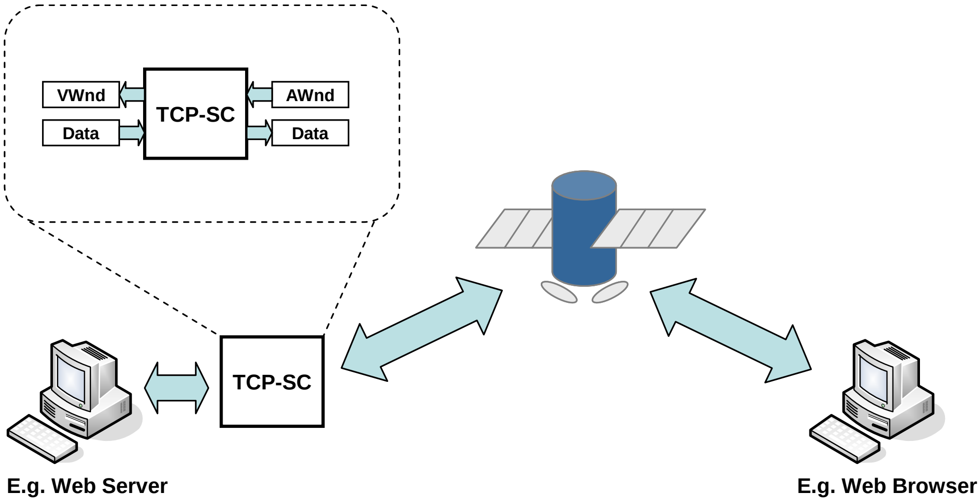


Fig. 1. A TCP connection modulated by a TCP-SC gateway.

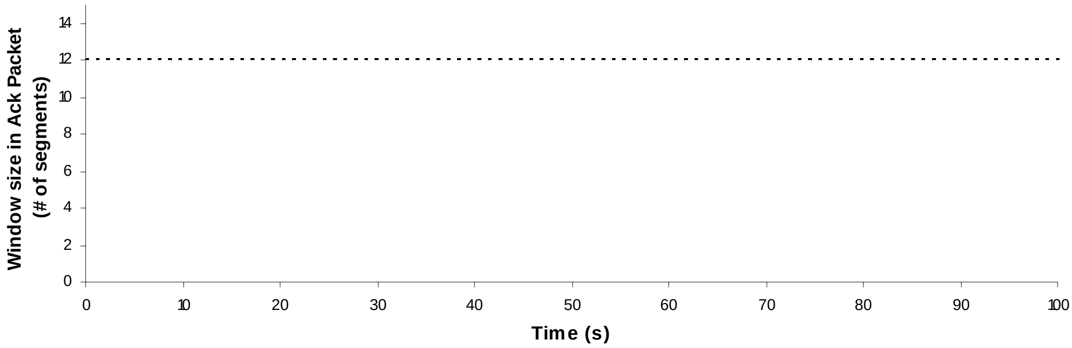


Fig. 2. Receiver's advertised window size remains at maximum throughout the experiment session.

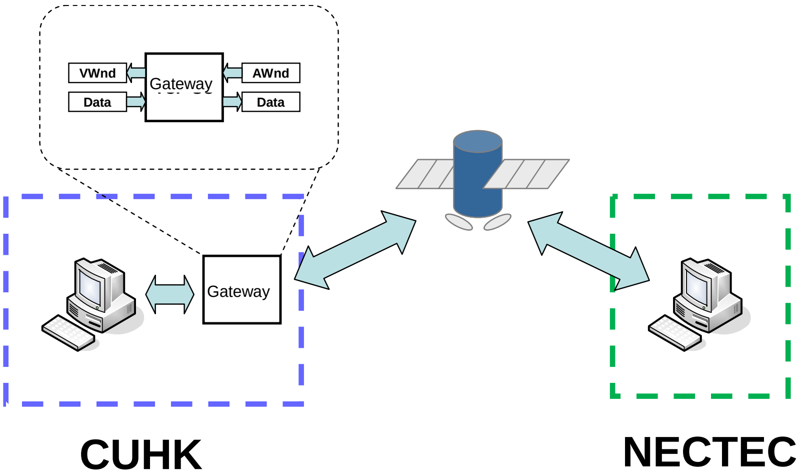


Fig. 3. Network Setup for the WINDS experiment between CUHK and NECTEC Thailand.

The principle of TCP-SC is that although the advertised windows cannot be adjusted without modification to the network application, it is possible to send more data than the amount specified by the advertised window size. For example, Fig. 2 plots the advertised window size for a browser downloading a file from a web server. It is clear that the application's advertised window size is small (12 TCP segments totaling 17KB) so the achievable throughput is far lower than is otherwise possible (only 256 Kbps on a 100 Mbps link). Now a more subtle observation is that the window size advertised by the receiver, although small, is consistently at the upper limit. In other words the receiver has no trouble processing all incoming TCP packets so that the receiving buffer is consistently near empty.

This observation opens up an entirely new interpretation of TCP's advertised window size – it is an indication of the receiver's processing capability instead of the absolute amount of buffer space available at the receiver. Consequently this implies that the sender is no longer constrained by the absolute value of the receiver's advertised window size and instead can transmit more data than the absolute value of the advertised window. In other words, this new approach enables the system to increase the window size without modifying either end of the communicating applications.

2.2 Protocol Performance over Bandwidth-Asymmetric Networks

Many existing networks exhibit bandwidth asymmetry between the uplink and the downlink. In almost all cases the downlink is allocated more bandwidth than the uplink to achieve efficient bandwidth utilization in client-server type applications. However the recent advances in and the wide-spread applications of distributed content distribution protocols such as peer-to-peer (P2P) rendered this assumption invalid. Specifically in these P2P protocols we observe that the protocol performance in the downlink can be negatively impacted by traffic condition in the uplink.

In particular if the upload data rate to other peers is too high it could severely degrade the download throughput in an asymmetric network, even if the downlink has abundant bandwidth available. Experiments revealed that the download throughput degradation is in fact not caused by congestion in the uplink, but caused by increased queuing delay in the uplink path during high upload data rates. This work tackled this problem by developing an adaptive algorithm to monitor the uplink queuing delay and adjust the upload data rate limit dynamically so that the download throughput will not be adversely affected. Experiments conducted using an open-source P2P software showed that the proposed algorithms can increase the downlink utilization over a wide range of network configurations (and over 200% increase in some cases) by automatically adjusting the upload data rate limit. The algorithms do not require any user intervention and can be readily incorporated into existing P2P systems.

2.3 Media Streaming over Multipath Networks

This work tackled the challenge of streaming high-bit-rate media over a complex networks with multiple sources, multiple paths, and even multiple overlays. We investigated two approaches: (a) a multi-overlay adaptive streaming protocol based on implicit achieveable bandwidth measurements; and (b) a joint-coding-scheduling approach to streaming media from multiple sources.

In the first approach we developed a new in-band bandwidth probing tool which can estimate achievable bandwidth, i.e., the data throughput that can be realized between two peers over the transport protocol employed (e.g., TCP). Unlike RTT this new tool can determine the amount of extra bandwidth available in the target network path so that excess data traffic can be diverted from congested path without causing new congestion in the target path. Moreover, the probing tool does not incur any bandwidth overhead as it piggybacks on the existing data flow. Our simulation results show that multi-overlay ALM networks constructed based on achievable bandwidth consistently out-performs RTT-based approach in terms of data delivery ratio and video playback continuity, and has comparable performance in terms of data delivery latency at high data rates. Moreover, the achievable bandwidth metric can prevent overlay topology oscillations and improve fairness among peers. The proposed bandwidth probing tool can be implemented entirely within the application and thus can be readily incorporated into existing ALM protocols.

In the second approach we investigated the use of erasure codes to encode the media data and then schedule multiple peers to stream the encoded data simultaneously to a receiver. By exploiting the order-invariant property of erasure codes this approach enables the sending peers to fully utilize their available bandwidth resources and yet does not need to estimate or predict their bandwidth availability. Moreover, we develop distributed scheduling algorithms to juxtapose the data transmissions from multiple peers so that the coding and storage complexities can be kept at practical level in scaling up the system.

3. RESULTS ACHIEVED

3.1 TCP Protocol Performance

We have successfully developed the proposed TCP-SC gateway and conducted multiple experiments using JAXA's WINDS satellite, in cooperation with NECTECH Thailand. The results were presented at the 27th International Symposium on Space Technology and Science held in Tokyo, Japan from 5-12 July 2009 [4]. We were also invited to submit the paper for possible publication in the Transactions of JSASS, Space Technology of Japan.

In addition we have filed a US Patent for the developed technology and have begun to investigate extending it to mobile data networks.

3.2 Protocol Performance over Bandwidth-Asymmetric Networks

We have successfully developed an adaptive algorithm to monitor the uplink queuing delay and adjust the upload data rate limit dynamically so that the download throughput will not be adversely affected. Experiments conducted using an open-source P2P software showed that the proposed algorithms can increase the downlink utilization over a wide range of network configurations (and over 200% increase in some cases) by automatically adjusting the upload data rate limit. The results were reported in [3].

In addition we have filed a US Patent for the developed technology.

3.3 Media Streaming over Multipath Networks

We have successfully developed solutions using the two approaches, i.e., a path selection algorithm and a joint-coding-scheduling protocol for distributed media streaming. The results were reported in [3] and [4] respectively.

We have also filed a US Patent for the technology in [3]. The research of [4] opens up a new way to enable the use of coding in distributed media streaming and we are now extending the developed techniques into network coding.

4. PUBLICATION AND AWARDS

[1] Jack Y. B. Lee, Peter T. S. Yum, W. S. Wan, "TCP-SuperCharger: A New Approach to High-Throughput Satellite Data Transfer," *Proc. 27th International Symposium on Space Technology and Science*, Epochal Tsukuba, Tokyo, Japan, July 5 to 12, 2009.

[2] Yuan Pan and Jack Y. B. Lee, "Adaptive Scheduling of Data Transfer in P2P Applications over Asymmetric Networks," *Proc. IEEE ICC 2010*, Cape Town, South Africa, 23-27 May 2010.

[3] Y. Y. Lin and Jack Y. B. Lee, "Path Selection in Streaming Video Over Multi-Overlay Application Layer Multicast," *IEEE Transactions on Circuits and Systems for Video Technology*, vol.20(7), July 2010, pp.1018-1031.

[4] M. L. Ma and Jack Y. B. Lee, "Multi-Source Scheduling in Streaming Erasure-Coded Video over P2P Networks," *Proc. 10th IEEE International Conference on Peer-to-Peer Computing*, Delft, Netherlands, 25-27 August, 2010.

AUTOMATIC VIDEO SEGMENTATION AND TRACKING FOR REAL TIME MULTIMEDIA SERVICES

Principal Investigator: Prof. King Ngi Ngan
Co-Investigator: Dr. Hongliang Li

Research Team Members: Liu Qiang, Research Assistant

Department: Electronic Engineering

Project Start Date: 1 June 2006
Completion Date: 30 May 2008

ABSTRACT

As an important technology, automatic video segmentation and tracking have great potential in a large number of application areas, including video monitoring and surveillance, video summarization and retrieval, video conferencing and videotelephony, computer vision and digital entertainment, etc. The purpose of this project is to develop techniques that are efficient and accurate in extracting interesting objects from videos in an unsupervised manner. Three key techniques, namely saliency model for video segmentation, saliency model for video tracking and boundary refinement technique, will be implemented. These include the use of perceptual features that will be modeled by several low-level and high-level cues for the object of interest. The outcomes of the project find applications in real time content-based multimedia services.

1. OBEJCTIVES AND SIGNIFICANCE

In recent years, there has been rapidly growing interest in content-based functionalities of video data, such as video editing, content-based image retrieval, video indexing, video event analysis, etc. Video segmentation as a challenging and active research area is a key technique for semantic object extraction and plays an important role in digital video processing and computer vision. The task of segmenting/tracking a video object emerges in many applications, such as traffic monitoring, surveillance and video conferencing, etc. Typically, the applications of video segmentation can be classified as follows:

- Video surveillance, where the segmentation result is used to allow the identification of an intruder or of an anomalous situation.
- Content-based video summarization, such as sports event summary, video skimming, video pattern mining, which requires the segmented semantic objects to perform the content classification.
- Content-based coding application.
- Computer vision, such as video matting, video toning, and rendering.
- Videotelephony and videoconferencing, where the segmentation can achieve better coding quality for the most relevant objects or to be able to store a specific object in a database, e.g., a face in a videotelephony directory.
- Digital entertainment, where some specific objects can be replaced by segmentation, such as the video games.

The objectives of the project are as follows:

- To investigate the saliency model for extracting objects of interest from videos.
- To develop novel automatic video object segmentation techniques for generic and specific (e.g., video surveillance, videoconferencing) applications based on saliency model.

- To develop real time object segmentation techniques based on saliency model and fast transform.

The successful completion of this project will provide new video segmentation tools for many applications in multimedia services such as videotelephony, videoconferencing, computer games and digital entertainment.

2. RESEARCH METHODOLOGY

Generally, object segmentation can be divided into two stages, i.e., desired object detection and object extraction, which are concerned with the pattern recognition and clustering techniques, respectively. According to the detection mode, object segmentation can be performed in two manners, i.e., supervised or unsupervised. The emphasis of this proposed research is to develop techniques that are efficient and accurate in extracting interesting objects from video sequences in an unsupervised manner. The major contribution of this project is on the use of perceptual saliency map that will be modeled by several low-level and high-level cues for the object of interest (OOI). These cues include color, orientation, intensity, location, and motion, etc. Figure 1 shows the framework of our implementation approach.

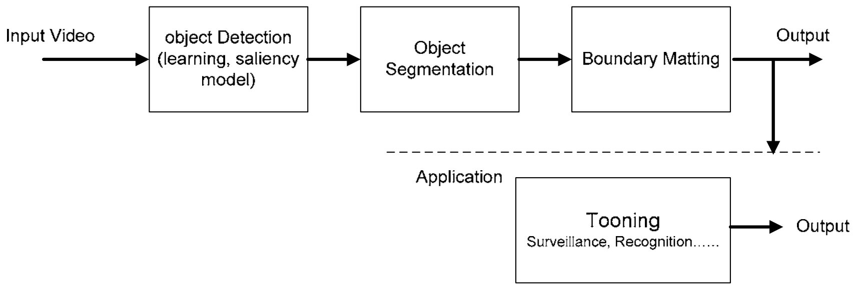


Figure 1. Framework of the developed object segmentation system.

For example, for the attention-based object segmentation method, we first construct the saliency map using the inherent features of the OOI. The general expression can be modeled by

$$S = f(oc, oi, op, os) \quad (1)$$

where S denotes the value of saliency map, and, oc, oi, op and os correspond to the color, intensity, position and structure of the object of interest, respectively.

Then, we employ non-linear filtering to eliminate the noise in the obtained saliency map, and use the classification approach to extract the object regions. Certainly, in order to refine the boundary of the objects, several optimization methods may be used, such as maximum a posteriori probability.

3. RESULTS ACHIEVED

3.1. Attention-based video segmentation: segmentation based on facial saliency map

We first generate the saliency map from the input head-and-shoulder type video image by our proposed facial attention model. Then, a geometric model and an eye-map built from the chrominance components are employed to localize the face region according to the saliency map. The final step involves the adaptive boundary correction and the final face contour extraction. There are three conspicuity maps corresponding to the chrominance, luminance, and position information used to construct the facial saliency map, which can be used to locate the potential face areas. Experimental evaluation on test sequences shows that the proposed method is capable of segmenting the face area effectively, which is evident in Figure 2.

Original Image Facial Saliency Map
Segmented Result

Figure 2. Segmentation based on facial saliency model.

3.2. Attention-based video segmentation: segmentation of defocused video image

In this work, an unsupervised segmentation algorithm based on matting model is proposed to extract the focused objects in the low depth of field video images. Based on the visual attention idea, we have constructed the focused saliency model to measure the focused region. Nonlinear filtering and matting approaches are employed to segment the object of interest accurately. An example can be found in Figure 3.

Figure 3. Segmentation result for defocused video image.
Left: The original image. Right: Segmentation result based on the proposed saliency model.

3.3. Object detection

3.3.1. Feature extraction employing fast and efficient method for block edge classification

In this work, we investigate difference properties from three coefficients in the non-normalized Haar transform (NHT) domain and present a fast and efficient method to classify block edge using these properties. The proposed method significantly reduces the number of computational operations in the edge models determination with no multiplications and less addition operations.

3.3.2. Object detection based on extracted features

In order to extract objects of interest (human) from videos, we first perform the object detection to obtain the location of the object. Based on the extracted features, we will use NHT coefficients and block edge information to detect the object by using cascade classifier. First, we manually segment the over 12,000 face samples from a large number of face databases. Then, a cascade of face classifiers is trained to detect the frontal face. Here, we use AdaBoost learning approach to perform the training process, which can capture the face saliency features for future face detection. Currently, we have finished six layer training work. For each layer, over 50% non-face samples will be rejected, while falsely eliminating only 0.1~0.3% face samples. The proposed face detector scans the image at many scales, looking for face locations within the scaled windows. It consists of three parts, i.e., skin color filtering, rejector cascade, and cascades of boosted face classifier, which is shown in Figure 4.

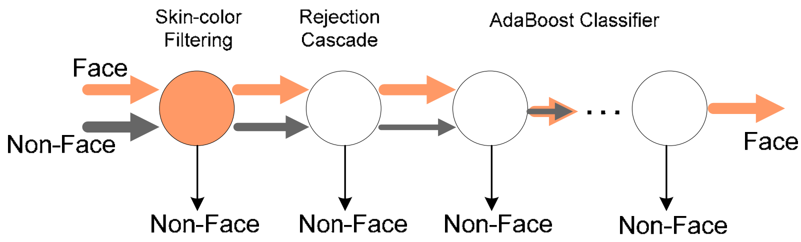


Figure 4. The flowchart of the human face detection procedure.

3.4. Face segmentation based on graph-cut

After the face detection, we can access the face locations in the input image rapidly. In this work, we use the min-cut optimization method to perform our face segmentation algorithm in the obtained face regions.

Generally, when the foreground or background is clearly defined by the user inputs, the segmentation process can be carried out based on graph cut optimization. Unfortunately, in the unsupervised manner, we only have the coarse object locations that are obtained from the boosting object detector. The corresponding sub-window may cover the complete object contour or only some parts of face regions. It means that the pixel outside the window may belong to the background, while the pixel inside the window is likely to be part of the face. We cannot determine with confidence which pixel should be marked as the background or foreground.

The proposed method works at two levels, i.e., coarse and fine scales. The initial segmentation is performed at the coarse level. There are some regions used for estimating foreground information, whilst four regions for the background. We take their means and variances as the initial clusters for the background, which means there are four components to model the GMM in the background. For each pixel in this region, we compute the weighted distances to the face and background clusters according to cost function and the similarity with other. Finally, we use the minimum cut to perform the global optimization.

The second level is the finer segmentation, which aims to refine the initial segmentation result. The corresponding foreground regions are defined as the set of the pixels belonging to the body terminal in the current window, while the background regions consist of those pixels outside the window that are classified to the background terminal. We use 8 and 8 components to describe the foreground and background colors, respectively. The mean and covariance of component k are estimated based on the K-means algorithm. Then, the similar method as the coarse scale can be used to compute the energy $E1$ and $E2$. Note that in the data cost function $E1$, the weight of each component is estimated based on the spatial samples within a defined window of 20×20 pixels that is centered with the current node. The min-cut is also used for the final optimization. Some examples based on this method are shown in Figure 5.

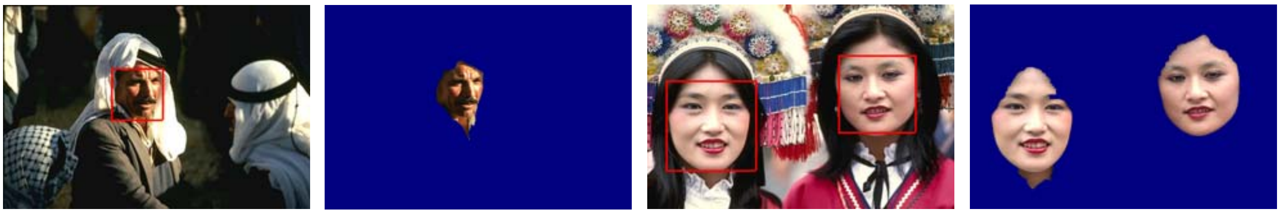


Figure 5. Graph cut based segmentation result for human faces.

3.5. Real-time human body segmentation and matting

Human body segmentation is generally considered a crucial step for human recognition, behavior analysis, or human-to-machine communication. For example, the extracted human object can be used to allow the identification of a suspicious behavior, and helps to detect their actions and alert the security center to the possible danger in time.

In this work, we aim to develop an automatic algorithm to segment human body from real-time video, which can be applied into videoconfereing, and video coding. Based on the framework given in Figure 1, we first employ the proposed face detection technique to locate the human faces in the input video. The coarse-to-fine object segmentation is then performed based on the pre-defined body region. In order to achieve the real-time segmentation, the tracking based segmentation method will be applied for the subsequent frames. As shown in Figure 6, the detection and body segmentation results are shown in the first and second frames, respectively. The following two frames illustrate the matting results after the body segmentation. It is means that a new background is used to replace the original one to compose a new scene. This technique is very important for future video communications, such as video chatting.

Figure 6. Segmentation result for human body.

4. PUBLICATIONS

[1] H. Li and K.N. Ngan, “Saliency Model based Face Segmentation and Tracking in Head-and-Shoulder Video Sequences,” *Journal of Visual Communications and Image Representation*, Europe, vol. 19, no. 5, July 2008, pp. 320-333.

[2] H. Li, K.N. Ngan and Z. Wei, “Fast and Efficient Method for Block Edge Classification and Its Application in H.264/AVC Video Coding,” *IEEE Transactions on Circuits and Systems for Video Technology*, U.S.A., vol. 18, no. 6, June 2008, pp. 756-768.

[3] H. Li and K.N. Ngan, “Unsupervised video segmentation with low depth of field,” *IEEE Transactions on Circuits and Systems for Video Technology*, U.S.A., vol. 17, no. 12, December 2007, pp. 1742-1751.

[4] H. Li and K.N. Ngan, “Automatic Video Segmentation and Tracking for Content-Based Applications”, *IEEE Communications Magazine*, U.S.A., vol. 45, no. 1, January 2007, pp. 27-33.

[5] H. Li and K.N. Ngan, “Unsupervised segmentation of defocused video based on matting model”, *IEEE International Conference on Image Processing*, Atlanta, U.S.A., October 2006, pp. 1825-1828.

[6] Q. Liu, C. Cai, K.N. Ngan and H. Li, "Camshift Based Real-Time Multiple Faces Match Tracking", *IEEE International Symposium on Intelligent Signal Processing and Communication Systems*, Xiamen, China, 28 November-1 December, 2007, pp. 726-729.

INFORMATION RETRIEVAL FROM MIXED-LANGUAGE SPOKEN DOCUMENTS

Principal Investigator: Tan Lee, Associate Professor (1)
Co-Investigator: P.C. Ching, Chair Professor (1)

Research Team Members: Wentao Gu, Research Associate (1)
Houwei Cao, PhD candidate (1)
Yu Ting Yeung, MPhil candidate (1)

(1) Department of Electronic Engineering

Project Start Date: 1st June 2006
Completion Date: 31st August 2008

ABSTRACT

This research aims to develop the key technologies for information retrieval from mixed-language spoken documents. With low-cost storage devices and rapid internet connection, we now have easy access to a huge amount of audio and video information. Effective indexing and retrieval techniques are needed to make this information useful. Existing spoken documents retrieval systems assume that the recording involves only one language with known identity. They are not able to handle mixed-language speech, which is very common nowadays. Hong Kong is an international city where many people are Cantonese and English bilinguals. English words are frequently embedded into spoken Cantonese. This presents great challenges to automatic speech transcription and indexing. This research addresses the problems of acoustic-phonetic variation, pronunciation and lexical variation in mixed-language speech. A speaker-independent speech recognition system is developed for automatic transcription of mixed-language recordings from radio programmes, lectures and meetings.

1. OBJECTIVES AND SIGNIFICANCE

With the advancement of digital electronics and computer technologies, we now have easy access to an “unmanageable” amount of multimedia information. Since speech is the most natural and convenient way of information exchange between people, audio recordings of human speech are among the most commonly available resources. Examples include radio and TV programmes, lectures, meetings, and interviews. Recently, online video/audio sharing, e.g., YouTube, has become increasingly popular. This leads to an explosive growth of public accessible audio information with virtually unrestricted content. Efficient indexing and retrieval techniques are needed to make the information useful to our daily life and work. Hong Kong is an international city where most people, especially the young generation, are Cantonese and English bilinguals. English words are frequently embedded into spoken Cantonese. For information retrieval from mixed-language spoken documents, conventional approaches based on monolingual speech recognition are not applicable, simply because there is no prior knowledge about when language switching would occur.

This research aims to develop the key technologies for information retrieval from mixed-language spoken documents. We analyse the variations of code-mixing speech at acoustic (signal), phonological, lexical and grammatical levels, and try to model them in a statistical way. A speaker-independent continuous speech recognition system is developed for automatic transcription of Cantonese-English code-mixing speech. The system can handle a large vocabulary of over ten thousand words, which cover the entire Cantonese lexicon and the most common code-switching English words. The speech transcription generated by the system contains multiple hypotheses to facilitate highly flexible and admissible indexing. Our research results will not only be useful in practical applications of audio information retrieval but also contribute to the broad research areas of natural language processing and multilingual speech communications.

2. RESEARCH METHODOLOGY

2.1. Understanding the Problem

We are interested in Cantonese-English code-mixing speech being spoken in Hong Kong. Cantonese is the primary language or matrix language and English is the secondary or embedded language. In a typical code-mixing sentence, a Chinese word is substituted by its English equivalent if exists. The grammatical structure is largely based on Cantonese. Since the speaker is usually a native speaker of Cantonese, the embedded English word carries strong Cantonese accent. The Cantonese accent may be realized in different ways. For example, the syllable structure of an English word may be modified to become an acceptable form in Cantonese. Phonemes that are unique to English may be replaced by Cantonese phones that sound similar. The stress pattern of an English word may be realized differently in code-mixing speech. Code-mixing happens mostly in casual conversational speech, in which speakers do not follow the pronunciations as specified in a standard dictionary. In terms of speech content, Cantonese is significantly different from standard Chinese. A large portion of the lexical items in spoken Cantonese are not found in standard Chinese or Mandarin, though the change of grammatical structure is not substantial.

2.2. Data Collection

2.2.1. Speech data

A large amount of speech data was collected to facilitate pronunciation variation analysis and acoustic modeling. The content of speech includes pre-selected code-mixing sentences as well as spontaneous speech with unpredictable content. Manual selection of code-mixing sentences was done based on the content of local newspapers and online resources, and the findings of relevant linguistic studies. The embedded English words are those with high frequency of occurrences in code-mixing speech and they cover different word categories and word lengths. Spontaneous speech data were collected in two different speaking scenarios: free talk and classroom teaching. In the case of free talk, each speaker was asked to freely speak about a given topic. For classroom teaching, recording of tutors' voice was conducted during small-group tutorials at the university. The recorded speech is very typical code-mixing speech.

2.2.2. Text data

Lexicon and language model are the key components of a large vocabulary continuous speech recognition system. Lexicon is often derived from dictionaries. Statistical language models are trained with large amounts of text data, which are usually collected from electronic versions of published materials, e.g., newspaper texts. There are assumptions that the lexicon contains all possible words in the language(s) and the training text data are reliable representations of the spoken language(s), i.e., they coincide with the true transcription of speech. These assumptions are not trivial. While most published text materials are able to represent formal speech, they can be quite different from conversational speech. This is especially true for Cantonese. Cantonese speech, when being written down, shows substantial differences from standard Chinese. Text of standard Chinese is not adequate to reflect the linguistic properties of Cantonese. This problem becomes even more severe for code-mixing speech. Code-mixing text is considered unofficial and found mainly in personal communication and discourses that involve contemporary and cross-cultural issues, e.g., computer, business, fashion, food, entertainment and showbiz. Two different approaches were adopted to increase the amount of Cantonese and code-mixing text data for statistical language modeling:

Translating standard Chinese text into written Cantonese. A rule-based method of translation has been devised and implemented. The translation rules can be learnt automatically from a small amount of parallel data.

Collecting data from Internet. An automatic text retrieving and filtering system has been developed for online data collection. The system is used to search for useful websites where substantial amount of colloquial Cantonese or code-mixing text can be found. To differentiate a Cantonese sentence from a standard Chinese one, we use a set of word identifiers and anti-identifiers, which are unique to Cantonese and standard Chinese respectively.

2.3 Analysis of Code-Mixing Speech

2.3.1. Pronunciation variation

A comprehensive analysis of pronunciation variation in code-mixing speech was carried out using an automatic phoneme recognition system. Three speech databases, namely, TIMIT – native English, CUSENT – read-style Cantonese, and CUMIX – Cantonese-English code-mixing speech, were used. Monolingual English and Cantonese acoustic models were trained with TIMIT and CUSENT respectively. These models were used to recognize utterances from matched and mismatched sources. Pronunciation variants caused by code-mixing were identified by comparing the phoneme confusion patterns of the mismatched condition with those of the matched condition.

2.3.2. Speech prosody

To investigate the variation of prosodic features caused specifically by code-mixing, we designed a controlled experiment to compare the prosody of a given stretch of speech occurring both in monolingual and in code-mixed utterances. We investigated five typical lexical stress patterns of English words. These English words (or their Cantonese equivalents) were carried by monolingual English, code-mixing and monolingual Cantonese sentences on different sentential positions. The sentences were spoken naturally by native Cantonese speakers. By analyzing the F0 variation of different parts of the sentences, the effect of code-mixing can be revealed.

2.4 System Design and Implementation

The major components of the code-mixing speech recognition system are acoustic models, pronunciation lexicon, and language models. The acoustic models are a set of hidden Markov models (HMMs) that characterize the statistical variation of the input acoustic observations. Each HMM represents a phonemic unit of English or Cantonese. The pronunciation lexicon and language models are used to define and constrain the way in which the phonemes can be concatenated to form words and sentences. The acoustic models are trained with the code-mixing speech databases as well as other available monolingual speech databases. The language models are trained by a large amount of Cantonese and code-mixing text data. The code-mixing pronunciation lexicon is constructed by combining a monolingual Cantonese lexicon with a set of commonly used code-mixing English words. The pronunciations of these English words are derived based on the findings of pronunciation analysis.

The recognition process is divided into two passes. In the first pass, the search space is significantly reduced from an all-included grammar network to a syllable/word graph. The basic elements of the graph are nodes and arcs. Each arc represents a hypothesized Cantonese syllable or a hypothesized English word. In the second pass, the syllable graph is further converted into a character graph from which the most likely code-mixing sentence is decoded.

3. RESULTS ACHIEVED

3.1 Code-Mixing Speech Databases

CUMIX, the first Cantonese-English code-mixing speech database, was developed. CUMIX contains not only code-mixing utterances but also monolingual English words, which are useful for the study of accented English. A total of 19000 utterances were recorded from 40 male and 40 female native Cantonese speakers. Most of the code-mixing utterances contain one English segment, which is either a single word or a sequence of words that occurs frequently in daily speech, e.g., “around”, “concern”, “CD”. All utterances were manually transcribed. Spontaneous speech data with unpredictable degree of code-mixing was collected from classroom teaching. About 20 tutorial sessions on various electronic engineering subjects were recorded. Each session contains about 30-minute speech data. The recordings were segmented into complete sentences and the time boundaries of code-mixing English words were marked manually.

3.2 Text Databases

3.2.1. Rule-based translation of standard Chinese text

A parallel corpus of standard Chinese and written Cantonese was developed by manual translation. It started with 8,300 complete sentences (90,000 characters) of standard Chinese (newspaper text). A few university students were asked to make necessary modification on these sentences so that they can be read out as natural Cantonese, with their meanings unchanged. Word segmentation was performed automatically for the parallel sentences. From word-level alignment of the parallel sentences, conversion rules are established by transformation-based learning. As a result, about 640 useful rules were obtained. They were applied to a text database of standard Chinese with 98 million characters. The translated text can be used as a database of formal Cantonese, which has a similar style to broadcast news. About 7.8% of the translated text are Cantonese terms, in comparison to 0.1% in the original standard Chinese data.

3.2.2. Online text data collection

A code-mixing text database with 30 million Chinese characters was completed. Part of the data was collected by extensive manual search of relevant online sources. The remaining part was collected in a fully automatic way using the newly developed text filtering system. The automatically collected text data are found to have fairly high quality, i.e., they are the transcriptions of typical code-mixing speech being spoken in Hong Kong.

3.3 Pronunciation Dictionary for Cantonese-Accented English

The results of pronunciation analysis confirm that English words spoken by Cantonese speakers indeed carry strong accent. Ten major pronunciation variations were identified, e.g., “tutor” → “tuta”, “bit” → “beat”, “three” → “free”. It was also observed that Cantonese speakers delete consonant syllable codas in most cases. Comparing casual with read-style Cantonese speech, the same patterns of pronunciation variation are observed. However, the degree of variation is much more severe in code-mixing speech than in read speech. These findings are used to generate the pronunciations of English words in the bilingual lexicon for code-mixing speech recognition.

3.4 Acoustic Modeling

In the context of Cantonese-English code-mixing, a fundamental question to be answered is whether we should have two sets of language-dependent phoneme models or a cross-lingual phoneme set that covers both Cantonese and English. Based on the pronunciation variation analysis, cross-lingual modeling approach is considered more appropriate. Although Cantonese and English are phonetically and phonologically very different, they share some common phonemes. These phonemes can share the same acoustic models in the cross-lingual modeling, so that better utilization of the training data can be achieved. We design the cross-lingual phoneme set by including all Cantonese phonemes and those English phonemes that have no Cantonese equivalents. The total number of phonemes is 70. The acoustic models are context-dependent hidden Markov models. Each model consists of three emitting states and each state is represented by a mixture of 32 Gaussian components. The training data are from CUSENT and CUMIX. The word-level recognition accuracy is about 60% for the code-mixing test utterances in CUMIX.

3.5 Language Modeling

The major problem in language modeling for code-mixing speech recognition is due to the limited amount of data. The English words in our text corpus generally have limited number of occurrences, which is not sufficient for probability estimation. To address this problem, we adopt a class-based approach. The English words are divided into 15 classes according to their parts of speech (POS) or meanings. It is noted that many of the words are nouns. Thus finer classification of nouns is implemented. The language models are evaluated in the phonetic-to-text (PTT) conversion task. Assuming that the true phoneme transcription is known, the language models are used to determine the most likely character/word sequence that matches the transcription, like in an LVCSR system. Using a held-out set of utterances from CUMIX, a conversion accuracy of 91.5% has been attained.

3.6 Speech Recognition Performance

A preliminary system for code-mixing speech recognition was developed based on cross-lingual acoustic models, bilingual pronunciation dictionary and class-based language models. It is assumed that the input utterance is either code-mixing speech with exactly one English segment, or monolingual Cantonese speech. The character accuracy for code-mixing utterances in CUMIX is 56.04%. It is found that the speaking style of Cantonese affects the recognition performance greatly. Formal Cantonese like in broadcast news is very different from casual conversation. With the language models trained from formal Cantonese (obtained by translation), the recognition accuracies for broadcast news speech and casual speech are 76% and 60% respectively.

3.7 Prosodic Properties of Code-Mixing Speech

The results of prosody analysis clearly show that the $F0$ level of a code-mixing English word tends to be lifted. There is a tendency that the prosody of the English speech is assimilated to that of Cantonese. The rhythmic pattern of English is shifted towards syllable timing, while the variations of $F0$ pattern are mainly on the word-final syllable. For a stressed syllable, the $F0$ contour turns flat. For a post-tonic unstressed syllable, the $F0$ contour falls steeply. Such $F0$ variations are the result of the phonological interaction of English lexical stress and Cantonese lexical tones. In contrast, the prosody of Cantonese is not much affected.

4. RESEARCH TEAM

This research was done at the DSP and Speech Technology Laboratory, Department of Electronic Engineering. The research team consists of the following members:

Tan Lee	Principal Investigator
P.C. Ching	Co-Investigator
Wentao Gu	Research Associate
Houwei Cao	PhD student
Yu Ting Yeung	MPhil student

5. PUBLICATIONS

[1] Joyce Y. C. Chan, P. C. Ching, Tan Lee and Houwei Cao, "Automatic speech recognition of Cantonese-English code-mixing utterances," in *Proc. INTERSPEECH 2006*, pp.113-116, Pittsburgh, USA, September 2006.

[2] Houwei Cao, P.C. Ching, Tan Lee and Ning Wang, "An Extended Cantonese-English Code-mixing Speech Corpus: exCUMIX," in *Proc. Oriental COCOSDA Workshop*, pp.1-5, Penang, Malaysia, December 2006.

[3] Wentao Gu and Tan Lee, "Effects of tonal context and focus on Cantonese F0," In *Proc. International Congress of Phonetic Sciences,* pp.1033-1036, Saarbrücken, Germany, August 2007.

[4] Wentao Gu and Tan Lee, "Effects of focus on prosody of Cantonese speech - a comparison of surface feature analysis and model-based analysis," in *Proc. ParaLing'07*, pp.59-64, Saarbrücken, Germany, August 2007.

[5] Houwei Cao, P.C. Ching and Tan Lee, "Pronunciation variation analysis for Cantonese-English code-mixing speech," in *Proc. Oriental COCOSDA Workshop*, pp.143-148, Hanoi, Vietnam, December 2007.
[6] Yu Ting Yeung, Houwei Cao, N.H. Zheng, Tan Lee and P. C. Ching, "Language modeling for speech recognition of spoken Cantonese," in *Proc. INTERSPEECH 2008*, pp.1570-1573, Brisbane, Australia, September 2008.
[7] Jiang Cao, Xiaojun Wu, Yu Ting Yeung, Tan Lee, Thomas Fang Zheng, "Automatic collecting of text data for Cantonese language modeling," to appear in *Proc. Oriental COCOSDA Workshop*, Kyoto, Japan, November 2008.
[8] Wentao Gu, Tan Lee and P.C. Ching, "Prosodic variation in Cantonese-English code-mixed speech," to appear in *Proc. ISCSLP 2008*, Kunming, China, December 2008.
[9] Nengheng Zheng, Houwei Cao, Tan Lee, P.C. Ching and Xia Li, "Deriving MFCC parameters from the dynamic spectrum for robust speech recognition," to appear in *Proc. ISCSLP 2008*, Kunming, China, December 2008.

Wireless Mesh Network Testbed

Principal Investigator: *Dah Ming Chiu**
Co-Investigator: *Wing C. Lau*, John CS Lui***
Research Team Members: CP Luk, KK Hui, Jerry Le, ZY Zhang, QL Zhao and others
*****Dept of Information Engineering
**Dept of Computer Science and Engineering

Project Start Date: 1, June 2006
Completion Date: 31, May 2008

Abstract

Wireless mesh networks is a hot research topic. The algorithms for routing and resource allocation for wireless mesh networks are very challenging problems because it is very difficult to precisely model the physical layer fading and interference exactly. Therefore a wireless mesh network for experimentation is very useful. The main goal of this project is to build a wireless mesh network using commodity hardware and public domain software components as much as possible. The usefulness of the wireless mesh will be tested by experimentation by a wireless network routing research project (using path aggregation).

1. Objectives and Significance

Most wireless networking research nowadays relies on simulation or theoretical analysis to evaluate the performance of the proposed designs. However, simulation and theoretical models often fail to capture the complex behavior and irregularities in a wireless environment. Such shortfall is particularly severe under indoor radio propagation environments, e.g. wireless networks deployed within densely populated buildings. This motivates us to design and implement a wireless mesh network testbed in CUHK to support ongoing and future research activities in wireless networking.

2. Results Achieved

We have established a Wireless Mesh Testbed with 30+ nodes equipped with multiple heterogeneous radio interfaces, spanning 4 floors of the two CUHK engineering buildings to support experimental wireless networking research. Software tools have been developed to support Network Management, Self-reconfigurations, and Traffic/ Performance Monitoring of the Testbed. The testbed has been used to conduct various experimental studies of wireless mesh research. For instance, in a project regarding optimizing wireless routing and traffic engineering via path aggregation, an adaptive distributed algorithm for finding spanning subgraphs are proto-typed using the open platform provided by our testbed.

Research findings related to this project have led to four MPhil theses and six publications in major international conferences including IEEE Infocom, ICNP, ICDCS and ICC. Six other papers are currently under submission to leading journals and conferences. A list of these publications can be found in Appendix A of this report.

Besides providing a platform to support wireless networking research, the testbed has also served as an important vehicle in the out-reaching effort of our engineering departments to the local community. Due to its highly tangible nature, the testbed has become a popular demo and the basis of a hands-on workshop for high-school visitors and during annual student-leader-camp and University Open-day. Recently, it has also been an interview subject for an upcoming TV program series produced by the Radio Television Hong Kong (RTHK) on Emerging Information and Communication Technologies.

3. Publications (* = graduate student trained)

1. Jerry Le*, John C.S. Lui, D.M. Chiu, "How many packets can we encode ? An analysis of Practical Wireless Network Coding," IEEE Infocom, 2008.

2. Jerry Le*, John C.S. Lui, D.M. Chiu, "DCAR: Distributed Coding-Aware Routing in Wireless Networks," IEEE ICDCS, 2008.

3. C.P. Luk*, W.C. Lau, O.C.Yue, "An analysis of Opportunistic Routing in Wireless Mesh Networks,", IEEE ICC 2008.

4. K.H. Hui*, O.C. Yue, W.C. Lau, "FRASA: Feedback Retransmission Approximation for the Stability Region of Finite-User Slotted Aloha," IEEE ICNP 2007.

5. K.H. Hui*, W.C. Lau, O.C. Yue, “Stability of Finite-User Slotted Aloha under Partial Interference in Wireless Mesh Networks,” IEEE PIMRC 2007.

6. K.H. Hui*, W.C. Lau, O.C. Yue, “Characterizing and Exploiting Partial Interference in Wireless Mesh Networks,” IEEE ICC 2007.

7. Jerry Le*, John C.S. Lui, D.M. Chiu, “DCAR: Distributed Coding-Aware Routing Protocol for Wireless Networks,” under journal submission.

8. K.H. Hui*, O.C. Yue, W.C. Lau, “FRASA: Feedback Retransmission Approximation for the Stability Region of Finite-User Slotted Aloha,” under journal submission.

9. K.H. Hui*, W.C. Lau, O.C. Yue, “Partial Interference and its Performance Impact on Wireless Multiple Access Networks,” under journal submission.

10. C.P. Luk*, W.C. Lau, O.C. Yue, “Threshold-based Opportunistic Routing Protocol in Multi-hop Wireless Networks,” under conference submission.

11. C.P. Luk*, W.C. Lau, O.C. Yue, “Opportunistic Routing with Directional Antennas in Wireless Mesh Networks,” under conference submission.

12. Z.Y. Zhang*, Yan Yuan*, D.M. Chiu, John C.S. Lui, “An Adaptive Distributed Algorithm for Path Aggregation,” under conference submission.

Disclaimer*: Though related to wireless mesh networks, most of the above papers are theoretical or simulation based, and have not directly made use of the testbed.*

Appendix: Architectural Design and Deployment of the Wireless Mesh Testbed

In this appendix, we describe in detail the design, implementation and deployment of the Wireless Mesh Network Testbed. We first introduce the hardware and software platforms used for realizing the testbed. The implementation of the ad-hoc routing protocol components, the traffic monitor and network management tools are then discussed. We conclude this appendix with a report on the current testbed architecture, deployment status and known system issues.

1. Hardware Components:

Two types of routers are used in the testbed. They are the Pepwave Manga Dual PCI unit and the Linksys WRT54GL router. The combination of these two types of routers forms a heterogeneous wireless mesh network.

a. PepWave MANGA Dual PCI Unit

The PepWave MANGA dual PCI unit is the major component of the network. It provides a fully configurable platform for the deployment of open source software such as MADWIFI driver. Moreover, its USB ports support the use of additional hardware components to further extend its functionality. It has two mini-PCI slots, allowing the installation of two 802.11a/b/g wireless interfaces. With multiple interfaces per node, multiple frequency channels can be assigned to a router. This reduces the interference between mesh nodes and allows more potential research project to be carried out in the network, like resources planning. The presence of 802.11a mode in the Atheros mini-PCI Wi-Fi cards of Pepwave router also allows us to avoid interference generated by numerous existing 802.11b/g devices in the engineering building. 30 Pepwave routers have been purchased. They form the backbone of the mesh testbed. The major drawbacks are their high cost and complexity in configuration. Currently, most of the 30 routers have been deployed with a few remaining ones serve as standby and/or testing units.

b. Linksys WRT54GL

The Linksys WRT54GL is a popular, inexpensive, off-the-shelf component to extend the network coverage. However, the degree of freedom to deploy and modify software components is much smaller than the PepWave counterpart. Although only equipped with one single network interface, the Linksys WRT54GL provides a low cost component for our testbed. It can be placed at the edge of the network to offer last mile wireless coverage. The number of wireless interface of Linksys router can be further extended by attaching a wireless access point to the LAN of the router. It has been tested that the wireless bridging function of WAP54G and DWL-7100AP allow us to construct point-to-point wireless links between routers. The limitation of this configuration is that the wireless links cannot form any loop.

Below is a comparison of the two major hardware components:

	PepWave MANGA Dual	Linksys WRT54GL

	PCI	
Cost	HKD$1550 (10/2007)	HKD$430 (10/2007)
No. of wireless interface	2	1
Machine Architecture	ARM	MIPSEL
CPU	ARM 822Tid at 166MHz	Broadcom 5352 at 200MHz
Memory	RAM: 32MB/ Flash:16MB	RAM: 16MB/ Flash: 4MB
Radio	802.11a/b/g	802.11b/g

2. Software Components:

Several software components are required in the mesh routers: an operating system (Embedded Linux), a routing daemon (OLSRD) and program for management and performance measurements (Iperf and management system daemon)
.

a. Embedded Linux

Each router is loaded with an embedded Linux, an operating system that specifically run on embedded devices, such as router. The embedded Linux provides a platform for the running of other programs, e.g. the OLSRD routing daemon. The embedded linux is included in the build root of the MANGA folder.

b. OLSRD

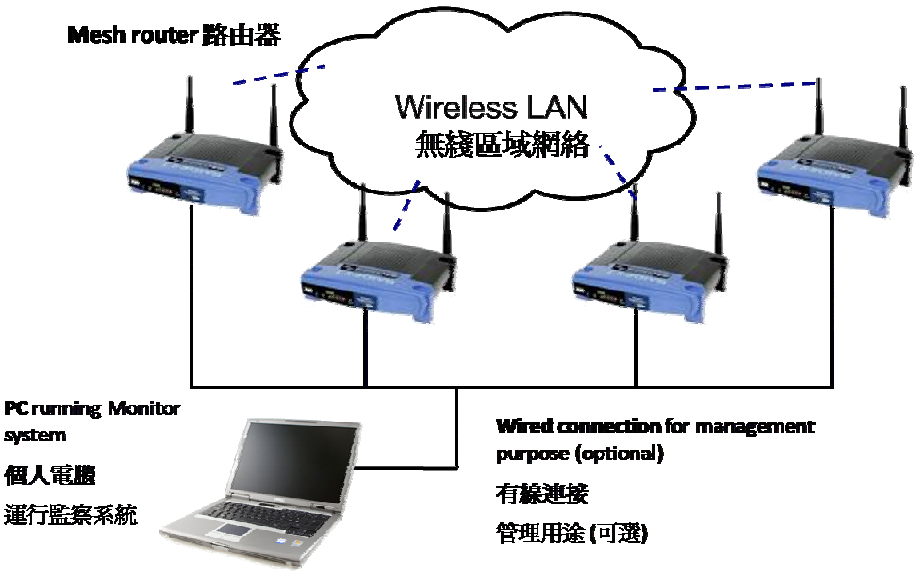
OLSRD is an implementation of the Optimized Link State Routing protocol. With the routing daemon installed, the routers can discover each other and compute the best path between any two nodes. There is a few reason of choosing OLSRD as our routing daemon. Firstly, as all of the routers in the testbed are stationary, OLSR which uses a proactive approach to discover route and maintain routing information can reduces route setup delay. OLSRD also has the gateway function, which allows a mesh router to act as a gateway to the Internet. Finally, it supports the use of multiple interfaces and multiple subnets, which increases the scalability of the network. It has been verified that OLSRD has no known problem to run in both Pepwave and Linksys routers by cross compiling it with the corresponding cross complier. Routing daemons in different system can also communicate with each other without any problem. The ad-hoc routing protocol used by our testbed is based on version 4.10 release of OLSRD, with additional customizations to support the aforementioned hardware and software platform.

c. IPERF

In addition, iperf is ported the routers. Iperf is a network utility which allows the user to measure the TCP and UDP bandwidth and performance. It is useful in network planning and research experiments. The iperf version used is 1.7.0. Since the size of iperf daemon is large, it is better to reduce the size of the daemon before deploying to the routers.

d. Management system daemon

We have also implemented a daemon used by the management system which collects information from the router and receive instructions from the central server. Besides the wireless interfaces, all routers and monitor server are all connected to a virtual LAN in the engineering building. The VLAN is separated from IE department network and cannot access the Internet due to security reasons. Each router has a controller daemon installed. The daemon collects information of the routers such as routing table and interface information and sends back the data back to the monitor server. The monitor server aggregates the information collected and publish it through a web interface. The diagram below shows the architecture of the monitor system.



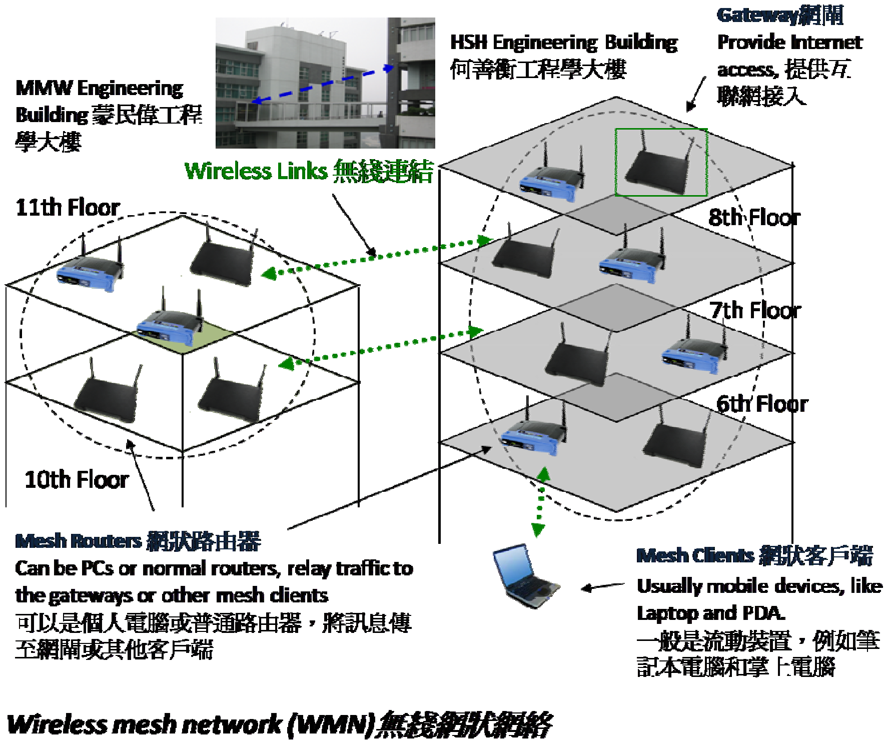
e. Patch System

A patch script is installed in each router. As the server and routers are connected by wired connection, from the central server, we can run scripts which upload files to all routers, configure the interfaces and subnet of the routers or run specific commands in a batch according to our requirements automatically.

3. The Testbed Architecture

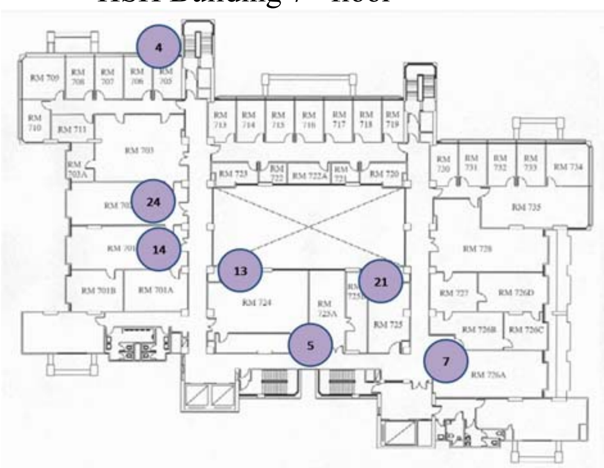
The routers are deployed in the two engineering buildings. Currently, they are deployed in 10th, 11th floors of MMW engineering and the 7th and 8th floors of HSH engineering buildings. The routers form a wireless mesh networks among themselves. In addition, there are cross-building wireless links between two buildings.

The diagram below shows the architecture of the testbed.

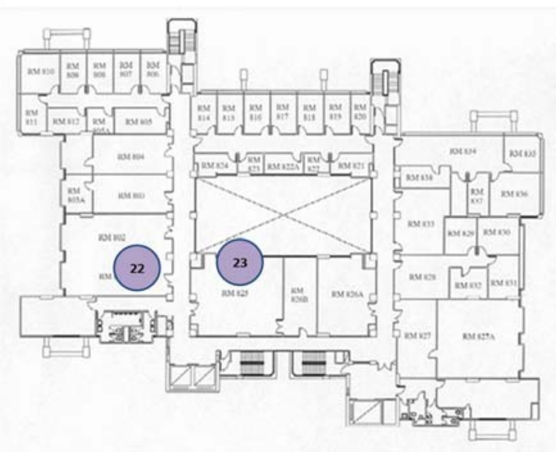


4. Current Deployment Floor Plan

HSH Building 7th floor

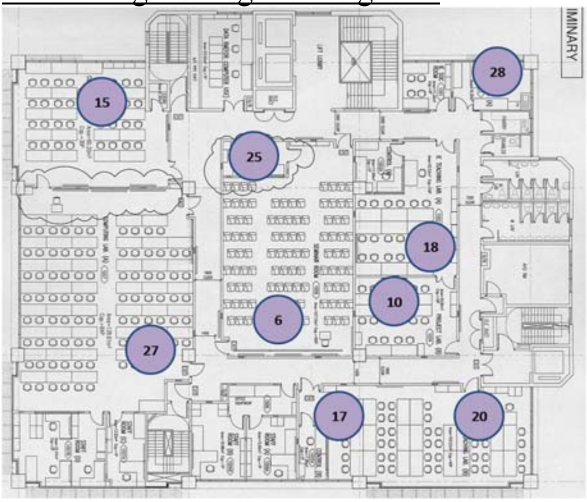


HSH Building 8th floor

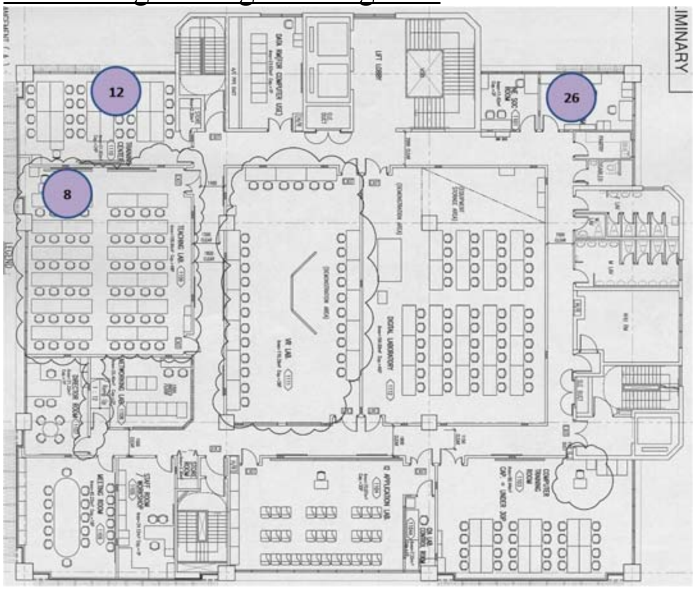


Note Room 817, 833 and 827 have port patched for VLAN but no router is deployed.

MMW Engineering Building 10/F



MMW Engineering Building 11/F



All other routers are reserved for testing and future deployment.

5. Known System Issues

Due to memory leak in the monitor server, the monitor server daemon: monitor_s needs to be restarted for every 3-4 hours. The problem is still not solved because of the limited resources and the complexity of the problem.