2009 Course Proceedings
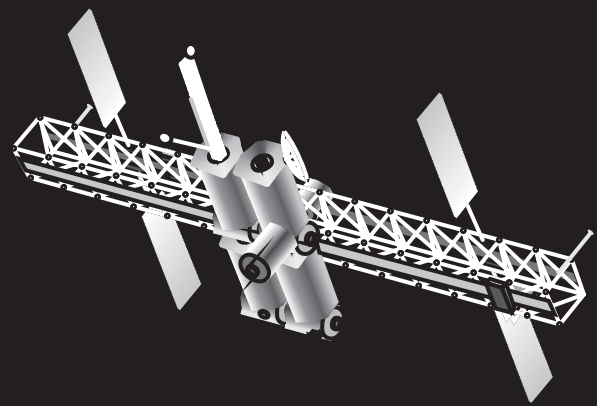
Communication:

An Exploration of Genres

Patterning of PMMA Microfluidic Parts using Screen Printing Process

Aminreza Ahari Kaleibar, Ash Parameswaran
Institute for Micromachine and Microfabrication
Research, Simon Fraser University, 8888 University Drive, V5A 1S6, BC Canada
Email: aaa83@sfu.ca

Abstract-- **An inexpensive and rapid micro-fabrication process has been presented and systematically characterized. Unlike the former efforts, the proposed process takes advantages of cheaper technologies such as the silk screen printing and UV opaqueness of the printing-ink to directly expose fluidic channel patterns on PMMA substrates. As a proof of principle, utilizes a homemade deep-UV, λ=254nm, source system, and the silk screen printing technology supplied through a local screen-printing shop. The prototyped devices were successfully bonded and sealed, and their proper operation was tested.**

Index Terms-- **biosensors, microfluidics, microfabrication, and screen printing process**

I. INTRODUCTION

T he microfluidic devices are inevitable building blocks of biosensors that find lots of applications in biological assays detection, and point-of-care diagnostic tools such as cancer cell detection, pathogenic bacterial cell detection in food, or the testing of water quality for possible contaminations [1, 2]. Nonetheless, their fabrication process requires precise and quite expensive equipments, and highly trained technicians, which makes their mass manufacturing expensive and time consuming; therefore, developing rapid and inexpensive micro-fabrication processes are highly demanded.

Depending on the applications of the microfluidic devices, silicon, glass, and polymer as the substrate materials are used for their construction. Given well- established microfabrication processes for silicon and glass in semiconductor industry and microelectromechanical systems (MEMS), silicon and glass as the substrate of the microfluidic devices received much attention. However, certain polymers are used as an economical alternative to glass and silicon [3]. In order for decreasing associated costs, a micro-fabrication technique has been reported in [3] which uses deep-UV for etching Poly Methyl Meth Acrylate (PMMA). Although, its process is relatively inexpensive, the process includes photolithographic techniques, the deposition of Au on PMMA, and etching it. Thus, the reported microfabrication process depends chiefly on clean room equipment. This paper describes a novel micro-fabrication process that takes advantages of cheaper technologies such as silk screen printing and UV opaqueness

of the printing-ink to directly expose fluidic channel patterns on PMMA substrates, and, as proof of principle, utilizes a homemade deep-UV source system, and the silk screen printing technology. Hence, this experimental set-up demonstrates the proposed process is fundamentally simple, rapid, and inexpensive.

The following sections of the paper are devoted to describe the proposed fabrication procedure, and then to present the experiments. Finally, it will bring results, and will discuss the results.

II. FABRICATION

In an earlier attempt at an economical micro-fabricarion process, a novel technique has been reported in [3], which begins with depositing a bi-layer of Au and Cr on PMMA. Then, the designed micro-channel is patterned on Au by the process of photolithography and etching. Indeed, Au performs as a hard mask for etching PMMA. In the next step of this process the long chains of the exposed PMMA are broken using deep-UV, λ=254nm. Shorted PMMA can be easily dissolved using IPA-water mixture [4]; therefore, the designed micro-channel is developed by a solution of 7:3 IPA: water. Whereas the reported technique is relatively inexpensive, it chiefly depends on the clean room equipments that constrain its mass manufacturing.

In order to optimize this method and alleviate the fabrication costs, we present a novel technique that takes advantage of cheaper technologies such as silk screen printing and UV opaqueness of ink that protects PMMA from deep-UV exposure and patterns micro-channel; therefore, the whole process of depositing the metal hard mask, Au, on PMMA, performing photolithography, pattering on the photoresist layer, and etching can be bypassed. In other words, the ink screen printing is substituted for those of [5], which are expensive and time consuming.

In Fig. 1, the proposed process has been depicted. First, the microfluidic device design is created using a CAD program and a negative emulsion image is created. Using this emulsion a silk-screen is created through a local screen-printing shop. Using a squeegee the printing ink is applied on a PMMA substrate. The printing ink shows excellent UV-Opaqueness. After the ink has dried on the PMMA substrate the sample is exposed using deep-UV. This process breaks the long chains of PMMA. Shorted PMMA can be easily dissolved using IPA-
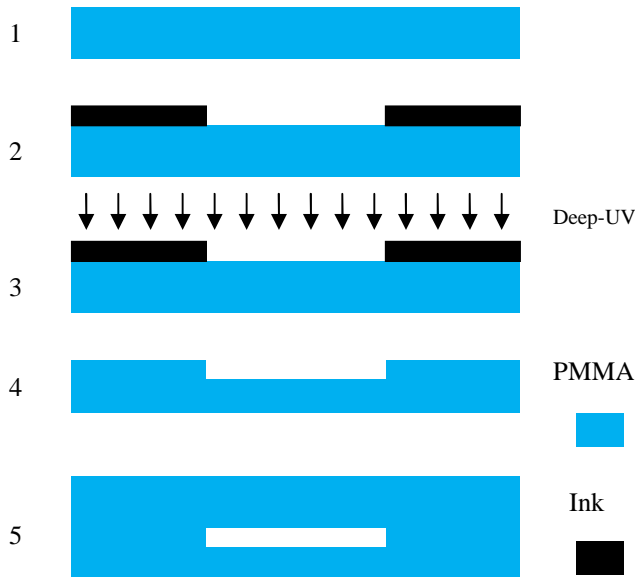
Fig. 1. The proposed micro-fabrication process. In order to fabricate a micro-channel on PMMA, (2) a negative image of the micro-channel is printed on PMMA by the silk screen printing technology, (3) then it is exposed to deep-UV, λ=254nm, using a homemade UV source, which leads to shorten the chemical bonds of exposed areas. (4) Given that PMMA with shortened chemical bonds and ink are dissolved in a mixture of 7:3 IPA: water, the channel can be micro-patterned in the same step, (5) an other PMMA film is attached on top of the pattered one to form the microfluidic device.

water mixture [4]. Interestingly enough, IPA also dissolves the ink; therefore, we need not employ a separate processing step to remove the ink. The development step, PMMA etching step, as well as the ink removal step are combined into one single processing step. Finally, the PMMA plate with patterned microchannels are bonded by another PMMA plate, and sealed carefully.

III. EXPERIMENT

We set up experiments aimed at systematically characterizing the proposed process. First, a negative emulsion image of different channel widths from 12μm to 2mm by 36μm increments has been designed using Cadence software. Using this emulsion a silk-screen is created by InkPlus [6], supplier of UV-opaqueness ink. For the fabrication purposes, PMMA from Plaskolite's OPTIX [7] has been used because of its good mechanical and optical properties, and its low price. For the purpose of our experiments, we built up a homemade deep-UV exposure system that utilizes an array of mercury vapor lamps as a source of deep-UV. The lamps are located inside a robust steel-made box, which shields them from outside, and are carefully sealed; therefore, the exposure system is user- and environmental-friendly equipment. The radiation power of the lamps is limited and fixed. The characterizing of the proposed process requires a variable exposure power, because it is expected to be proportional with etch depth; thus, we vary the exposure time, which accounts for the different energy received by the exposed PMMA samples. The chosen exposure times, for this purpose, are 5, 10, and 15 hours. Subsequently, the development step can be done at clean room

and a solution of 7: 3 IPA: water is used for developing the exposed PMMA and removing the printed ink. The temperature inside the solution is required to be kept constant at 28°C. We prepare the mentioned solution by mixing 280ml IPA with 120ml water, and set its temperature at 28°C. During the development, the depth and width of the micro-channels are regularly measured using a TENCOR profilometer, Alpha-Step 500.

IV. RESULTS and DISSCUSSION

Our measurement results are depicted in Fig.2, 3, 4.The summary of the measured etch depth versus exposure time is depicted in Fig. 5. This graph shows that 70μm channel depth can be achieved after 15 hours exposure by deep-UV source, and the development time is 30 minute. On the other hand, after 5 hours exposure, the etch depth will be 25μm, and 60 minutes required to develop it. The longer exposure time is the faster developing rate is achieved. However, the increased exposure time results in rough microchannel surfaces that is undesirable for some applications.
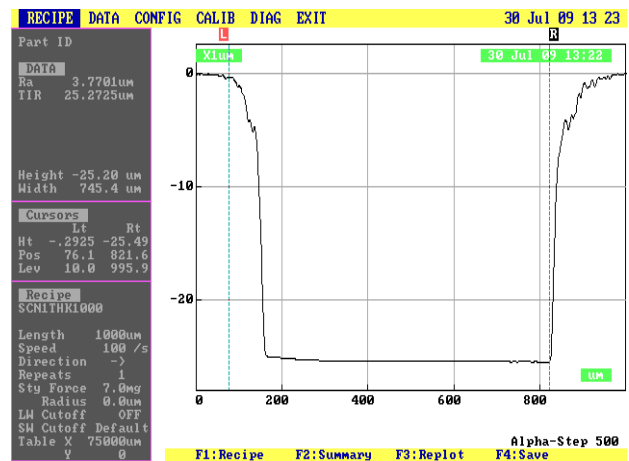

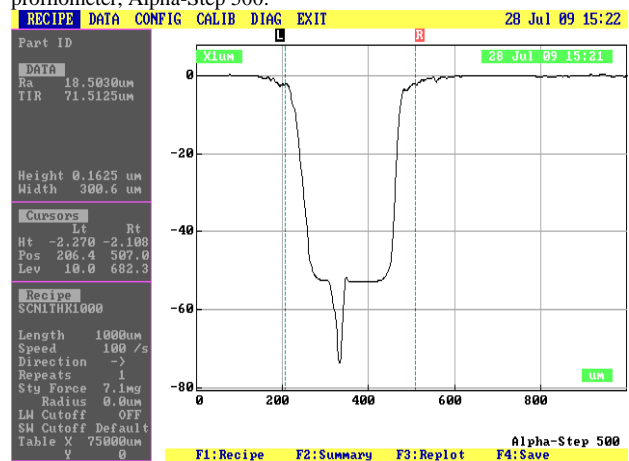**Fig.2.** The measured channel depth for 5 hour exposure using TENCOR profilometer, Alpha-Step 500.

Fig.3. The measured channel depth for 10 hour exposure using TENCOR profilometer, Alpha-Step 500.
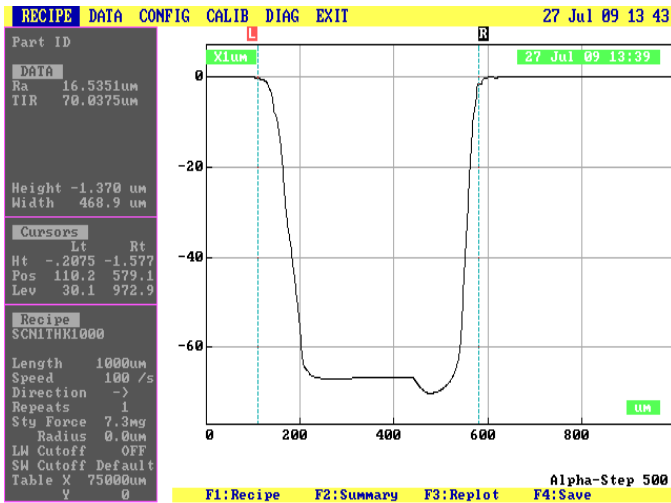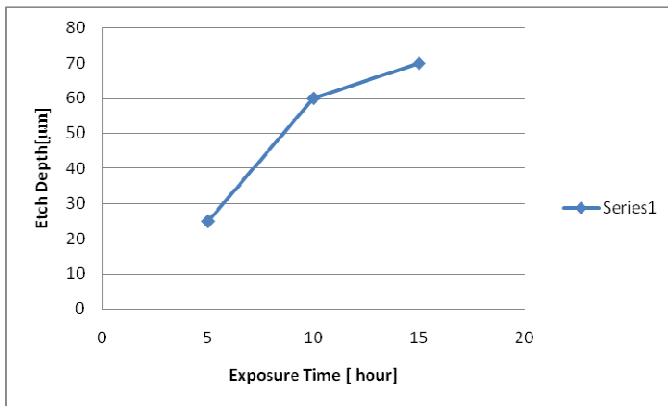
[4] R.W. Johnstone et al "IPA/Water as a Developer for Poly (dimethylglutarimide)," Journal of Micro/Nanolithography, MEMS, and MOEMS, 2008
[5] M.Haiducu et al "Deep-UV patterning of commercial grade PMMA for low-cost, large-scale microfluidic" Journal of Micromechanics and Microengineering, 2008
[6] Optix acrylic sheet www.plaskolite.com/productguide/04.pdf
[7] InkPlus www.inkplus.ca

Fig.4. The measured channel depth for 15 hour exposure using TENCOR profilometer, Alpha-Step 500.

Fig.5. The etch depth in micro-meters versus exposure time in hours

Aminreza Ahari Kaleibar was born in Tabriz, Iran in 1978. He received his BSc in electrical engineering –electronics from Iran University of Science and Technology in 2000. Since January 2009, he joined school of engineering science at Simon Fraser University, Burnaby, Canada, and has been working toward his MASc degree in electrical engineering.

V. CONCLUSION

In this paper, an inexpensive and rapid microfluidic fabrication process presented, by taking advantages of silk screen printing process and UV-opaqueness of ink. The process was systematically characterized by an experimental set-up. We demonstrated that the proposed process can be ultra-economical alternative to current microfluidic fabrication techniques, and in particular make their mass manufacturing feasible. The promising results of this work encourage us examining its applicability to a microfluidic device such as micromixer in future.

ACKNOWLEDGEMENT

REFERENCES

[1] M.Javanmard et al "Direct Electrical Detection of Target Cells on a Microfluidic Biochip" Proc. of SPIE Vol. 6886, 68860B, (2008)
[2]J.Liu et al "Real-time Protein Detection Using ZnO Nanowire/Thin Film Bio-sensor Integrated with Microfluidic System" IEEE Electronic Components and Technology Conference, 2008
[3]H. Dong et al "Screen-printed microfluidic device for electrochemical immunoassay" Lab on a Chip 2007

Power and Energy Management in Data Centers

Data centers are facilities that house a large number of servers and associated components like storage systems and communication infrastructure which provide infrastructure supporting a variety of Internet services. These services require high availability (i.e. close to 99.9% up time). In order to provide high availability and fault tolerance, servers are replicated and spare capacity is always maintained. Naturally, data centers tend to consume large amounts of power.

Power consumption has become a key concern in today's computing landscape and is considered as a first order constraint in the design, deployment and operation of computing facilities. Consequently, power and energy management have become key areas of research. Reduction in power consumption results in reduced total cost of operation (TCO) and also allows the data center operator to invest in newer hardware and supporting infrastructure to provide improved services to clients. Significant research has been undertaken in this area. This paper details the motivation for the research and provides a short survey of previous work related to power management on server clusters.

The introduction of virtualization poses new problems and challenges. The virtual machine monitor (or hypervisor) must be capable of handling and mediating between (potentially) opposing power management requests issued by the guest virtual machines. We investigate how virtualization has affected data center power & Quality of Service (QoS) management. We discuss techniques that have been proposed for power management in hypervisors on individual machines and how they can be applied to data centers.

QoS requirements and power management requirements typically tend to be of a conflicting nature. The consequence of power management is reduced performance. This reduction in performance can affect the QoS by increasing latency, reducing throughput, and in the case of Web hosting, resulting in poor user experience. Our survey investigates prior work in QoS-aware job scheduling and power management in data centers.

Cooling forms a considerable part of the power consumption of a data center. Thermal-aware scheduling can help reduce the cooling costs associated with data center operation. Such a scheme would require information on the physical location of the servers and cooling systems to be incorporated into the job scheduling algorithms. We investigate the previous work on mechanisms and policies for thermal aware job scheduling.

Categories and Subject Descriptors: 1 [**Power Management**]: —*Power management, Job Scheduling, Quality of Service, Data centers*

1. INTRODUCTION

A data center houses a large number of servers and their associated components such as storage and communication systems. Typical data centers house hundreds, maybe even thousands of servers and serve as the backbone for a variety of Internet services. A data center can be broadly viewed as consisting of hardware infrastructure, software infrastructure and a cooling infrastructure. The hardware infrastructure can be further broken down as consisting of the servers that provide computational power, storage equipment, and the networking infrastructure that provides local and remote network access. Software infrastructure includes Operating Systems (OS), administrative software for data center monitoring and management, and cluster level infrastructure (or middle-ware) that provide transparent access of resources to application level software. Finally, the cooling infrastructure includes equipment that provides cooling for the facility.

1.1   Types of data centers

Based on the layout, data centers can be broadly classified into one of the 4 tiers listed in Table I. This classification however, is not precise and commercial data centers typically fall between Tiers 3 and 4[Barroso and Hölzle 2009]. A higher tier implies an improvement in resource availability and reliability, but is accompanied with an increase in power consumption.

Table I.   Data Center Classification

Type	Description
Tier 1	Single path for power and cooling; no redundant components.
Tier 2	Redundancy added to Tier 1 thereby improving availability.
Tier 3	Multiple power and cooling distribution paths of which one is active.
Tier 4	Two active power and cooling paths and redundant components on each path.

1.2   The Case for Power Management in Data Centers

Data centers host services that require high availability, close to 99.99%. Fault tolerance, therefore, becomes imperative. The loss of one or more components must not cause the data center to terminate its services to clients. Consequently, data centers feature hardware redundancy. Furthermore, data centers are designed for a peak load which might be observed only occasionally and for short bursts of time. This conservative design results in over-provisioning of hardware in a data center. All these factors combined contribute to the high power consumption of data centers. Studies [Koomey 2007; Shah and Krishnan 2008] have reported the global electricity costs for data centers running into billions of dollars. Increasing numbers of data centers are being setup to serve the needs of an expanding Internet, and we can expect the total costs of data centers to increase in coming years. Improved power management can, therefore, reduce electricity costs for the data center operator and allow the operator to invest in providing improved services.

2.   PLATFORM POWER MANAGEMENT

Power management is a feature that allows devices to reduce power consumption when under low-utilization or idle. Today, power is considered as a first order constraint in the design of computing systems, therefore, power management support has become mandatory to all components on a computing platform. The components include processors, interconnect buses, device buses like Peripheral Component Interconnect (PCI/PCI-Express) and Universal Serial Bus (USB), and peripheral devices (hard disks, networking cards etc). Processor power management features include Dynamic Voltage and Frequency Switching (DVFS)[1] which can be used to vary processor supply voltage and operating frequency. Processors also support sleep states in which they consume a very small amount of power(processor C-States). Interconnect buses may also support multiple operating frequencies, thereby extending the concept of DVFS to be applicable on them. Device buses like PCI-e support Active State Power Management (ASPM) features; peripheral

---

[1]DFS refers to Dynamic Frequency Switching.

devices like disks and networking cards (both wired and wireless) support low power D-states.

3. DATA CENTER POWER & THERMAL MANAGEMENT

A data center can be viewed as a facility that houses computing components, auxiliary computing components, and non-computing components. The computing components are the server clusters which perform computational tasks and serve requests. We use the term auxiliary components to refer to the communication network and network-attached storage that the servers use. Non computing components include Computer Room Air Conditioning (CRAC) units. Power consumption of the data center is the aggregate power consumed by all these components. A unit of computation would require power at the processor, for IO operations, and for removing the heat generated during the computation.

3.1 Power Management for Server Clusters

Power management and energy efficiency in server clusters has been an active area of research. The mechanisms discussed in Section 2, though not many, have been used in a variety of ways to provide power savings – DVFS being the most popular approach used. An Operating System can vary the frequency of a server based on CPU utilization. At high utilization, the OS raises the processor operating point (voltage-frequency value), providing better performance, and at low utilization, the operating point is reduced to conserve power. Each server can vary its voltage and frequency in response to the load observed locally. This approach is termed Independent Voltage Scaling (IVS). An alternate approach is to measure the load across the entire cluster, identify the frequency for the entire cluster, and set each server to the operating point closest to the cluster frequency. This approach is Coordinated Voltage Scaling (CVS) [Elnozahy et al. 2002]. While DVFS can provide power savings, placing idle servers either in a low power state or turning them off completely are more effective [Elnozahy et al. 2002; 2003; Rajamani and Lefurgy 2003]. [Elnozahy et al. 2002] use Vary On–Vary Off (VOVO) policy combined with IVS and CVS to improve the power footprint of the server cluster. [Elnozahy et al. 2003] propose holding requests in memory, waking up the server to handle all the requests, and then placing the server back in a deep sleep state. They term it *request batching* because the server executes batches of requests.

Schemes that turn off servers require a request distribution scheme that can consolidate load onto fewer servers, thereby allowing idle servers to be turned off. Load concentration can provide improved performance per Watt for the entire cluster. [Pinheiro et al. 2001] investigate the effects of load balancing and consolidation by implementing the balancing scheme at the application level, OS level, and through application-OS negotiation. Power can also be conserved by exploiting the heterogeneity of the servers on a cluster and the nature of the jobs executing on it. Jobs that are CPU intensive can be scheduled on the "fast" servers whereas jobs that are memory intensive can be scheduled on the "slow" processors. The alternative approach, proposed by [Kotla et al. 2005] varies processor frequency on individual servers based on application characteristics, namely the Instructions Per Cycle (IPC) metric. Applications that are mostly memory-bound do not benefit from increased frequency and, therefore, can be executed at lower frequency without

significant loss of performance.

Load concentration and power management schemes must provide acceptable performance, measured as "high throughput" for batch jobs and "low latency" for interactive jobs. Power management entails a reduction in performance due to the latency associated with bringing the components into a fully functional state. This latency can lead to a perceivable loss of Quality of Service (QoS). [Sharma et al. 2003] propose a framework for power management that takes into account an application's QoS requirements. [Elnozahy et al. 2003] combine their request batching scheme with DVFS to provide additional power savings without significant loss of QoS. Theoretical models have also been studied. [Mastroleon et al. 2005] provide a Markov decision process model consisting of available CPUs, jobs, and the thermal conditions. Each job assignment has an associated cost. The optimal schedule involves identifying the number of CPU's to assign to a given job in the current time slot, given thermal constraints. An important limitation of their model is that jobs are dropped if the job buffer is full. [Raghavendra et al. 2008] use a control theoretic approach for coordinated power management in data centers. Data centers have multiple 'controllers' which are taking input from the system and initiating actions to control power consumption, provide QoS and maintain system stability. The individual controllers are local/global DVFS schemes, temperature based control, and VM based power management. The authors argue that a lack of coordination between these elements would result in instability. Their simulation results show that coordinated power management causes fewer violations and less performance loss compared to an un-coordinated mechanism.

Virtualization allows data center operators to consolidate resources and improve security, availability, and reliability. A virtualization layer hosts multiple Operating System instances and manages the underlying platform resources. When running on virtualization, each guest Operating System assumes that it has complete control of the underlying resources. Each guest OS can issue independent power management requests to the hardware based on its local information. The virtual machine monitor or host OS must arbitrate between potentially conflicting requests and use global information to take decisions. [Tian et al. 2007] capture some of the challenges in power management with virtualization. [Nathuji and Schwan 2007] describe a power management solution for virtual clients. Their solution allows virtual guests to control the power states of the underlying hardware by intercepting the power management requests made by guest VMs. Metrics like CPU utilization and delays must account for changes in timing due to virtualization.

Our discussion has considered the case that all servers in the cluster execute the same kind of workloads. Commercial data centers would be different. The setup of a data center would consist of multiple server tiers. Each tier handles a certain type of request. For example, in a multi-tiered setup for an e-commerce website, the first tier might handle requests for the most commonly used HTML pages and images. The second tier handles secure connections and dynamic content, the third handles database operations and provides access to networked attached storage. Note that the tiered setup mentioned here is different from the tier-scheme for data centers presented in Section 1. In a multi-tiered setup, it becomes necessary to identify how many machines must be kept active in each tier such that the QoS constraints

are met. [Horvath and Skadron 2008] investigate the problem. They formulate a linear optimization problem to optimize the power consumption of the cluster. The constraints for the optimization are the number of machines per tier, the number of sleep states supported, and the operating frequencies supported.

3.2 Heterogeneity

Most research work assumes homogeneous clusters for evaluation because they provide simplicity in setup and analysis. This observation is also made by [Bianchini and Rajamony 2004]. Homogeneity, in this case, refers to similarity in processor capabilities (from a computational standpoint) and similarity in platform capabilities (from a power management standpoint). Homogeneity may be prevalent in a data center that is either newly established or owned, operated, and utilized by the same company (data centers owned by Google, Yahoo, Microsoft etc). However, it is highly likely that a data center contains computing components that are heterogeneous. Heterogeneity can stem from

—Differences in platform computational capabilities caused due to differences in processor features, platform architecture, memory architecture, etc.

—Differences in platform power management capabilities, namely operating states supported, support for low power sleep states, device power states, resume latency, etc.

[Heath et al. 2005] investigate request redistribution mechanisms for heterogeneous clusters. They dynamically reconfigure clusters and the request distributions in order to optimize power/energy or latency metrics by using analytical models to predict cluster throughput and power consumption. An analytical model is used to predict cluster throughput & power, both being a function of the request distribution scheme. In their approach, [Horvath and Skadron 2008] account for heterogeneity between servers in different tiers; however, servers in a given tier are assumed to be homogeneous.

4. THERMAL AWARE SCHEDULING

4.1 Cooling infrastructure of a data center

A data center has Computer Room Air Conditioning(CRAC) units that circulate temperature and humidity controlled air. The CRAC units push cold air through vents in the floor to all parts of the data center. The cool air is sucked by processor cooling fans from the front of the server racks, and hot air is expelled to the back of the racks. The arrangement of server racks is such that hot and cold aisles are formed. This arrangement reduces mixing of hot and cold air and helps improve cooling efficiency. [Schmidt 2005] capture many of the challenges faced in data center cooling design.

Different kinds of workloads stress different parts of a computing platform, consequently exhibiting different heat generation profiles. A computationally intensive workload would cause CPU heat to increase significantly compared to a memory or disk intensive workload. Thermal awareness allows jobs to be scheduled to create an improved heat distribution profile in the data center so that cooling can be more

efficient and effective. It also reduces the probability of critical server shutdown caused due to violation of thermal limits.

[Moore et al. 2005] explore solutions to control heat generation by temperature aware workload placement. Their Zone Based Discretization (ZBD) policy schedules tasks while attempting to create a uniform exhaust profile and reduce local hot spots. The Minimum Heat Recirculation (MinHR) policy attempts to minimize the heat recirculation and maximize the power budget available for individual machines. [Moore et al. 2006] use machine learning techniques to predict the thermal profile of a data center based on workload characteristics, cooling configuration, and physical topology. [Parolini et al. 2008] visualize the data center to be composed of a computational network and a thermal network. The computational network describes tasks and the servers. The thermal network captures the heat exchange in the data center and accounts for non-computational components (like the CRAC units). The authors model the problem as a Constrained Markov Decision Process (CMDP), where they attempt to minimize the power consumption. Approaches for thermal aware workload placement can provide improvements if task properties [Vanderster et al. 2007; Rajamani et al. 2006] are also taken into account.

5. CONCLUSION

Our paper discussed various algorithms and techniques aimed at reducing server cluster power consumption. However, most work assumes homogeneous hardware and does not take workload characteristics into account. We also discussed algorithms for thermal aware scheduling that could reduce cooling costs. The thermal aware scheduling algorithms do not measure the impact on cluster performance. Thermal aware job scheduling attempts to spread load across machines to create a more uniform thermal profile. The idea of spreading the load (to generate a uniform thermal profile) intuitively goes against load consolidation algorithms (for saving processor power). This conflit provides us scope to investigate the tradeoffs between scheduling for power savings and performance.

Each component in the data center including software and hardware could individually and independently attempt power/thermal management. As pointed out by [Raghavendra et al. 2008], lack of proper coordination between these components could lead to system instability. Standardised interfaces have been defined in the Advanced Configuration and Power Interface (ACPI) specification for interaction between operating system and hardware. However, interfaces can also be defined between para virtualized and guest OSes and between guest OSes. An approach that uses application variety and hardware variety can provide improved power savings. Naturally, such a solution cannot be expected to be trivial and offers considerable scope for future research work.

REFERENCES

BARROSO, L. A. AND HÖLZLE, U. 2009. The Datacenter as a Computer: An Introduction to the Design of Warehouse-Scale Machines. *Synthesis Lectures on Computer Architecture 4,* 1, 1–108.

BIANCHINI, R. AND RAJAMONY, R. 2004. Power and energy management for server systems. *Computer 37,* 11 (Nov.), 68–76.

CASSATT CORPORATION. Cassatt white paper. Whitepaper, Cassatt Corporation.

CHASE, J. S., ANDERSON, D. C., THAKAR, P. N., VAHDAT, A. M., AND DOYLE, R. P. 2001. Managing energy and server resources in hosting centers. In *SOSP '01: Proceedings of the eighteenth ACM symposium on Operating systems principles.* ACM, New York, NY, USA, 103–116.

CHASE, J. S. AND DOYLE, R. P. 2001. Balance of power: Energy management for server clusters. In *In Proceedings of the 8th Workshop on Hot Topics in Operating Systems (HotOS.*

CURTIS-MAURY, M., SHAH, A., BLAGOJEVIC, F., NIKOLOPOULOS, D. S., DE SUPINSKI, B. R., AND SCHULZ, M. 2008. Prediction models for multi-dimensional power-performance optimization on many cores. In *PACT '08: Proceedings of the 17th international conference on Parallel architectures and compilation techniques.* ACM, New York, NY, USA, 250–259.

ELNOZAHY, E. M., KISTLER, M., AND RAJAMONY, R. 2002. Energy-efficient server clusters. In *In Proceedings of the 2nd Workshop on Power-Aware Computing Systems.* 179–196.

ELNOZAHY, M., KISTLER, M., AND RAJAMONY, R. 2003. Energy conservation policies for web servers. In *USITS'03: Proceedings of the 4th conference on USENIX Symposium on Internet Technologies and Systems.* USENIX Association, Berkeley, CA, USA, 8–8.

GE, R., FENG, X., AND CAMERON, K. W. 2005. Performance-constrained distributed dvs scheduling for scientific applications on power-aware clusters. In *SC '05: Proceedings of the 2005 ACM/IEEE conference on Supercomputing.* IEEE Computer Society, Washington, DC, USA, 34.

GHIASI, S., KELLER, T., AND RAWSON, F. 2005. Scheduling for heterogeneous processors in server systems. In *CF '05: Proceedings of the 2nd conference on Computing frontiers.* ACM, New York, NY, USA, 199–210.

HEATH, T., DINIZ, B., CARRERA, E. V., JR., W. M., AND BIANCHINI, R. 2005. Energy conservation in heterogeneous server clusters. In *PPoPP '05: Proceedings of the tenth ACM SIGPLAN symposium on Principles and practice of parallel programming.* ACM, New York, NY, USA, 186–195.

HORVATH, T. AND SKADRON, K. 2008. Multi-mode energy management for multi-tier server clusters. In *PACT '08: Proceedings of the 17th international conference on Parallel architectures and compilation techniques.* ACM, New York, NY, USA, 270–279.

INTEL CORPORATION. 2004. Addressing Power and Thermal Challenges in the Datacenter. Whitepaper, Intel Corporation.

KOOMEY, J. 2007. Estimating total power consumption by servers in the US and the world. Tech. rep., Lawrence Berkeley National Laboratory. February.

KOTLA, R., GHIASI, S., KELLER, T., AND RAWSON, F. 2005. Scheduling processor voltage and frequency in server and cluster systems. *Parallel and Distributed Processing Symposium, 2005. Proceedings. 19th IEEE International*, 8 pp.–.

KOTOV, V. 2001. On virtual data centers and their operating environments. Tech. Rep. HPL-2001-44, HP.

LEFURGY, C., RAJAMANI, K., RAWSON, F., FELTER, W., KISTLER, M., AND KELLER, T. 2003. Energy management for commercial servers. *Computer 36,* 12 (Dec.), 39–48.

LI, T., BAUMBERGER, D., KOUFATY, D. A., AND HAHN, S. 2007. Efficient operating system scheduling for performance-asymmetric multi-core architectures. In *SC '07: Proceedings of the 2007 ACM/IEEE conference on Supercomputing.* ACM, New York, NY, USA, 1–11.

MASTROLEON, L., BAMBOS, N., KOZYRAKIS, C., AND ECONOMOU, D. 2005. Automatic power management schemes for internet servers and data centers. *Global Telecommunications Conference, 2005. GLOBECOM '05. IEEE 2*, 5 pp.–.

MOORE, J. AND CHASE, J. 2005. Data center workload monitoring, analysis, and emulation. In *in Eighth Workshop on Computer Architecture Evaluation using Commercial Workloads*.

MOORE, J., CHASE, J., AND RANGANATHAN, P. 2006. Weatherman: Automated, online and predictive thermal mapping and management for data centers. *Autonomic Computing, 2006. ICAC '06. IEEE International Conference on*, 155–164.

MOORE, J., CHASE, J., RANGANATHAN, P., AND SHARMA, R. 2005. Making scheduling "cool": temperature-aware workload placement in data centers. In *ATEC '05: Proceedings of the annual conference on USENIX Annual Technical Conference*. USENIX Association, Berkeley, CA, USA, 5–5.

NATHUJI, R., ISCI, C., AND GORBATOV, E. 2007. Exploiting platform heterogeneity for power efficient data centers. In *ICAC '07: Proceedings of the Fourth International Conference on Autonomic Computing*. IEEE Computer Society, Washington, DC, USA, 5.

NATHUJI, R. AND SCHWAN, K. 2007. VirtualPower: coordinated power management in virtualized enterprise systems. *SIGOPS Oper. Syst. Rev. 41,* 6, 265–278.

PAI, V. S., ARON, M., BANGA, G., SVENDSEN, M., DRUSCHEL, P., ZWAENEPOEL, W., AND NAHUM, E. 1998. Locality-aware request distribution in cluster-based network servers. *SIGPLAN Not. 33,* 11, 205–216.

PAROLINI, L., SINOPOLI, B., AND KROGH, B. H. 2008. Reducing data center energy consumption via coordinated cooling and load management. In *HotPower '08: Workshop on Power Aware Computing and Systems*. ACM.

PINHEIRO, E., BIANCHINI, R., CARRERA, E., AND HEATH, T. 2001. Load balancing and unbalancing for power and performance in cluster-based systems. Tech. Rep. DCS-TR-440, Department of Computer Science, Rutgers University, New Jersey. May.

RAGHAVENDRA, R., RANGANATHAN, P., TALWAR, V., WANG, Z., AND ZHU, X. 2008. No "power" struggles: coordinated multi-level power management for the data center. *SIGPLAN Not. 43,* 3, 48–59.

RAJAMANI, K., HANSON, H., RUBIO, J., GHIASI, S., AND RAWSON, F. 2006. Application-aware power management. In *Workload Characterization, 2006 IEEE International Symposium on*. 39–48.

RAJAMANI, K. AND LEFURGY, C. 2003. On evaluating request-distribution schemes for saving energy in server clusters. *Performance Analysis of Systems and Software, 2003. ISPASS. 2003 IEEE International Symposium on*, 111–122.

RANGANATHAN, P., LEECH, P., IRWIN, D., AND CHASE, J. 2006. Ensemble-level power management for dense blade servers. *Computer Architecture, 2006. ISCA '06. 33rd International Symposium on*, 66–77.

SCHMIDT, R. R. 2005. Challenges of data center thermal management. *IBM Journal of Research and Development 49,* 4, 709.

SHAH, A. AND KRISHNAN, N. 2008. Optimization of global data center thermal management workload for minimal environmental and economic burden. *Components and Packaging Technologies, IEEE Transactions on 31,* 1 (March), 39–45.

SHARMA, R., BASH, C., PATEL, C., FRIEDRICH, R., AND CHASE, J. 2005. Balance of power: dynamic thermal management for internet data centers. *Internet Computing, IEEE 9,* 1 (Jan.-Feb.), 42–49.

SHARMA, V., THOMAS, A., ABDELZAHER, T., SKADRON, K., AND LU, Z. 2003. Power-aware QoS management in Web servers. *Real-Time Systems Symposium, 2003. RTSS 2003. 24th IEEE*, 63–72.

TIAN, K., YU, K., NAKAJIMA, J., AND WANG, W. 2007. How virtualization makes power management different. In *Ottawa Linux Symposium, Proceedings of*.

VANDERSTER, D., BANIASADI, A., AND DIMOPOULOS, N. 2007. Exploiting task temperature profiling in temperature-aware task scheduling for computational clusters. In *Lecture Notes in Computer Science - Advances in Computer Systems Architecture, 2007*. Springer, 10.1007/978-3540-74309-5 18, 175–185.

Panoramic Video Based Augmented Environments

A.   Dalvandi, T. Calvert

School of Interactive Arts and Technology, Simon Fraser University Surrey, Central City

250 - 13450 102nd Avenue

Surrey, BC V3T 0A3 CANADA

Abstract- **Conventionally virtual environments (VE's) are created by literally building a computer model of every object in the environment. Creating a realistic VE comprising the objects found in a street scene, for example, can involve a lot of detailed modeling work and requires a high speed computer and display system to support real-time animation of multiple moving objects. Instead of using a VE with synthetic objects, the environment could also be created with panoramic still images captured from a real environment, and augmenting these scenes with virtual objects like human avatars. These augmented VE's can work very well but become complicated if there are moving objects in the scene and if the point of observation is moving.**

Our interest focuses on urban street environments and we are actively evaluating the use of panoramic video for creating augmented environments for simulating pedestrian navigation. However, the obvious limitation of this approach is that movement through the environment is restricted to a prerecorded prescribed path. A major question relates to the granularity of directional choices that is necessary to maintain immersion and suspension of disbelief in such a virtual environment.

Index Terms— **Augmented Environments, Panoramic Video, Sense of Presence, Virtual Reality**

I. INTRODUCTION AND RELATED WORKS

Virtual Environments (VE's) are increasingly used to represent real or imaginary scenes in applications that include games, virtual worlds, education and training etc. Traditionally VE's are created by literally building a computer model of every object in the environment [1] that includes static solid objects (buildings, etc), moving solid objects (automobiles, etc), human avatars and other phenomena such as smoke, clouds, fire, etc. VE's comprising these objects can be quite realistic and convincing. However, real world scenes contain enormously rich details which need laborious modeling and expensive special purpose rendering hardware. Therefore, creating a realistic VE of a street scene, for example, can be very difficult and expensive.

Not surprisingly, a number of researchers have suggested that instead of using a VE with synthetic objects much of the environment could be created with panoramic images or video of a real environment [2][3]. Those images or videos can then be augmented with virtual objects or human avatars. In an early example of image-based environments [4], sequences of images were taken by single frame cameras at 10-foot intervals in the streets of Aspen, Colorado.

To simulate the effects of walking on the streets and to allow users to deviate from the straight path, separate sequences of all the possible turns through all the intersections were filmed. Two videodisk players then displayed this movie at a rate controlled by the user.

In another work, Chen [5] used 360-degree cylindrical panoramic images of a real environment to compose a virtual environment and connected a number of these images to form a walkthrough sequence. Walking in this VE is accomplished by "hopping" to different panoramic points. The use of orientation-independent images allows a greater degree of freedom in interactive viewing and navigation. However, using still images and moving in discontinuous consecutive hops creates a lack of realism in the interaction experienced by the user.

Miller et al [6] took a similar approach to create a Virtual Museum. In this example, at selected points in the museum, a 360-degree panning movie was rendered to let the user look around. Walking from one of the points to another was simulated with a bi-directional transition movie, which contained a frame for each step in both directions along the path connecting the two points.

Quite recently Sato, et al. [7] used an omnidirectional camera system mounted on a car moving at a constant speed to obtain panoramic images of streets at 15 fps. After geometrical calibrations, the movie is projected to a 3-screen display to provide the user's view to the front, right, and left. A treadmill is used as the locomotion interface to detect a user's locomotion speed and the scene in the recorded movies is appropriately changed according to the user's walking.

Many examples of VE's are created with panoramic images and regular or panoramic video [8][9][10]. Most of these VE's can work very well if there are no moving objects in the scene or if the point of observation does not move. If there are moving objects, the panoramic image can be replaced by panoramic video ([11][12]).

II. PROBLEM

Depending on the level of interactivity to be provided, creating a video based VE can range from being

very simple to very difficult. The simplest situation involves recording a linear video along a specific path at a specific speed, while the most difficult model to implement is a VE which gives the user the ability to navigate in any direction and at any desired speed.

While changing the sampling rate of video playback can provide a range of different navigation speeds, the main question relates to the granularity of the directional choices that are necessary to maintain suspension of disbelief in users.

In real street environments pedestrians can jay walk at any point along their way and at any angle from 0° to 360°. In a virtual street environment, it is impossible to prerecord all possible pathways. Therefore, jay walking has to be restricted to a number n of different directions at angles of $\alpha_1, \alpha_2, \alpha_3, ...\alpha_n$ to the street, and it can only be initiated every d meters. It needs to be determined by experiment what are reasonable values for n and d. It is possible, for example, that a lack of choice in where to initiate a jaywalk is much more upsetting than a lack of choice in the angles at which jaywalking can be carried out.

The problem addressed here is to investigate whether currently available technology will support a system where the point of observation can move at a useful range of speeds and in a useful variety of directions. Specifically, we are initially investigating a VE of urban streets, which puts some limitations on movement.

III. CREATING PANORAMIC VIEW OF THE STREET SCENES

In our simplified VE, a pedestrian must walk on a sidewalk (available on both sides of the street) in a direction parallel to the direction of the street or cross the street at a cross-walk in a direction perpendicular to the direction of the street. At any intersection pedestrians can stop and decide on their route. Then, they can choose any of the intersecting paths/streets to continue their navigation. This idea is inspired by the natural and regular way pedestrians walk in real streets [1] and will be set up first and tested experimentally for acceptability.

To create a VE that gives these abilities to users, the following steps are undertaken. A system comprising 8 regular cameras connected to a PC has been designed and implemented so that the combined fields of view of the cameras cover the whole 360-degree horizontal field of view. Cameras are mounted on a trolley moving along the streets at the regular speed of a pedestrian (about 1 m/s) and video is captured at 30 fps.

A software system takes the digital video output from the multiple cameras, separates this into individual frames, stitches the synchronized frames into composite panoramas, sequences panorama frames and creates the panoramic video.

To create the street scene panoramic video we moved the trolley on which the cameras are mounted, along each street on both sidewalks and in both directions. At intersections, we captured video of crossing all streets by all possible crosswalks in addition to the video of
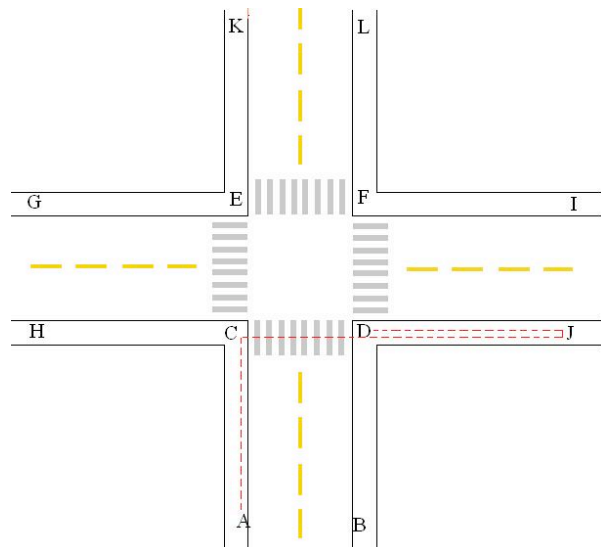

FIG. 1. Video capture paths and an example of a pedestrian's navigation route in the VE

walking along all the intersecting streets. In order to allow users to have some time to decide on the direction of their route, the trolley stopped at the intersections and cameras captured video of the stationary situation for about 20 seconds. This video is then looped to give users as much time as they desire to stand and look around without interrupting the video playback.

As illustrated in an example map in Figure 1, video of all of the AC, BD, CE, CD, DF, EF, EG, CH, FI, DJ, EK, FL paths is captured in both directions. Videos of stationary situations are also captured at points C, D, E, and F. An example of a navigated path chosen by a user can be to start from A to C, to stand for a while at C, to cross the street to D, then to continue to J and then come back to D.

IV. BUILDING AN AUGMENTED ENVIRONMENT USING PANORAMIC VIDEO

To increase the level of interactivity that our VE presents, we augment it by adding computer-generated virtual objects, especially human figures. We are using a game engine as a powerful tool for mixing video panoramas with 3D modeled objects and avatars. This augmented environment has the potential for use in games and many other interactive and simulation applications.

To build an augmented world, panoramic video is mapped onto a cylindrical wall surrounding a virtual navigator located at the center of the cylinder. Other objects and avatars are also placed inside the cylinder and in front of the panorama. Consequently, the panoramic video acts as the background texture of the environment providing a realistic view and all the details available in a real street scene.

The user sees the VE from the virtual navigator's perspective and can interact with the Augmented VE by looking around, navigating in the panoramic video space and interacting with 3D models. We simulated movement

in this environment by displaying video at the frame rate specified by the user's navigation speed. 3D objects are also programmed to move, appear or disappear, and respond to the user's actions according to the situation.

V. CONCLUSION

Panoramic video based Augmented Reality Environments (ARE's) have the potential to allow us to create realistic and inexpensive environments. They allow inclusion of very detailed backgrounds and objects that are difficult to create in conventional VE's. Our experiments will show how effective and efficient these VE's are and in particular what are the limitations on relatively unstructured pedestrian navigation.

Subjects in our experiments are asked to stand in front of a large screen and navigate the ARE by using a controller and a map of the environment. The controller allows them to navigate in different directions (at the designated points) and at different speeds. Subjects will be asked to complete a navigation task by following a route to a specified target. In terms of their navigation, they will be able to look around, go forward along the sidewalks, turn into the crossing streets and cross at the pedestrians cross walks. Initially they are asked to follow sidewalks and designated crosswalks. Then they repeat the navigation using jaywalking where they wish.

Three methods of data collection are used: think aloud, a presence questionnaire, and a post interview. These are employed to gather information about subjects' sense of presence, their general satisfaction with the navigation and their ability to choose a desired navigation route and speed. The experiments are actively underway to demonstrate how well our augmented VE supports the sense of presence and immersion in human subjects.

For future studies, a major objective is to explore how dense the prerecorded routes must be if the subjects are to maintain immersion and sense of presence. The number of choices users can make for the directions and speed of navigation or the best points for street crossings are among the important parameters we intend to explore.

REFERENCES

[1] A. Park, T. Calvert, P. L. Brantingham, P. J. Brantingham, "The use of virtual and mixed reality environments for urban behavioural studies", *PsychNology Journal,* 6(2), pp. 119 – 130, 2008.

[2] P. Liu, X. Sun, N.D. Georganas, E. Dubois, "Augmented reality: a novel approach for navigating in panorama-based virtual environments (PBVE)", *2nd IEEE International Workshop on Haptic Audio and Visual Environments and Their Applications, HAVE 2003 Proceedings*, pp. 13- 18, 2003.

[3] S. Dasgupta, A. Banerjee, "An augmented-reality-based real-time panoramic vision system for autonomous navigation", *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, pp. 154 – 161, 2006.

[4] A. Lippman, "Movie-maps: An application of the optical videodisc to computer graphics", *Proceedings of the 7th annual conference on Computer graphics and interactive techniques*, ACM, Seattle USA, 1980, pp. 32–42, 1980.

[5] S. E. Chen, "Quicktime VR – an image-based approach to virtual environment navigation", *SIGGRAPH '95: Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*, ACM, New York USA, pp. 29-38, 1995.

[6] G. Miller, E. Hoffert, S. E. Chen, E. Patterson, D. Blackketter, S. Rubin, S. A. Applin, D. Yim, J. Hanan, "The Virtual Museum: Interactive 3D Navigation of a Multimedia Database", *The Journal of Visualization and Computer Animation*, pp.193-196, 1992.

[7] T. Sato, M. Kanbara, N. Yokoya, S. Ikeda, "An Immersive Telepresence System with a Locomotion Interface Using High-Resolution Omnidirectional Movies", *17th International Conference on Pattern Recognition*, IEEE Computer Society Press, Los Alamitos USA, pp.396-399, 2004.

[8] D. Southwell, A. Basu, M. Fiala, and J. Reyda, "Panoramic Stereo", *International Conference on Pattern Recognition*, volume A, IEEE Computer Society Press, Vienna Austria, pp. 378–382, 1996.

[9] R. Szeliski, H. Y. Shum, "Creating full view panoramic image mosaics and environment maps", *SIGGRAPH '97: Proceedings of the 24th annual conference on Computer graphics and interactive techniques*, ACM Press/Addison-Wesley Publishing Co. , New York USA, pp.251-258, 1997.

[10] A. Agarwala, C. Zheng, C. Pal, M. Agrawala, M. Cohen, B. Curless, D. Salesin, R. Szeliski, "Panoramic Video Textures" *ACM Transactions on Graphics (Proceedings of SIGGRAPH 2005)*, ACM, New York USA, pp. 821- 827, 2005.

[11] http://www.immersivemedia.com/#demo18

[12]http://gardengnomesoftware.com/pano2vr_sample.php?demo=video

Arefe Dalvandi is a master student in the school of Interactive Arts and Technology (SIAT) at Simon Fraser University, Canada. She has got her bachelor degree in Information Technology in Sharif University of Technology, Iran. She is currently actively working on Video Based Virtual Environments under the supervision of Tom Calvert in SIAT.

Generating Application Specific SIMPPL Controllers

Jian Li, *Simon Fraser University,* Lesley Shannon, *Member, IEEE*

Abstract—The rapid development of semiconductor technology makes it possible to integrate an entire electronic system onto one single chip — System-on-Chip (SoC). However, the effort and time devoted to the design process increases significantly with the exponential growth of the SoC's complexity. By reusing previously designed Intellectual Property (IP) Modules, SoC designers can shorten the SoC design cycles. SIMPPL (Systems Integrating Modules with Predefined Physical Links) is an SoC design model that facilitates the design process by leveraging the reuse of existing IPs. SIMPPL controllers are lightweight micro-controllers used within the SIMPPL framework to serve as interfaces to IP modules.

In this paper, we present an automated generator for quickly customizing the SIMPPL controller. It enables SoC designers to generate different SIMPPL controllers with customized instruction sets, registers and status bits, while automatically optimizing the area and performance of the controllers to meet the requirement of the system. We tested the auto generated controller with Xilinx's ISE^{TM} (Integrated Simulation Environment) simulator and the results show an average improvement of 42% in area reduction and about 16% average increase in maximum frequency over the original controller.

Index Terms—SIMPPL, design reuse, instruction set, design automation.

I. Introduction

THE advances in semiconductor technologies have now enabled SoC designers to build complex digital systems with an immense scale, which in turn demand SoC designers to increase their productivity in order to keep up with the pace. SIMPPL (Systems Integrating Modules with Predefined Physical Links) is a framework for reusing existing IP modules for the purpose of shortening SoC design time. SIMPPL facilitates the implementation SoC for both ASIC and FPGA platforms.

SIMPPL separates the data-path of IP modules from control logic by using SIMPPL controllers as interfaces. The SIMPPL controller is a lightweight micro-controller with a simple instruction set for integrating the IP modules into the system. The SIMPPL framework simplifies the re-design of IP modules for specific applications and expedites system integration [1].

The SIMPPL controller has an instruction set aimed at providing sufficient abilities for the IP cores to be reconfigured to various potential applications. In this paper, we present a toolset to further customize the SIMPPL instruction set for optimized area and speed. We use the ISE^{TM}simulator [2] to

J. Li and L. Shannon are with the Faculty of Applied Science, Simon Fraser University, Burnaby, V5A 1S6, Canada
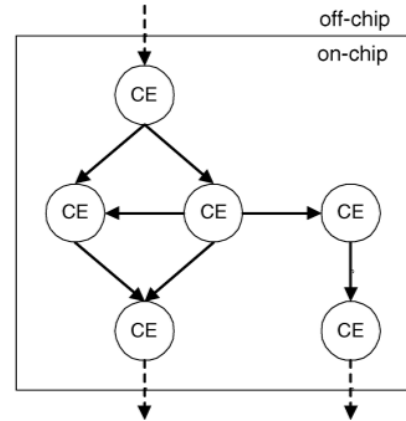

Fig. 1.   CE abstraction of the SIMPPL framework.

estimate the device area cost and maximum clock frequency of the generated controller with a customized instruction set, and compare those to the original SIMPPL controller.

The remainder of our paper is structured as follows: Section II briefly summarizes the SIMPPL framework and the motivation of the SIMPPL generator, Section III discusses the SIMPPL generator in detail, Section IV outlines the experimental setup, Section V provides the results for the experiment, and finally, Section VI concludes the paper and discusses future extensions.

II. Background

This section first briefly discusses of the SIMPPL framework, then describes our motivation for developing the controller generator.

A. The SIMPPL Framework

SIMPPL defines a new abstraction for SoC as a collection of CEs (Computing Element) with interconnections to perform specific functions [1]. Figure 1 shows a typical SoC abstraction using the SIMPPL framework. CEs are connected to each other through communication links (usually asynchronous First-In-First-Out [3]). As denoted in Figure 2, a CE comprises three elements, a SIMPPL controller, a SIMPPL controller sequencer (SCS), and an IP core (denoted as Programming Element (PE) from now on). The abstraction of CE decouples the IP data path (PE) from the control logic and the internal system communication links, thus simplifying IP reuse.
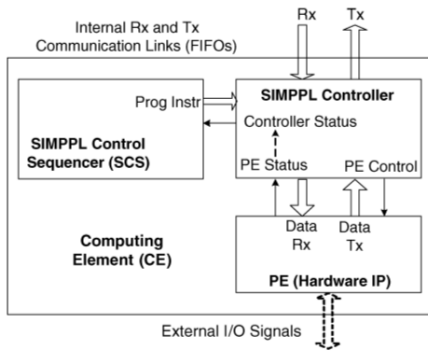
Fig. 2. Elements of a CE.

The instruction set of the SIMPPL controller consists of twelve instructions, each of which performs a distinct function that briefly falls into one of the two categories: instructions for control operations and those for data transfer. However, typical applications for implementing the SIMPPL controllers usually require a subset of the twelve instructions; because, in most cases, each PE (IP module) is aimed at accomplishing specific tasks, and its adaptation to the SIMPPL framework only has simple communication and control demands. In other words, the SIMPPL controllers for typical applications usually contains "redundant" logic for unnecessary instructions, which means potential benefits in logic optimization exist if such redundancy can be "squeezed out".

B. Motivation for Customizable Instruction Set

The motivation for our work could be summarized as follows, (a) allow user to specify instruction set, (b) automatically generate customized SIMPPL controllers, (c) reduce area and increase clock frequency by removing unnecessary logic. Because the current instruction set keeps SIMPPL controllers adaptable to a wide variety of PEs, we intend to introduce the feature of customization without re-defining the existing instruction set.

C. Possible Extensions

We hope to extend instruction customization such that instructions with unique requirements could be "generated on the fly" without introducing significant design penalties, such as an increase in design time and effort or a deterioration in area and/or performance.

III. THE SIMPPL GENERATOR

The details for the SIMPPL generator are discussed in this section.

A. Basics about the SIMPPL Controller

The current instruction sets of SIMPPL controllers can be divided into six groups,

(a) Immediate Data Transfer
(b) Immediate Data Transfer + Immediate Address

(c) Immediate Data Transfer + Indirect Addressing
(d) Immediate Data Transfer + Auto-increment
(e) Register Operations (initialization, arithmetic)
(f) Miscellaneous (reset, no-op, bypass)

Each instruction has a unique 8-bit binary code (opcode), with each bit or the combination of bits representing the different internal functions of the instruction [1].

For the purpose of reducing the controller's complexity, the SIMPPL controller is designed as a "single issue" controller, meaning that only one instruction is allowed to be "on the fly" (in any of the three stages) at any given time.

The SIMPPL controller is also a three-stage micro-controller — the three stage being "fetch", "decode" and "execute". In the "fetch" stage, the controller first fetches one instruction from either the local memory or the receive-link (the FIFO [3] shown in Figure 1); then in the "decode" stage, the controller interprets the instructions and translate them into the a combination of flags; and finally in the "execute" stage, the corresponding actions are carried out according to the flags' combination, signaling a successful instruction execution.

The pseudo code in Figure 3 shows the "decode" and "execute" stage of the controller (marked as "Decode Block" and "Execution Block" respectively).

```
/* Decode Block */
if (opcode == INST0) {
  flag_0 <= 1;
  flag_1 <= 1;
    ...
  flag_n <= 0;
}
 ...

/* Execution Block */
if (flag_0 == 1) {
  action_A; clear flag_0;
}
if (flag_1 == 0) {
  action_B; clear flag_1;
}
...
```

Fig. 3. Pseudo Code for Decoding and Executing SIMPPL instructions.

B. How the SIMPPL Generator Works

By detecting the specified decoding and executing logic of the target instructions, the SIMPPL generator is able to perform logic optimization for the new controller (e.g., re-organizing the decoding logic for flag signals and/or removing unnecessary execute blocks).

Figure 4 shows the work flow of the SIMPPL generator. The SIMPPL generator starts by taking a list user-specified instructions as its input file, then detects the corresponding "Decode Block" and "Execution Block" in the *original* con-
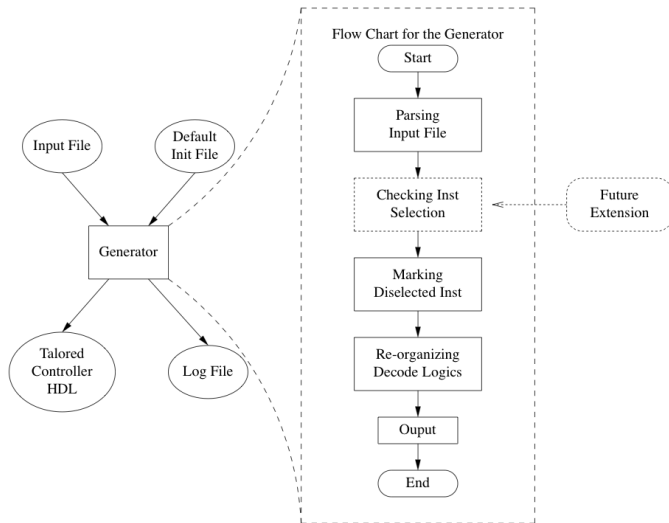
Fig. 4.   Work-flow and Algorithm Flowchart for Generator

Fig. 5.   LUT Usage

Fig. 6.   FF Usage

troller HDL code, and finally re-organizes the decoding logic of the customized controller.

The generator assumes that the correctness of the given instruction combination has been checked by the user, and enforces no sanity check of the combination at this moment. Given the lightweight and simple nature of the controller and its instruction set, we believe that it won't impose significant extra burden for instruction checking, and we plan to include this feature in future extensions.

IV. EXPERIMENTAL SETUP

We simulated the generated versions of the SIMPPL controllers with a customized instruction set using the Modelsim^{TM} simulator from Synopsis [4]. The test cases covered typical SIMPPL applications, including computation-centric PEs such as FFT transforms, and control-centric PEs such as an normal arbiter.

We obtained the area and maximum clock frequency for each customized controller in the ISE^{TM} (Integrated Simulation Environment) from Xilinx. The results are compared to those of the original controller.

The targeted FPGA platform is the Xilinx Virtex4 LX40-12 because it is the highest speed grade device of 90nm fabrication from Xilinx, and the ISE^{TM} suite version is 10.1.3.

V. RESULTS AND ANALYSIS

We chose to compare the original controller with full instruction support to a customized controller which only supports immediate data transfer, because instructions for immediate data transfer are common for most SIMPPL applications.

The device area is measured with the Look-Up-Table (LUT) usage and Flip-Flop (FF) usage on the FPGA platform. We performed tests on three kinds of SIMPPL controllers, namely the full controller, the producer and the consumer. As suggested by their names, the full controller supports all
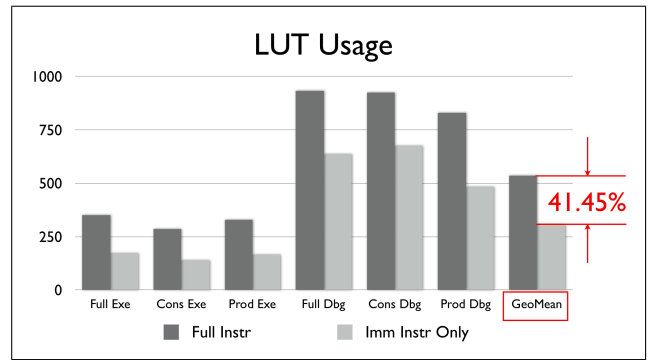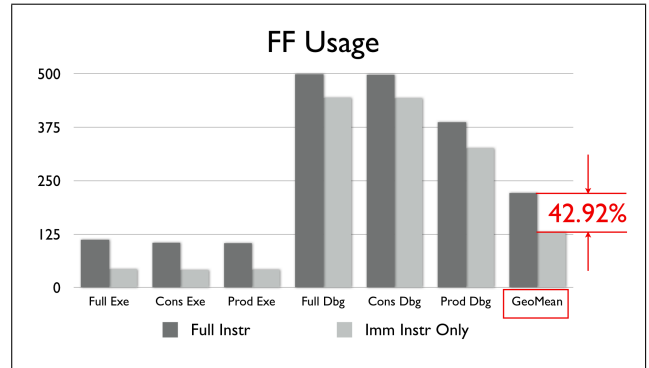
the SIMPPL instructions, whereas the producer supports only instructions for producing data and the consumer supports only instructions for consuming data. It should be noted that SIMPPL controllers come with an optional built-in debug infrastructure for uploading debugging information for designers to monitor the controllers' internal status.

We present the simulated results for the three kinds controllers with the optional debug feature (denoted as "dbg") and without the debug feature (denoted as "exe"), which results in a combination of six groups. As shown in Figure 5, the bars with darker grey represent the six combinations of the original full controllers, whereas the bars with lighter grey represent the corresponding customized version that only supports immediate data transfer instructions. Geometric means of the previous six groups for both the original controller and the customized version are presented in the rightmost group of the figure, marked with the percentile difference. Similarly for Figure 6 and 7.

In summary, we achieved an average area reduction by 41.45% in LUT usage and by 42.92% in FF usage, meanwhile we obtained an average system performance upgrade in maximum clock frequency by 15.76%.

VI. CONCLUSIONS AND FUTURE WORK

In this paper, we present an automated HDL generator for customizing application specific SIMPPL controllers. The generator provides the ability to quickly optimize performance for SIMPPL controllers in terms of area and maximum clock
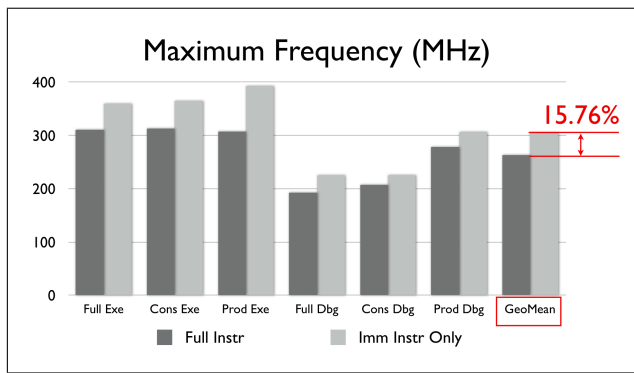
Fig. 7.   Maximum Clock Frequency

Jian Li received his B.Sc. degree in electronic engineering from Zhejiang University, Hangzhou, China, in 2005. He then worked as an electronic system engineer in Renesas Semiconductor Design, Suzhou, China until 2008.

In September of 2008, he joined the Reconfigurable Computing Lab of the School of Engineering Science, Simon Fraser University, Burnaby, BC, Canada, as a M.A.Sc. student. His research interests include reconfigurable system design, system-on-chip architectures and IC CAD tools.

frequency, which further expedites the system integration of SoCs by using the SIMPPL framework.

We tested the functionality of the generated controller in typical use cases with the Modelsim$^{\text{TM}}$ simulator. The area consumption and maximum clock speed of the generated controller is obtained using the Xilinx ISE$^{\text{TM}}$ suite. A comparison between the customized controller and the original controller shows an average area reduction of about 42% and an average speed increase by 15.76%.

We plan to extend the functionality of the generator to include a sanity checker for enforcing instruction correctness, and we also hope to increase the flexibility of the logic reorganizer of generator ton provide more freedom for the user in instruction customization.

Lesley Shannon (S'03–M'06) received the B.Sc.Eng. degree in electrical engineering from the University of New Brunswick, Fredericton, NB, Canada, in 1999, and the M.A.Sc. and Ph.D. degrees from the University of Toronto, Toronto, ON, Canada, in 2001 and 2006, respectively.

In 2006, she joined the School of Engineering Science, Simon Fraser University, Burnaby, BC, Canada, as an Assistant Professor. Her current research interests include computing system design, particularly, system-on-chip architectures, embedded computing systems, and reconfigurable computing, as well as computing system design methodologies and CAD tools.

ACKNOWLEDGMENT

REFERENCES

[1] L. Shannon and P. Chow, "SIMPPL: An adaptable SoC framework using a programmable controller IP interface to facilitate design reuse," *IEEE Transactions on Very Large Scale Integration Systems*, vol. 15, no. 4, p. 377, 2007.

[2] I. Xilinx, "Software Manuals and Help," *Xilinx.[J/OL]. http: www. xilinx. com/support/sw _ manuals/xilinx6/index. htm*.

[3] Xilinx, "Fast simplex link (fsl) bus v2.11a," *Document DS449*, June 25, 2007.

[4] S. ModelSim, "users manual, Mentor Graphics Inc."

Active Vibration Control of a Flexible Cantilever Beam with a Translating Base using Piezoelectric Material

Masih A. Hosseini

Abstract—**Study of light weight flexible cantilever beams as a benchmark, helps engineers in field of design and control to understand the vibrational properties of machine elements in rotating machineries, aerospace structures, light weight robots, etc. In this paper modeling, simulation and control of a cantilever smart beam attached to a translating base is studied. A plant model is developed and a simple vibration control method is proposed. Simulations demonstrate that the method is effective in suppressing the tip vibrations and can keep the beam flat.**

Index Terms—**Smart Material, Vibration, Modeling, Simulation, Piezoelectric.**

I. Introduction

LIGHT weight flexible structures might experience harsh vibrations due to dynamic loadings. Flexible beams, attached to a support body, draw attention in different fields such as design of robotic arms, aircraft propeller and helicopter rotor blade dynamics, satellite truss and solar panels, rotating machineries, etc. These structures typically suffer from long vibration settling times. Generally, vibrations are undesirable as they cause unpleasant noise, unwanted structural stresses, and malfunctions or system failures. This fact makes vibration control an attractive and challenging problem to be solved both for industry and academia.

Using smart materials (e.g. piezoelectric ceramics) in these structures, helps us design light-weight, intelligent, and controllable structures with regulated and desirable behaviour. Actuators from piezoceramic materials have many applications ranging from active vibration control to nano-scale positioning tasks. The reason that piezocermic materials are used in many applications lies in their high-frequency response behaviour and their infinite resolution [17], [8]. Sensors and actuators built based on piezo technology, typically posses other favorable design characteristics such as: (1) small volume, (2) mechanical simplicity, (3) high strain sensitivity, (4) high stiffness, (5) low noise, (6) good linearity, (7) temperature insensitivity, (8) ease of implementation, (9) low power consumption, (10) low weight, (11) large bandwidth, (12) efficient conversion between electrical energy and mechanical energy, and (13) easy integration with various metallic and composite structures. Piezoelectric materials are used for dynamic stability applications, such as resonant vibrations, active vibration

M.A. Hosseini is a graduate student at Vibration and Smart Materials Lab in Engineering Science Department of Simon Fraser University, British Columbia, Canada. E-mail: masih_hosseini@sfu.ca.

suppression and dynamic buckling, because they have bandwidths beyond the frequency range of structural and acoustic control applications [15].

This project targets to suppress the base-excitation induced vibrations of a long, thin and flexible metallic cantilever beam. The clamped end of the beam is attached to a shaker in a way that normal to the plate is set parallel to the horizon, so that the plate would not be deflected by the gravity force loading. Two piezoelectric patches are attached to the top and bottom surface of the plate, close to the clamped side. The shaker's lateral movements excite the beam and can induce different mode shapes. The shaker excites the beam and induces the 1st vibrational mode. We derived the governing equations of motion and proposed a method to control the residual vibration of the tip and the curvature. The control objective is to force the tip of the beam to follow its base motion and thus suppress the vibration. The chosen control method uses piezoelectric patches as actuators. Beam deflection is sensed using a strain gauge which is placed near the piezoelectric patches.

II. Literature Review

Initially, De Luis at MIT [5] proposed the piezoelectric materials to be built in laminated beams. Fanson and Caughey at CalTech [6], [7] performed feedback control to suppress structural vibration with piezoelectric actuators and sensors. They introduced the positive position feedback (PPF) control scheme to control the vibrating structure. Newman [13] used the strain rate feedback (SRF) for active damping of a flexible space structure. Sung and Chen [18] proposed a method to suppress elastodynamic responses of flexible linkages in high-speed motion by using state feedback optimal control and using several pairs of suitably located piezoelectric sensors and actuators.

In PPF strategy, the structural position is directly fed to the compensator and a scalar gain is positively fed back to the structure. PPF provides quick damping for a particular known modal frequency and it is easy to implement. The insensitivity of PPF to varying modal frequency is experimentally validated [16].

In SRF strategy, the structural velocity is fed back to the compensator and the compensator position coordinate multiplied by a negative gain is fed to the structure. SRF has a wider active damping region and it can stabilize more than one mode if it has sufficient bandwidth [16].

Song *et. al.* [15] investigated the vibration control of a composite beam using piezoelectric patches as sensors and

actuators. Theoretical, experimental, and numerical (FEA) methods are used to find the fundamental modal frequencies and shapes. Based on the concluded model, SRF and PPF control schemes have been implemented for the single-mode and multi-mode vibration suppression.

Using Euler-Bernoulli beam theory, natural frequencies of the beam can be approximated. Many approximate models are developed to describe the behavior of a flexible beam [8]. Choi and Lee [2] presented the derivation for the modeling of a beam with a piezoceramic actuator affixed near the base.

Song *et. al.* [17] designed a modular control patch (MCP) using smart material and a miniaturized digital controller. Different control algorithms for vibration suppression of a aluminium cantilever beam has been tested using MPC. Different combinations of PPF and SRF were used to suppress multiple mode vibration problem. Their work showed that PPF is most effective for single-mode vibration suppression and two PPF filters in parallel are most effective for multimode vibration suppression.

Fei [9] presents results on active control schemes for vibration suppression of flexible steel cantilever beam with bonded PZT (lead zirconate titanate) piezoelectric actuators. He derived an analytical model and investigated different active vibration control methods such as optimized parameter PID and strain rate feedback control.

Dadfarnia *et. al.* [4], [3], [12] proposed a method for regulating of a Cartesian robot arm which was modeled as a flexible beam with translating beam and a mass at the tip. A model of the the system was developed using Lagrange and assumed mode expansion method. A PD controller was used for the base control and a Lyaponuv-based controller was utilized to suppress the residual vibration.

Halim and Moheimani [10] introduced a class of resonant controllers that can be used to minimize structural vibration using collocated piezoelectric actuator-sensor pairs. Barczyk and Lynch [1] experimentally validated a closed-loop flatness-based tracking control of a rotating flexible beam using a boundary-input Euler-Bernoulli partial differential equation.

III. BACKGROUND AND SYSTEM MODELLING

Continuous systems also known as distributed parameter or infinite dimensional systems, are a class of multi-degrees of freedom vibration problems where the vibrating points are not considered as lumped masses but rather as continuous elements, spatially defined by continuous function of relative position.

In this vibration problem a continuous vibration of a beam perpendicular to its length is considered; this vibration is often called transverse or flexural vibration. Equation of motion for a beam subjected to external forcing functions based on Euler-Bernoulli beam theory is

$$\rho A(x)\frac{\partial^2 w(x,t)}{\partial t^2} + \frac{\partial^2}{\partial x^2}\left[EI(x)\frac{\partial^2 w(x,t)}{\partial x^2}\right] = f(x,t), \quad (1)$$

where $A(x)$ is the cross sectional area of the beam and is constant in our problem. ρ is the density of the beam. $w(x,t)$ is the displacement function indicating the deflection from the neutral axis along the length of the beam. E and I are the

Young's modulus and moment of inertia respectively. $f(x,t)$ is the external forcing function which is distributed on the beam.

This general form of equation should be adapted to the problem being considered in this study. To solve the equation of motion (1) using the separation of variables method, four boundary conditions and two initial conditions are required. Boundary conditions to solve the spatial equation for a cantilever beam are found by examining the deflection $w(x,t)$, the slope of deflection $\partial w(x,t)/\partial x$, the bending moment $EI(x)\partial^2 w(x,t)/\partial x^2$, and the shear force $\partial\left[EI(x)\partial^2 w(x,t)/\partial x^2\right]/\partial x$ at each end of the beam [11]. For a cantilever beam (clamped-free), at the clamped end, where the beam is fixed, deflection and slope are zero (equations (2) and (3)).

$$deflection = w(L,t) = 0 \tag{2}$$

$$slope = \frac{\partial w(L,t)}{\partial x} = 0 \tag{3}$$

At free-end the shear forces and bending moments are zero.

$$moment = EI\frac{\partial^2 w(L,t)}{\partial x^2} = 0 \tag{4}$$

$$shear = \frac{\partial}{\partial x}\left[EI\frac{\partial^2 w(L,t)}{\partial x^2}\right] = 0 \tag{5}$$

The general form of the solution to the Euler-Bernoulli beam problem can be found in standard tables (or by solving for the boundary conditions). By solving the Euler equation (1) of a beam with fixed-free end-condition (3,2, 4, 5), the first three roots of the equation $(\lambda_i l)$ is found as 4.6940, 7.8547 and 2.3650. l is the length of the beam. The corresponding natural frequencies are found by plugging in the previous λ_i values in $\omega_i = \lambda_i^2\sqrt{EI/\rho A}$. The mode shape for this special case is found using the following equation:

$$\begin{aligned}\phi_i(x) &= a\sin\lambda_i x + b\sinh\lambda_i x \\ &\quad + \alpha_i\left[c\cos\lambda_i x + d\cosh\lambda_i x\right],\end{aligned} \tag{6}$$

and

$$\alpha_i = \frac{\sinh\lambda_i l + \sin\lambda_i l}{\cosh\lambda_i l + \cos\lambda_i l}. \tag{7}$$

In equation (6) the coefficients are ($a = 1$, $b = -1$, $c = -1$, and $d = 1$). Fig. 1 shows the schematic of our problem with some important notations on it. We can find the equation of motion of the beam by mode-summation method [19]. To use the mode expansion method, the deflection function should be expressed as

$$w(x,t) = \sum_i \phi_i(x)q_i(t), \tag{8}$$

where ϕ_i is the mode shape and $q_i(t)$ is the generalized coordinate. Using equation (8) and establishing the kinetic and potential energies of the system, Lagrange equation can be utilized to obtain the equation of motion of the system in terms of the general coordinate, q_i. From the kinetic energy equation
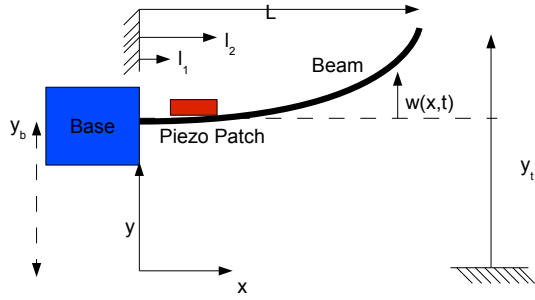
Fig. 1. Beam with piezoelectric patch and a moving base.

and the modal orthogonality relation, the kinetic energy can be expressed as

$$
\begin{aligned}
T &= \frac{1}{2}\int_0^L \dot{w}^2(x,t)\rho A\, dx \\
&= \frac{1}{2}\sum_i\sum_j \dot{q}_i\dot{q}_j \int_0^L \phi_i\phi_j\rho A\, dx \\
&= \frac{1}{2}\sum_i M_i\dot{q}_i^2,
\end{aligned}
\tag{9}
$$

where the generalized mass M_i is defined as

$$
M_i = \int_0^L \phi_i^2(x)\rho A\, dx.
\tag{10}
$$

Similarly, the potential energy is

$$
\begin{aligned}
U &= \frac{1}{2}\int_0^L EI w''^2(x,t)\, dx \\
&= \frac{1}{2}\sum_i\sum_j q_i q_j \int_0^L EI\phi_i''\phi_j''\, dx \\
&= \frac{1}{2}\sum_i K_i q_i^2.
\end{aligned}
\tag{11}
$$

In equation (11) the generalized stiffness is

$$
K_i = \int_0^L EI\left[\phi_i''(x)\right]^2\, dx.
\tag{12}
$$

The generalized force Q_i is the work done by the applied force $f(x,t)$ in the virtual displacement δq_i is

$$
\begin{aligned}
\delta W &= \int_0^L f(x,t)\left(\sum_i \phi_i\delta q_i\right) dx \\
&= \sum_i \delta q_i \int_0^L f(x,t)\phi_i\, dx.
\end{aligned}
\tag{13}
$$

The generalized force is

$$
Q_i = \int_0^L f(x,t)\phi_i(x)\, dx.
\tag{14}
$$

Substituting (9), (11) and (14) into Lagrange equation,

$$
\frac{d}{dt}\left(\frac{\partial T}{\partial \dot{q}_i}\right) - \frac{\partial T}{\partial q_i} + \frac{\partial U}{\partial q_i} = Q_i,
\tag{15}
$$

we find the differential equation for $q_i(t)$, which is

$$
\ddot{q}_i + \omega_i^2 q_i = \frac{1}{M_i}\int_0^L f(x,t)\phi_i(x)\, dx.
\tag{16}
$$

If instead of distributed loads, we had a concentrated moment $M(a,t)$, at $x = a$, the virtual work and generalized force for such a load would be

$$
\delta W = M(a,t)\delta w'(a,t) = M(a,t)\sum_i \phi_i'(a)\delta q_i,
\tag{17}
$$

and

$$
Q_i = \frac{\delta W}{\delta q_i} = M(a,t)\phi_i'(a).
\tag{18}
$$

Based on the background that is developed so far, the response of a cantilever beam with a base excitation given as $y_b(t)$, as shown in Fig. 1, can be written by the following differential equation:

$$
\left[EI w''(x,t)\right]'' + \rho A\left[\ddot{y}_b(t) + \ddot{w}(x,t)\right] = 0,
\tag{19}
$$

and can be rearranged to

$$
\left[EI w''(x,t)\right]'' + \rho A\left[\ddot{w}(x,t)\right] = -\rho A\ddot{y}_b(t).
\tag{20}
$$

Comparing (16) and (1) with (20), the equation of motion for the generalized force can be written as:

$$
\ddot{q}_i + \omega_i^2 q_i = -\ddot{y}_b(t)\frac{1}{M_i}\int_0^L \phi_i(x)\, dx.
\tag{21}
$$

In differential equation (21), $\ddot{y}_b$ is known, and ϕ_i is found using (6) and (7). M_i is calculated using (10). Solving this equation needs two initial conditions, $y_b(0)$ and $\dot{y}_b(0)$ and both can be assumed as zero. To control the residual vibrations caused by inertial forces a piezoelectric patch is used as an actuator. The contribution of the piezoelectric patch to the motion of the beam should be found. The beam model is linear and using the superposition property of the linear systems, the contribution of PZT actuator on the beam can be added to the equation of motion in (21).

The distributed actuator is equivalent to adding concentrated moments M_p at the boundaries of the actuator [14], where M_p is:

$$
M_p = -\frac{1}{2}bE^p d_{31}(t_p + t_b)v(t)S(x) = M_{p_0}v(t)S(x),
\tag{22}
$$

where $v(t)$ is the input voltage to the patch and $S(x)$ is a piecewise function as:

$$
S(x) = H(x - l_1) - H(x - l_2).
\tag{23}
$$

Adding the contribution of the actuator force to the beam equation (21), we get the following equation:

$$
\begin{aligned}
\ddot{q}_i + \omega_i^2 q_i ={}& \frac{1}{M_i}\left[\int_{l_1}^{l_2} M_p(x,t)\phi_i'(x)\, dx\right] \\
& -\ddot{y}_b(t)\frac{1}{M_i}\int_0^L \phi_i(x)\, dx,
\end{aligned}
\tag{24}
$$

Properties	Symbol	Value	Unit
Beam thickness	t_b	0.8125	mm
Beam Young's modulus	E	69×10^9	N/m^2
Beam and PZT width	b	20	mm
Beam length	L	300	mm
Beam density	ρ	3960.0	kg/m^3
PZT Young's modulus	E^p	66.47×10^9	N/m^2
PZT length	$l_2 - l_1$	33.655	mm
PZT position on beam	l_1	44.64	mm
PZT thickness	t_p	0.2032	mm

Fig. 2. Block diagram of the beam simulation.

which can be rewritten as:

$$\ddot{q}_i + \omega_i^2 q_i = \frac{M_{p_0} v(t)}{M_i} \left[\int_{l_1}^{l_2} \phi_i'(x)\,dx \right]$$
$$- \ddot{y}_b(t) \frac{1}{M_i} \int_0^L \phi_i(x)\,dx. \qquad (25)$$

The application being studied in this paper is only concerned with the first mode of vibration ($i = 1$), while the other modes are being neglected. The parameters needed for the simulation are found using Matlab programming based on the values in Table. I. From equation (6) and (7) we get the first mode shape, ($i = 1$), as:

$$\phi_1(x) = 0.9819\,[-\cos 15.6467x + \cosh 15.6467x]$$
$$+ \sin 15.6467x - \sinh 15.6467x, \qquad (26)$$

and its derivative is:

$$\phi_1'(x) = 15.6467 \cos 15.6467x$$
$$-15.6467 \cosh 15.6467x$$
$$+15.363 \sin 15.6467x$$
$$+15.363 \sinh 15.6467x. \qquad (27)$$

From (10), the equivalent mass is $M_1 = 0.0186$ (Kg) and $\omega_1^2 = 239.6919$. Finally the equation of motion in terms of the generalized coordinate system can be written as:

$$\ddot{q} + 239.6919q = 0.003v(t) - 6.8691\ddot{y}_b(t). \qquad (28)$$

Let's assume that the system also has a damping (ζ), that can be measured experimentally. Then equation (28) can be rewritten as (29) and it approximately resembles the damped equation of motion as the following:

$$\ddot{q} + 30.9640\zeta\dot{q} + 239.6919q = 0.003v(t) - 6.8691\ddot{y}_b(t). \quad (29)$$

The equation (29) can be written as:

$$\ddot{q} + 2\omega_1\zeta\dot{q} + \omega_1^2 q = G_v v(t) + G_b \ddot{y}_b(t), \qquad (30)$$

in terms of natural frequency and damping.

IV. SYSTEM ANALYSIS AND CONTROL DESIGN

To find the displacement of the tip of the beam the temporal term, $q_1(t)$, should be multiplied by $\phi_1(L)$. Let's multiply both

sides of equation (30) by the spatial term $\phi_1(x)$, or $\phi_1(L)$. The new form of the equation of motion will be:

$$\ddot{w}(x,t) + 2\omega_1\zeta\dot{w}(x,t) + \omega_1^2 w(x,t) =$$
$$G_v v(t)\phi_1(x) + G_b \ddot{y}_b(t)\phi_1(x). \qquad (31)$$

Let's transfer equation (31) from the time domain to the Laplace domain as the following:

$$s^2 W(x,s) + 2\omega_1\zeta s W(x,s) + \omega_1^2 W(x,s) =$$
$$G_v V(s)\phi_1(x) + G_b s^2 Y_b(s)\phi_1(x). \qquad (32)$$

Contribution of the base displacement and piezoelectric voltage on the deflection of the beam can be expressed separately by W_1/Y_b, and W_2/V respectively, and the total displacement can be expressed by $W = W_1 + W_2$. The respective transfer functions are

$$H^*(s) = \frac{W_1}{Y_b} = \frac{G_b s^2 \phi(x)}{s^2 + 2\omega\zeta s + \omega^2}, \qquad (33)$$

and

$$L(s) = \frac{W_2}{V} = \frac{G_v \phi(x)}{s^2 + 2\omega\zeta s + \omega^2}. \qquad (34)$$

For the first mode of vibration we can obtain the absolute displacement of the tip of the beam due to the base excitation as:

$$H(s) = \frac{W_1 + Y_b}{Y_b} = \frac{(G_b\phi(x) + 1)s^2 + 2\omega\zeta s + \omega^2}{s^2 + 2\omega\zeta s + \omega^2}. \quad (35)$$

Assume that by an experiment, we found the modal damping of the first mode as $\zeta = 0.1$. From previous relations we calculate the other coefficients as $G_v = 0.003$, $G_b = -6.8691$, $\phi(x) = \phi(L) = -1.9636$ and $\omega = 15.4820$. The tip displacement in Laplace domain can be expressed by:

$$H(s)Y_b(s) + L(s)V(s) = Y_t(s), \qquad (36)$$

and is depicted schematically in Fig. 2. The control strategy is to make $Y_t(s)$ and $Y_b(s)$ equal. Y_b is the known input of the system and we have no control on it. To have a perfect matching between Y and Y_b a control system should be designed to change the voltage input to the piezoelectric patch. From equation (36) we can see that, for obtaining a perfect control, an input which is proportional to the base motion
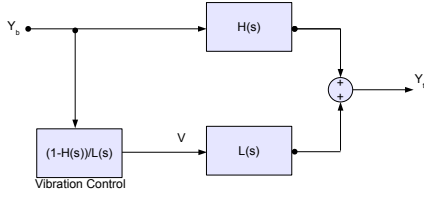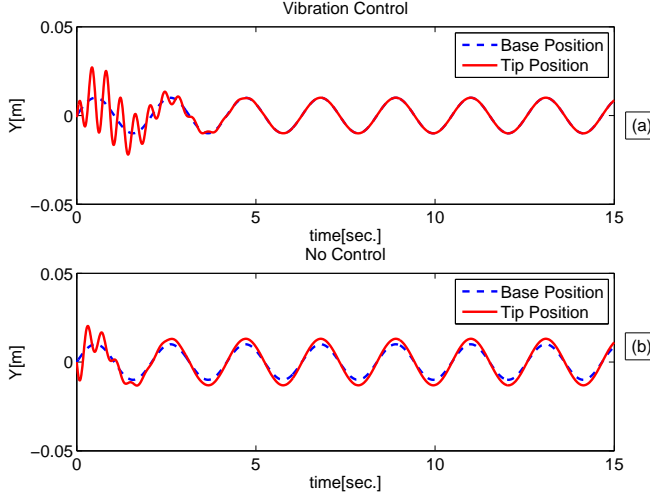
Fig. 3. Block diagram of the vibration control method.

Fig. 4. Base and tip displacement of the beam. (a) with feedforward control, (b) without any control.
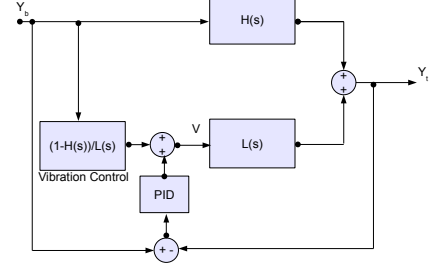

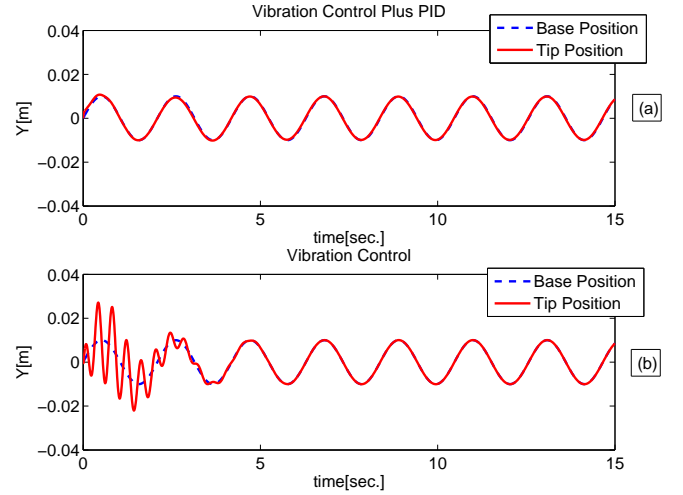Fig. 5. Block diagram of the PID control added to the previous vibration control method.

Fig. 6. Comparison of basic vibration control to vibration control plus PID control. (a) Feedforward vibration control plus PID feedback. (b) Feedforward vibration control.

should be commanded to the actuator. So we can rewrite the equation (36) and set the tip and base displacement equal as the following:

$$H(s)Y_b(s) + G_c(s)L(s)Y_b(s) = Y_b(s), \qquad (37)$$

so that

$$G_c(s)L(s) = 1 - H(s). \qquad (38)$$

That means, the desired controller is:

$$G_c(s) = \frac{1 - H(s)}{L(s)}. \qquad (39)$$

The controller is depicted schematically in Fig. 3.

Plugging in the values, the resulting transfer functions of the controller are:

$$L(s) = \frac{W_2}{V} = \frac{0.0035}{s^2 + 3.1s + 239.7}, \qquad (40)$$

and

$$H(s) = \frac{W_1 + Yb}{Y_b} = \frac{-6.9536s^2 + 3.1s + 239.7}{s^2 + 3.1s + 239.7}. \qquad (41)$$

V. SIMULATIONS AND IMPLEMENTATION

After computing the beam's vibrational property using MATLAB programing, the simulation of the dynamics and control of the beam is performed in Simulink. In Fig. 4, beam vibration control is compared with a case without control. As it is shown in Fig. 4, steady state tracking error of the

tip vanishes but the transient response of the the beam is affected. To cope with the steady state tracking problem a PID control is added to the basic control system and is depicted in Fig. 5. Proportional gain, derivative gain and the integral gain are selected as K_p=10e4, as K_d=10e5 and K_i=10e5 respectively. System response is shown and compared with the basic vibration control strategy in Fig. 6 . Looking at Fig. 6 the transient response is improved particularly. Steady state error of the PID controller is depicted in Fig. 7 and is compared to steady state error of the basic controller. Steady state error of the basic controller is slightly better than the basic controller plus PID scheme but this advantage is negligible. The control effort of the new controller (PID and vibration control) is less than the basic vibration controller. Considering these reasons, the PID control plus vibration control is proposed as a solution to our problem.

To validate the method a test bed is prepared (Fig. 8). Preliminary works for providing a xPC target based rapid prototyping solution is done too. The beam depicted in Fig. 8 has two collocated *ACX mide* patch close to the fixture side. Two strain gauges are attached to the beam to measure the deflections of the beam. The Simulink model needs information about the position/velocity/acceleration of vibrating base
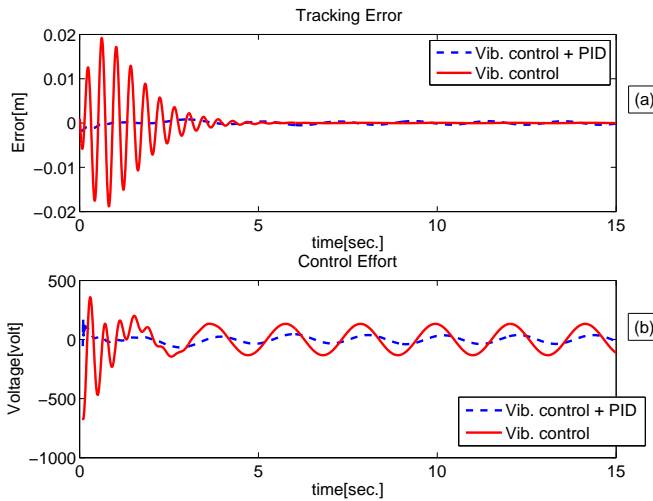
Fig. 7. (a) Comparison of the tracking error of vibration control scheme and vibration control plus PID scheme and control effort. (b) Comparison of the control effort of the two control schemes.
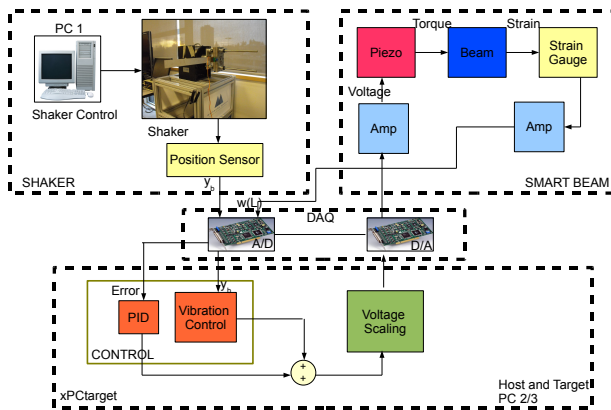

Fig. 8. Schematics of the xPCtarget based smart beam control.

which needs to be fed into the control loop. These information can be read by the accelerometer attached to the shaker. A scheme of the vibration control system is depicted in Fig. 8. Validificaion of the control method is planned and will be addressed in near future.

VI. CONCLUSION

In this research, we derived an analytical model for the first vibration mode of a cantilever beam. A control method capable of damping the residual oscillations of the beam is designed and analyzed. Simulation results show high potentials of the proposed control method in suppressing the vibrations and improving the tracking behavior. Future work will address the experiments and implementation of our method on a test setup.

REFERENCES

[1] M. Barczyk and A.F. Lynch. Flatness-based closed-loop control of a rotating euler-bernoulli beam: experimental results. In *American Control Conference, 2006*, pages 6 pp.–, 2006.

[2] Seung-Bok Choi and Chul-Hee Lee. Force tracking control of a flexible gripper driven by piezoceramic actuators. *Journal of Dynamic Systems, Measurement, and Control*, 119(3):439–446, 1997.

[3] Mohsen Dadfarnia, Nader Jalili, Zeyu Liu, and Darren M. Dawson. An observer-based piezoelectric control of flexible cartesian robot arms: theory and experiment. *Control Engineering Practice*, 12(8):1041–1053, August 2004.

[4] Mohsen Dadfarnia, Nader Jalili, Bin Xian, and Darren M. Dawson. A lyapunov-based piezoelectric controller for flexible cartesian robot manipulators. *Journal of Dynamic Systems, Measurement, and Control*, 126(2):347–358, 2004.

[5] Javier DeLuis. Use of piezoelectric actuators as elements of intelligent structures. *AIAA*, 25(10):1373–1385, 1987.

[6] J. L. Fanson and T. K. Caughey. Positive position feedback control for large space structures. *AIAA*, 1987.

[7] J. L. Fanson and T. K. Caughey. Positive position feedback control for large space structures. *AIAA*, 28:717–24, 1990.

[8] Juntao Fei. Active vibration control of flexible steel cantilever beam using piezoelectric actuators. *System Theory, 2005. SSST '05. Proceedings of the Thirty-Seventh Southeastern Symposium on*, pages 35–39, 2005.

[9] Juntao Fei and Yunmei Fang. Active feedback vibration suppression of a flexible steel cantilever beam using smart materials. In *Innovative Computing, Information and Control, 2006. ICICIC '06. First International Conference on*, volume 1, pages 89–92, 2006.

[10] D. Halim and S.O.R. Moheimani. Spatial resonant control of flexible structures-application to a piezoelectric laminate beam. *Control Systems Technology, IEEE Transactions on*, 9(1):37–53, 2001.

[11] Daniel J. Inman. *Engineering Vibration*. Prentice Hall, New Jersey, 2000.

[12] Nader Jalili and Ebrahim Esmailzadeh. *Vibration Damping, Control, and Design, Editor: Clarence W. de Silva*, chapter 5 on Vibration Control, pages 5–1 to 5–46. CRC Press, 2007.

[13] S.M. Newman. Active damping control of a flexible space structure using piezoelectric sensors and actuators. Master's thesis, U.S. Naval Postgraduate School, 1992.

[14] Andre Premont. *Vibration Control of Active Structures: An Introduction*. Kluwer Academic Publishers, 1997.

[15] G. Song, P. Qiao, and W. K. Binienda. Active vibration damping of a composite beam using smart sensors and actuators. *ASCE Journal of Aerospace Engineering*, 15(3):97–103, July 2002.

[16] G. Song, S. P. Schmidt, and B. N. Agrawal. Experimental study of vibration suppression of flexible structure using modular control patch. *Proc. IEEE Aerospace Conference, Snowmass, Colorado*, 1998.

[17] G. Song, S. P. Schmidt, and B. N. Agrawal. Active vibration suppression of a flexible structure using smart material and a modular control patch. *Proc. Inst. Mech. Eng. (IMechE 2000)*, 214(Part G):217–229, 2000.

[18] C. K. Sung and Y. C. Chen. Vibration control of the elastodynamic response of high sped fleible linkage mechanisms. *ASME Journal of Vibration and Acoustics*, 113(14-21), 1991.

[19] William T. Thomson and Marie Dillon Dahleh. *Theory of Vibration with Applications*. Prentice Hall, New Jersey, 1998.

Masih Ali Hosseini was born in 1982. Masih earned his B.Sc. (2005) in aerospace engineering from Sharif University of Technology, Teheran, Iran, and M.Sc. (2008) in biomedical Engineering from RWTH–Aachen University of Technology, Nordrhein-Westfalen, Germany.

He worked on intelligent control of nonlinear hydraulic servos for his undergraduate thesis with Dr. A. Khayyat at Sharif University. He was a part time research assistant (2006-2007) at mediTEC in Helmhloltz Institute Aachen and a full-time research student (2007-2008) at Institute of Robotics and Mechatronics of German Aerospace Centre (DLR–Hirzinger's Lab), in Oberpfaffehofen, Bayern, Germany. He wrote his masters thesis on control of teleoperated master/slave robots with force feedback over the Internet and satellite links. He is currently working on his second M.A.Sc degree in mechatronics at Simon Fraser University, Vancouver, British Columbia. His current research topic is design of active engine mounts for variable displacement engines under supervision of Dr. M.F.Golnarafghi, Dr. S.Arzanpour and Dr. A.Parameswaran.

Compliant Constant Force Microgripper for Microassembly and Micromanipulation

Mohamadsadegh Hajhashemi, Farshad Barazandeh, and Hadi Parsaiyan

Abstract— **The characterization process of micro- and submicrometer particles demands their careful handling and precise placement. In most cases, a well designed microgripper can address these requirements. The following research is focused on the process of design, finite element analysis, and microfabrication of an innovative compliant constant force microgripper. This new architecture of microgripper enables it to apply an invariable force to the handled micro components during the micro-assembly process. UV-assisted vertical etching was used to realize the proposed design on the Polyethylene Terephthalate (PET) substrate. This process is a well established process to address requirements of high resolution and high-aspect-ratio in microfabrication of plastic structures. Finally, performance of the prototyped sample has been successfully verified during the micro-assembly process.**

Index Terms— **Compliant Mechanism, Constant Force, Microgripper, UV-assisted vertical etching**

I. INTRODUCTION

Micro-manipulation is a vital step in the process of assembly and maintenance of a micro-machine and its components. Efficient, reliable and flexible handling is still very challenging in the micro-assembly process. Due to the fragility and sensitivity of micro parts, applied force is an important factor that should be under controlled. Force guided assembly has been studied extensively in the last two decades (for example, Whitney [1], Lozano-Perez, Mason, and Taylor, [2], Erdmann [3], McCarragher and Asada [4], Peshkin [5]). For monitoring an assembly process and guiding a hand-held part, the force and moment acting at the contact points are useful information directly reflecting the assembly process state. Unlike visual information, the force information provides real-time, in-situ information during parts mating operations. Therefore, the force information is indispensable for closed-loop assembly process control as well as for in-situ monitoring. To make robotic assembly reliable and robust in the face of a high degree of uncertainties, the close loop control based on force information and in-situ monitoring is critically important. The question is how to obtain reliable force despite inherent difficulties. In the previous research, intricate control algorithms have been used to measure and manage this force [6], but the proposed mechanism in this paper has been designed based on the

M. Hajhashemi is with the Engineering Science Department, Simon Fraser University, Vancouver, Canada (e-mail: mhajhash@sfu.ca).
F. Barazandeh is with the Mechanical Engineering Department, Amirkabir University of Technology, Tehran, Iran.
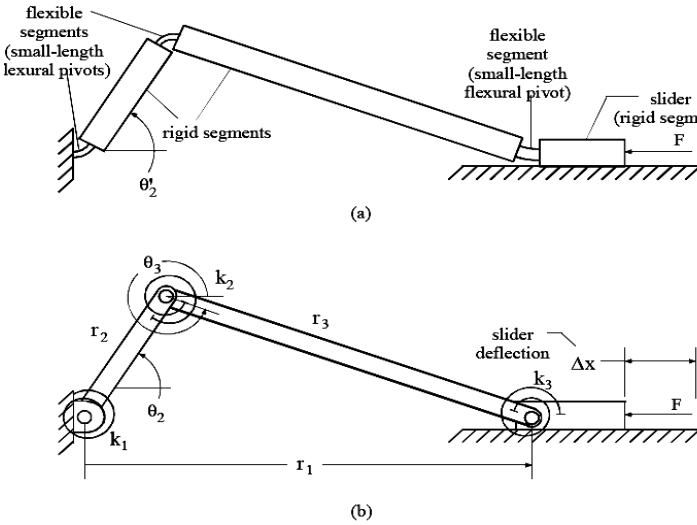
optimized set of dimensions such that regardless of the micropart size, grasping force remains constant [7].

In a compliant mechanism, motion comes from the deflection of the parts of the mechanism. Compliant mechanisms offer several advantages over rigid-body mechanisms. For example the deflection characteristics allow direct energy storage within a flexible member, which eliminates the necessity of additional energy storage devices (e.g., springs). Constant-force mechanisms can be compliant mechanisms or rigid-body mechanisms with either linear or torsional springs or a combination of them. In general, they use the principles of stored strain energy to produce a near-constant grasping force over a large range of jaw displacement. This constant force is constant force accomplished by determining specific geometric ratios which allow for equal increases in the stored strain energy. Therefore, the output force remains constant throughout the displacement. Moreover, the microgripper dimensions were carefully selected to be compatible with micro-manufacturing technologies. The prototyped microgripper is capable of manipulating components with dimensions from a few to hundreds of micrometers. In this design, the microgripper tips move in parallel, and thus, they are always mutually aligned.

II. STRUCTURE AND PRINCIPLE OF OPERATION

A. *Slider-Crank Model*

The original behavioral model is based upon a simple compliant slider-crank mechanism. By solving PRBM and standard kinematic equations, the position of the slider-crank is determined. The variables used in the equations and the mechanism orientation are shown in Fig. 1[7]. Knowing r_2, r_3 and $\varDelta x$ as given parameters, the angles θ_2 and θ_3 and the length can be determined from following equations (1, 2, and 3):

$$\theta_2 = a\cos\frac{r_1^2 + r_2^2 - r_3^2}{2r_1.r_2}, \quad \theta_3 = a\cos\frac{-r_2.\sin\theta_2}{r_3}, \quad r_1 = r_2.\cos\theta_2 + r_3.\cos\theta_3 \quad (1)$$

$$\theta_3 = a\cos\frac{-r_2.\sin\theta_2}{r_3} \quad (2)$$

$$r_1 = r_2.\cos\theta_2 + r_3.\cos\theta_3 \quad (3)$$

(a)

(b)

Fig. 1. Compliant and rigid-body slider crank model [7].

B. *Principle of Virtual Work*

The principle of virtual work and the PRBM can be used to determine the static force for a given amount of deflection. In order to determine the relationship between static force and a given amount of deflection, a set of equations must be developed to relate displacement, compliant member deflection, and the static input force (F_j). Using the principle of virtual work and the PRBM, the virtual work (δw) can be calculated from (4):

$$\delta w = \sum_i \overrightarrow{F_i}.\overrightarrow{\delta z_i} + \sum_j \overrightarrow{M_j}.\overrightarrow{\delta \theta_j} - \sum_k \frac{dV_k}{dq_k}\delta q_k \qquad (4)$$

Where F and M are the exerted force and moment respectively, V is the potential energy, q is the generalized coordinate and $\overrightarrow{\delta z}$ and $\overrightarrow{\delta \theta}$ are virtual displacements. Using (2), the amount of virtual work associated with each torsional spring was determined. Another equation was developed relating an unknown static input force applied to the slider in the horizontal direction. Applying the principle of virtual work, this set of equations can be solved together. Equation 5 shows the calculated force determined by this method.

$$F = \frac{r_3 \cos\theta_3 [k_1\theta_2 + k_2(2\pi + \theta_2 - \theta_3)] + r_2 \cos\theta_2 [k_2(2\pi + \theta_2 - \theta_3) + k_3(2\pi - \theta_3)]}{r_2 r_3 (\sin\theta_2 \cos\theta_3 - \sin\theta_3 \cos\theta_2)} \quad (5)$$

This equation determines the relationship between the amount of exerted force, spring constants, and geometric parameters of the design such as the length of the links and corresponding angles as defined in Fig. 1(b) [7]. For the purpose of optimization, now an objective function should be defined. Parameters Ξ and d in (6) and the optimization objective in (7) are defined and developed for this purpose.

$$\Xi = \frac{\max(\overline{F})}{\min(\overline{F})}, \quad Deflection\,parameter(d) = \frac{(r_2 + r_3) - r_1}{r_2 + r_3} \times 100 \quad (6)$$

$$Minimize\,\Xi \qquad (7)$$

The optimum parameters for the above problem are shown in (8). Fig. 2 illustrates the relationship between non-dimensional parameter φ, defined in (9), versus the deflection parameter. It should be noted that the initial deflection of the mechanism results in a large force spike. However, after the initial deflection, the results are completely reliable.

$$\frac{r_3}{r_2} = 2.0821 \qquad \frac{k_2}{k_1} = 1 \qquad \frac{k_3}{k_1} = 9.3816 \qquad (8)$$
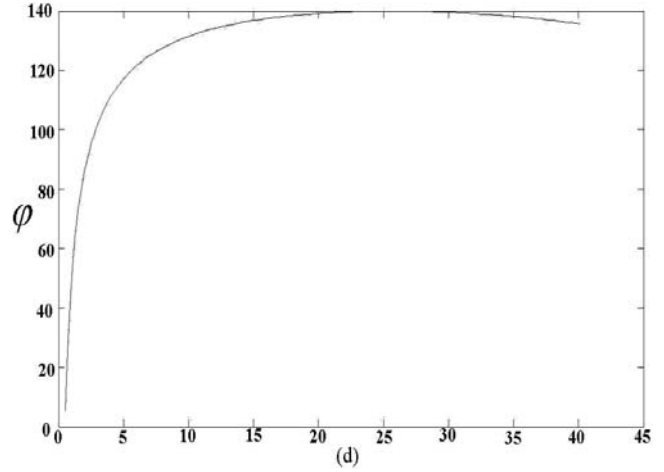
$$F = \frac{k_1}{r_2}\varphi \qquad (9)$$

Fig. 2. Deviation of the applied force versus the deflection parameter.

III. FINITE ELEMENT SIMULATION

The optimized compliant constant force calculated in the previous section was successfully used to design the Microgripper's structure. Knowing limitations of the existing fabrication processes, proper dimensional considerations has been applied to the design, and for the purpose of investigation, a Finite element analysis has been undertaken with the ABAQUS software package. Three dimensional elements (3D) were used in the FE model of the microgripper mechanism to predict the stress concentration and displacements of the microgripper mechanism. Because this microgripper is composed of two symmetrical jaws, the results derived for a symmetrical half-model condition is shown in fig. 3. These results are correspondingly applicable to the other side of the mechanism. As shown in fig. 3, bulk stress concentration occurs at the notch hinge points. The maximum value of concentrated stress was found to be 65 MPa which is still less than the yield stress of the material.
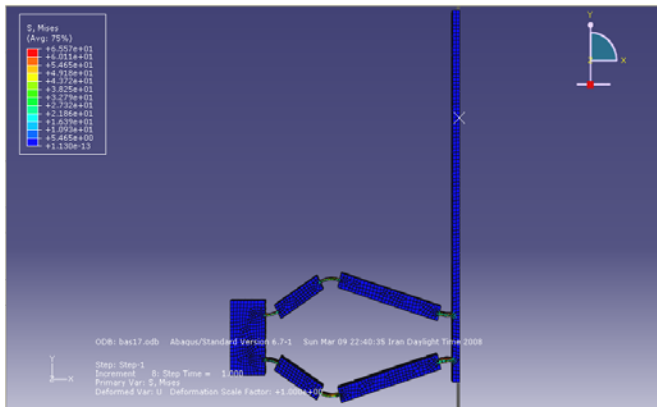
Fig. 3. Finite element analysis of constant-force microgripper.

IV. FABRICATION PROCESS

Polyethylene terephthalate (PET) is a robust material and chemically resistant to many solvents. These characteristics make it a suitable choice for many large areas in electronic applications. Thus, in the following study, the microgripper is fabricated of PET using a UV assisted vertical etching process [8]. The presence of the UV light promotes etching in the direction determined by the incident photons. Contrary to the etching methods described in literature [9], [10], where either an X-ray radiation or energetic ion bombardment is used to accelerate the etching process, in the adopted fabrication method, only a UV source is utilized during the chemical etching of the substrate. Using Di-Methyl-Formamide (DMF) solution and proper UV illumination, a micro-machining process on the PET substrates has been completed successfully. The masking layer used in these experiments is an Si layer deposited with the Plasma Enhanced Chemical vapor deposition. This mask has been found to be suitable in order to protect the PET substrate from the UV exposure (as well as the solvent in the protected areas). Using photolithography, the desired patterns can be made in the masking layer. Fig. 4 illustrates a schematic view of the setup used in this study. This setup is also equipped with a water-cooled glass container and a heater to control the temperature of the solvent during the etching process [8].
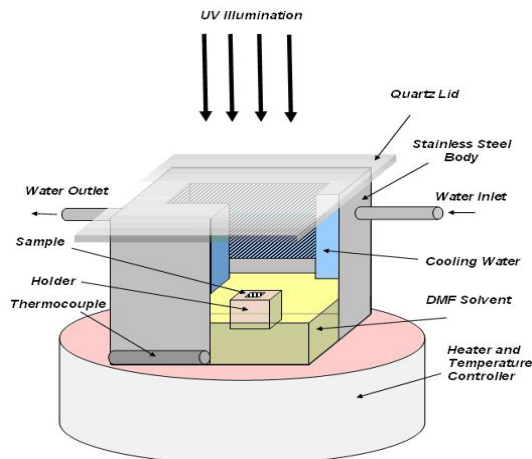
Fig. 4. Schematic of the setup used for ultraviolet assisted micromachining [8].

The micro-fabrication process is shown schematically in table 1. The deposited a-silicon layer protects the PET surface from exposure to the UV and DMF, and so acts as a mask. Following the photolithography process, the sample is placed in a Reactive Ion Etching (RIE) device to dry-etch the silicon layer in the SF_6 plasma.

Table1. Process flow for fabricating the microgripper [11].

1.Cleaning the PET substrate in the DCM	5.Etching of uncovered silicon layer with SO_2
2.Depositing the silicon layer	6.Removing the photoresist layer
3.Creating the $1-2\,\mu m$ photoresist layer	7.Vertical etching of PET in DMF
4.Patterning the photoresist layer	8.Removing the silicon layer

Based on the simulation and optimization results, several designs with different dimensions of the critical features were selected. Fig. 5 shows a magnified scanning electron microscopy (SEM) image of the final microgripper's structure. The minimum feature size of this microgripper is 20×100 μm.
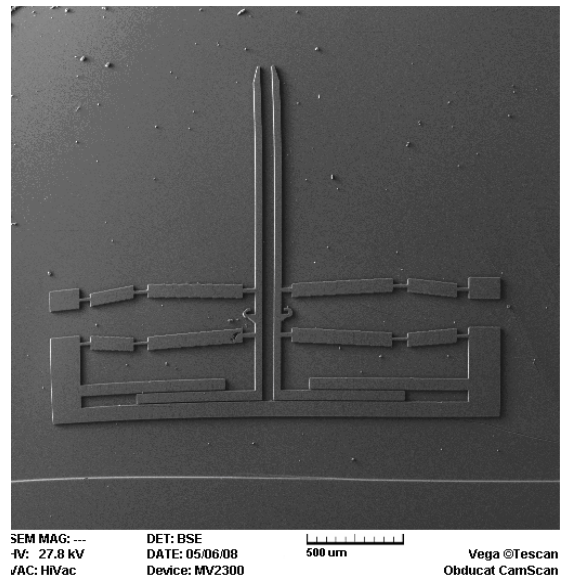

Fig. 5. SEM images of the fabricated microgripper.

V. PERFORMANCE TESTING

A micro manipulation test was performed using the proposed microgripper device. The assembly of miniaturized

gear systems, typically with diameters below 2 mm, requires the use of special micro-gripping tools. A miniature rotor's dimensions are typically a few hundreds of a micrometer. Using Pick and place steps, miniature wrist watch rotors can be assembled as shown in fig. 6. This figure is showing images taken from the jaws of the microgripper approaching to the miniature rotor and grasping it. In order to depict the grasping process, the direction of the mechanical contact between the miniature rotor and the microgripper is selected to be perpendicular to the rotation axis of the rotor.
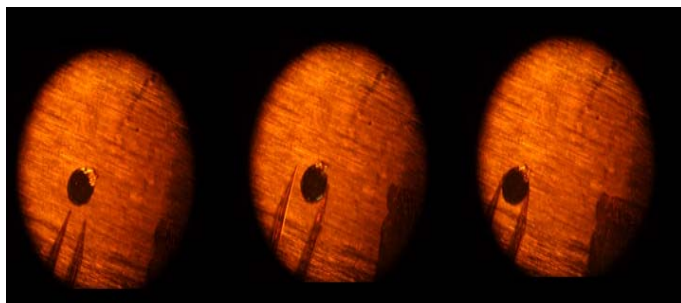
Fig. 6. Grasping action of the microgripper on the miniature rotor.

VI. CONCLUSION

The processes of design, microfabrication and testing of a novel monolithic compliant constant force microgripper have been described in this study. After the finite element analysis (FEA), the prototype of this microgripper was fabricated from a PET substrate using ultraviolet assisted vertical etching. The successful working stroke of this microgripper is 150 to 500 micrometers. Due to the constant force which is exerted by a compliant constant force mechanism, controlling the applied force is no longer needed. Possible applications are the handling of microparts with complex shapes where adhesive or vacuum gripping is not suitable. Additionally, this device can facilitate a micro-assembly process in restricted and difficult to access environments. Finally, micromanipulation tests were conducted to confirm the potential applications of this microgripper in handling micro-components.

ACKNOWLEDGMENT

REFERENCES

[1] D.E. Whitney, "Force Feedback Control of Manipulator Fine Motions", ASME J. of DSMC, vo1.99, no.2, pp91-97, 1977.
[2] T. Lozano-Perez, M.T. Mason and Taylor, "Automatic Synthesis of Fine-Motion Strategies for Robots", Int. J. of Robotics Research, vo1.3 no.1, pp.3-24, 1984.
[3] M. Erdmann, "Using Backprojections for fine Motion Planning with the Uncertainty", International Journal of Robotics Research, vo1.5, no.1, pp.19-45, 1986.
[4] B.J. McCarragher and H. Asada, "Qualitative Template Matching Using Dynamic Process Models for State Transition Recognition of Robotic Assembly", ASME Journal of Dynamic Systems , Measurement, and Control, 1993.
[5] M.A. Peshikin, "Admittance Matrix Design for Force-Guided Assembly", IEEE Transactions on Robotics and Automation, Vol. 8.No.2, 1992
[6] Jungyul Park, Sangmin Kim, Deok-Ho Kim, Identification and Control of a Sensorized Microgripper for Micromanipulation, IEEE/ASME Transactions on mechatronics, Vol. 10, No. 5, October 2005
[7] Howell, L.L., Midha, A., and Murphy, M.D, 1994, "Dimensional Synthesis of Compliant Constant-Force Slider Mechanisms," *Machine Elements and Machine Dynamics*, DE, Vol.71, pp. 509-515.
[8] T. Maleki, S. Mohajerzadeh, and A. Afzali-Kousha,"Plastic micromachining assisted by ultraviolet illumination," *IEEE Trans. Electron Devices*, vol. 50, no. 8, pp. 1813–1815, Aug. 2003.
[9] A. I. Vilenskey, D. I. Zagorski, and S. A. Bystrove, "Investigation of latent tracks in poly-ethylene-terephthalate and their etching," *Surface Science*, vol. 507–510, pp. 911–915, 2002.
[10] V. P. Nazmov *et al.*, "Influence of irradiation of poly-ethylene-terephthalate films with X-ray upon the rate of dissolution in alkali water solution," in *Nuclear Instruments and Methods in Physics Research*, 2001, vol. B173, pp. 311–318.
[11] A.Jahanshahi, A.H.Tamaddon, M.Sadeghi, S.Mohajerzadeh, "Enhanced UV-Assisted Vertical Etching of PET Suitable for the Fabrication of Plastic Micro Systems ", in Proceedings of the MME 2008 Workshop, pp. 197-201, Aachen, Germany.

Mohamadsadegh Hajhashemi. was born in 1984. He recieved the B.S. and M.S. degree in mechanical engineering from the Isfahan University of Technology and Amirkabir University of Technology respectively. He is presently enrolled in the Ph.D. program with the department of engineering science at Simon Fraser University.
His main research interests are in the areas of microelectromechanical systems(MEMS), compliant mechanisms , microassembly and topology optimization methods.

Biogeography Based Optimization Algorithm for Computationally Efficient Symbol Detection in Multi-Device STBC-MIMO System

Saeed Ashrafinia

Communication Research Lab

Simon Fraser University

Burnaby, Canada

Saeed_Ashrafinia@sfu.ca

Abstract— **In this paper, a new Evolutionary Algorithm is proposed to detect symbols in Multi-Device (MD) Space-Time Block Coded (STBC) Multi Input Multi Output (MIMO) Communication System. Exhaustive search for finding an optimal detection – Maximum Likelihood (ML) detection – has a computational complexity, that increases exponentially with the number of mobile devices, transmit antennas per mobile device, and the number of bits per symbols (complex modulation schemes e.g. MPSK or M-QAM). We apply Biogeography-Based Optimization Algorithm (BBO) to this system to find a nearly optimal solution in real time. It requires very low amount of computation in comparison with exhaustive ML search, likewise a nearly optimum detector like Sphere Detection (SD). BBO algorithm is an evolutionary computation, which updates its chosen popu-lation during each iteration based on migration and mutation of the current individuals. The simulation results is also presented to compare the performance of the proposed BBO algorithm with other detection methods such as SD, Zero Forcing (ZF), MMSE, V-BLAST, Semi-definite Relaxation (SDR), as well as another heuristic algorithm like Genetic Algorithm (GA), in terms of Bit Error Rate (BER) and complexity. The effectiveness of BBO is verified through these simulation results.**

Keywords- Biogeography-Based Optimization, Space Time Block Code, Evolutionary Algorithm

I. INTRODUCTION

Multi-Input-Multi-Output (MIMO) communication systems are current theme of international wireless research. This technology offers significant increases in data throughput and link range without additional bandwidth or transmit power, thus has attracted attention in wireless communication. Recent research also shows that MIMO communication systems have significantly higher channel capacity than the Single-Input-Single-Output (SISO) systems for the same total transmission power and bandwidth [1,2]. Moreover, Space Time Block Code (STBC) can increase the capacity of MIMO systems, and consequently improves data throughput and spectral efficiency [3].

The proposed system in this paper has multiple transmitters and one receiver. The transmitter section comprises of multiple devices, each of them has multiple antennas that applies Space-Time Block Code (STBC). Joint detection is performed at the receiver side, which has multiple antennas in the front. Such a system is called a Multi-Device (MD) STBC-MIMO system,

which is illustrated in Fig.1. One example of such systems is an uplink multiple access communication system. Generally in such systems, number of receive antennas are smaller than the total number of transmit antennas (the sum of antennas of all mobile devices).

This paper is focused on symbol detection in MD-STBC-MIMO systems. As will be discussed in Section III, at the presence of a-priori equally likely blocks of symbols, the probability of error will be minimized by the Maximum Likelihood (ML) detection, which is optimal. However, ML cannot be performed with a computationally efficient algorithm. Studies with the near-optimal Sphere Decoding (SD) algorithm exhibit its expected computational complexity grows polynomially for small constellation sizes and exponentially for large constellation sizes (Section IV, [4]). Nevertheless, an efficient algorithm with polynomial growth of expected complexity has not been found. In fact, even the expected computational complexity of the sphere decoding grows exponentially with the problem size in MIMO communication systems [5]. Verdú also proposed the single antenna-based ML Multiuser Detector that offers the best Bit Error Rate (BER) performance among all other detectors [6].

Computationally complex problems can be treated as heuristic problems, especially with the use of evolutionary algorithms. We applied the Biogeography-Based Optimization (BBO) and Genetic Algorithm (GA) as heuristic algorithms to the problem. GA is an artificial intelligence technique based on the biological concepts of genetic combination and natural selection, inspired by Darwin's theory of evolution. In GA Parameters to be optimized (known as genes) are concatenated into data strings (called chromosomes). These chromosomes undergo the operations of selection, crossover and mutation [7].

Another type of evolutionary optimization methods recently has been presented by Simon D. is Biogeography-Based Optimization (BBO) [8], which considers the mathematics of biogeography as the basis for the development of a new-developed evolutionary algorithm. BBO has features in common with other biology-based optimization methods, such as Genetic Algorithm and Particle Swarm Optimization [9], which makes BBO applicable to many problems of the same types. Simulation results show that BBO can meet near-optimal SD while it is more computationally efficient than SD and optimal

ML, and have considerably better BER performance than Zero Forcing (ZF), Minimum Mean Square Error (MMSE), Vertical Bell Laboratories Layered Space-Time (V-BLAST) [10], Semi-Definite Relaxation (SDR) [11], and GA.

In the rest of this paper, the system model is presented in section II. The application of existing symbol detection algorithms is discussed in section III. In Section IV the idea of applying BBO on symbol detection is presented. Section V compares the computational complexities; and the simulation results are presented in section VI.

II. SYSTEM MODEL

Fig. 1 shows an MD-STBC-MIMO system with one receiver and multiple transmit devices. Each of the K mobile devices at the transmitter section has N_T transmit antennas that applies STBC, whereas the receiver front end has N_R receiving antennas. The multiple devices in the proposed systems can cause co-channel interference. An IQ-modulation scheme (e.g. MPSK, M-QAM, etc.) maps source symbols into complex numbers. Even if each signal employs an orthogonal space-time code, the absence of coding across different mobile devices cannot guarantee the orthogonality among their signals.

First, let us consider the case of single mobile device; i.e. $K=1$. Hence the mobile device transmits using N_T transmit antennas and communicates with a receiver with N_R antennas. T is denoted as the number of timeslots in the space-time code block. The channel model assumed to be quasi-static; i.e. the channel gain remains constant during each block of data. The receiver also knows the Channel State Information (CSI). This assumption is reasonable if training or pilot signals are used to learn the channel. MIMO channel is denoted by a complex $N_T \times N_R$ dimensional matrix H (for one mobile device), and a complex $T \times N_T$ dimensional matrix S represents the input signal in a space-time code block. So the relation between the input and output signal is

$$\tilde{Y} = S.H + \tilde{Z} \tag{1}$$

where $\tilde{Y}$ is the $T \times N_R$ dimensional complex output matrix and $\tilde{Z}$ represents the additive white noise matrix.

Generally, for the linear dispersion space-time coding approach, the relation between input and output of the channel is often expressed in another form rather than (1) [12]. The input signal in STBC (matrix S) can be expressed as:

$$S = \sum_{q=1}^{Q} \left[\left(\alpha_q + j\beta_q \right) C_q + \left(\alpha_q - j\beta_q \right) D_q \right] \tag{2}$$

where Q is the number of symbols conveyed in a space time code block, and $\alpha_q + j\beta_q$, $q=1,...,Q$ are complex numbers that represent the Q symbols. (Note that α_q and β_q denote the real and imaginary parts of a symbol.) Then, a $2Q$-dimensional real-valued row vector χ is defined with the components constituted by α_q and β_q, $q=1,...,Q$. Therefore, the real and imaginary parts of the matrix $\tilde{Y}$'s components can be rearranged as a $2TN_R$-dimensional real-valued row vector y. Alternatively; χ and y have a new relation as following:
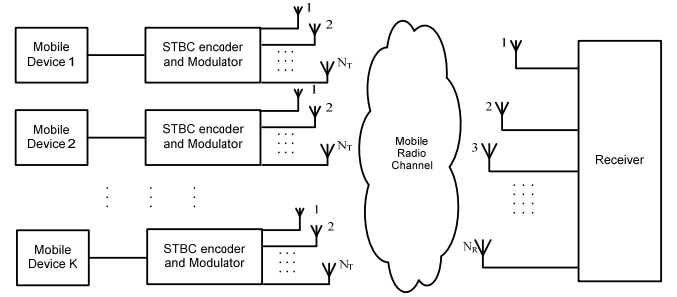
$$y = \chi \, \Omega + Z \tag{3}$$

Figure 1. A block diagram of MD-STBC-MIMO system

The new $2Q \times TN_R$ real-valued matrix Ω contains the component of matrices H, C_q, D_q, $q=1,...,Q$, and Z (the $2TN_R$-dimensional real-valued vector representing noise). In the case of multiple mobile devices, equation (1) is generalized to

$$\tilde{Y} = \sum_{k=1}^{K} S_k . H_k + \tilde{Z} \tag{4}$$

where the $T \times N_T$-dimensional complex matrix, S_k, is the input signal from mobile device k, and the $N_T \times N_R$-dimensional complex matrix H_k represents the channel from device k to the receiver. Correspondingly, matrix y can be written as

$$y = \begin{bmatrix} \chi_1 & \chi_2 & \cdots & \chi_K \end{bmatrix} \begin{bmatrix} \Omega_1 \\ \Omega_2 \\ \vdots \\ \Omega_K \end{bmatrix} + Z \tag{5}$$

Here χ_k is a $2Q$-dimensional real-valued row vector representing the Q_k complex symbols sent from mobile device k in a space-time code block. The total number of $N_S = \sum_{k=1}^{K} Q_k$ symbols (from all mobile devices) are transmitted in a space-time coded block, through all transmit antennas.

III. SIGNAL DETECTION

The detector at the receiver has to choose among M^{N_S} possible transmitted symbols, where M is the size of the symbol constellation associated with the modulating scheme. ML detection is recognized to yield the lowest BER for a-priori equally likely symbols. In fact, ML detection chooses transmitted symbols to maximize $P \left(y \mid \chi_1 \quad \chi_2 \quad \cdots \quad \chi_K \right)$.

In the presence of additive white Gaussian noise, Z, ML detection chooses the vector $\begin{bmatrix} \chi_1 & \chi_2 & \cdots & \chi_K \end{bmatrix}$ from M^{N_S} possibilities that has the shortest Euclidean distance:

$$\left\| y - \sum_{k=1}^{K} \chi_k . \Omega_k \right\| \tag{6}$$

Defining $b \triangleq \log_2 M$, and M as the size of symbol constellation, the ML detection scheme can be implemented by searching through all $M^{N_S} = 2^{bN_S}$ symbol combinations. Performing such an exhaustive search to find the minimum of (6) is computationally inefficient, especially for large N_s. Compu-

tational complexity increases exponentially with N_s, number of bits per symbol, transmit antennas per device, and number of mobile devices (K). High-speed communication requirements demand a low-complexity detection scheme.

As low-complexity near-optimal detectors, two evolutionary algorithms, GA and BBO are applied to the MD-STBC-MIMO detection problem in this paper. Therefore, the problem is converted into discrete optimization problems that search through a portion of M^{N_s} symbol combinations.

IV. BIOGEOGRAPHY-BASED OPTIMIZATION ALGORITHM

Evolutionary Algorithms (EAs) are one group of well-known heuristic algorithms, which have been often used to solve difficult optimization problems. Usually EAs are inspired by the theory of biological evolution (e.g. selection, crossover, mutation, recombination and reproduction). Generally, candidate solutions to an optimization problem are represented as individuals in the population. The cost function value of a candidate solution indicates fitness of the individual. BBO algorithm, just like some EAs, uses crossover (migration) and mutation operators to generate a new population of individuals during each iteration. These operators based on a probabilistic manner estimated from the best-selected individuals of the previous iteration.

BBO is a population-based EA based on the mathematics of biogeography. Biogeography is the study of the geographical distribution of biological organisms. In BBO, islands represent problem solutions, along with emigration and immigration of species between islands that represent the sharing of features between these solutions. Mathematical models of biogeography describe the migration, speciation, and extinction of species. Species migrate between islands. Islands that are well suited as habitats for species are said to have a high Habitat Suitability Index (HSI), i.e. the goodness of the solution is high. Each solution feature is called a Suitability Index Variable (SIV). SIVs are the elements of an island (solution vector) that can be considered as its independent variables. Likewise, HSI is the fitness function that can be considered the dependent variable.

In original BBO large number of species in an island (high HSI) causes more species to emigrate, so its emigration rate, denoted by μ, is correspondingly large. On the other hand, the immigration rate for this island, denoted by λ, is smaller because it's less likely to sustain further immigration of species due to the high demand on its existing environmental resources. In practice, λ and μ are assumed linear, having the maximum immigration and emigration rate of I and E, respectively.

Suppose in some problem, there are certain numbers of candidate solutions. A good solution is the one analogous to an island with a high HSI. High HSI solutions are more likely to share their features with other solutions, and low HSI solutions are more likely to accept shared features from other solutions. In BBO, just like every other evolutionary algorithm, each solution may also have some probability of mutation, although mutation is not an essential feature of BBO.

Various algorithms have been developed to use migration rate of each solution probabilistically in order to share features between them. In this paper, a special case of the partial immigration-based BBO is used [13]. In partial immigration-based BBO, the immigration curve is used probabilistically for each SIV to decide whether to immigrate. If immigration is selected for an SIV, then the emigrating island is selected probabilistically.

In this approach, λ and μ are assumed linear. However, in order to reduce the complexity of the algorithm to beat the other evolutionary algorithms complexities, μ curve assumed to be constant, which is shown in Fig. 2. Linear immigration rate increases the probability of poor solutions to enhance, just like it prevents good solutions to be changed. Immigration curve can be written as $\mu_k = k\,(E/n)$, where k is the species count and $n = S_{max}$. Subsequently, setting a constant emigration rate uniformly distributes islands, thus each solution equally likely selects an island to immigrate SIVs from it. A conceptual description of one generation of BBO using this approach is presented in Fig. 3.

In order to mutate the individuals, the mutation upper bound, m_{max} is fixed with a small number. In the mutation procedure for each SIV, m_{max}, as an upper limit, is compared with a random number to decide whether to mutate. Although this simple method is less efficient for mutation, compare to the one proposed in [8], it yields lower computational complexity.
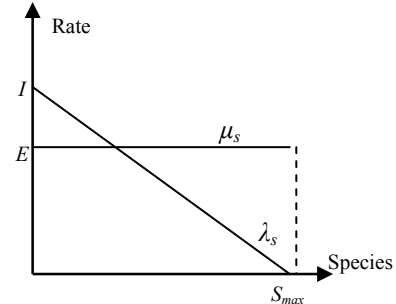

Figure 2.   A block diagram of MD-STBC-MIMO system

```
For each iteration g_i
    Calculate HSI
    Sort population                             Migration
    For each island x_i
        Use λ_i to decide whether to immigrate to x_i(s)
        If immigrating then
            Use μ_i to select the emerging island x_j
            x_i(s) → x_j(s)
        End if
    NEXT island                                 Mutation
    For each island x_i
        Use m_max to decide to mutate each SIV
        If mutating x_i(s) then
            Replace x_i(s) with a randomly generated SIV
        End if
    NEXT island
NEXT generation
```

Figure 3.   One generation of BBO pseudo code

As with other population-based optimization algorithms, elitism is also incorporated in order to retain the best solutions in the population. In this case, the algorithm should not migrate or mutate the elite solutions. In the proposed algorithm, since $\mu_n = 0$, there is no chance for the best solution to be changed throughout the migration. Also the mutation probability can be set to zero for the best solutions to avoid mutating. Throughout this procedure, having at least one elite solution which hasn't been modified during an iteration is assured.

V. COMPUTATIONAL COMPLEXITY

A motivation for applying the proposed near-optimal algorithms to a MD-STBC-MIMO problem is their low computational complexity. In this section, the computational complexity of BBO for MD-STBC-MIMO symbol detection is compared with that of ZF, MMSE, SD, SDR, V-BLAST, GA and the exhaustive search. The computational complexity of exhaustive search (an implementation of the ML detector) is well-known to be $O(2^n)$ where $n = N_S \, log_2(M)$ [14]. However, exhaustive search is usually impractical for real-time operations of symbol detection. In order to overcome this complexity, a number of suboptimal detection schemes with better computational complexity are presented.

The worst case complexity of sphere decoder is exponential and its expected complexity depends on problem size and SNR [15]. At low SNR SD has high complexity that is $O(\tilde{n}^6)$ [16], however at high SNR it has polynomial complexity, often roughly cubic [15]. ZF is one of the sub-optimal detectors that involve inverting a matrix, and its computational complexity is $O(\tilde{n}^3)$ where $\tilde{n} = n/log_2(M)$. The Computational complexity of MMSE detector is also $O(\tilde{n}^3)$ [17]. In V-BLAST, each iteration requires either ZF or MMSE operations, and the number of iterations is equal to the total number of transmitted antennas. Hence the upper bound on computational complexity of V-BLAST has two cases: If the total number of transmit antennas in the system is equal to the number of receiving antennas, then complexity is $O(\tilde{n}^3)$. If the total number of transmit antennas in the system is greater than the number of receive antennas, then complexity is $O(\tilde{n}^4)$ [18]. The computational complexity of single-SDR is $O(N_t^{3.5})$ where N_t stands for number of transmit antennas [19].

In each iteration of GA, the most computational complex parts are calculating the fitness function, natural selection and crossover, which have no more complexity than $O(g.n)$, where g is the number of iterations and n is the number of population per iteration. Therefore the computational complexity of GA would be $O(g.n)$ which is particularly smaller than ML [20].

The computational complexity of BBO also can be written as a function of number of iterations (g) and population size per iteration (n). According to the BBO pseudo code presented in Fig.2, using the migration approach of partial immigration-based BBO empowers immigration for all SIVs inside each island. In this algorithm, first immigration rates are used probabilistically to decide for all SIVs whether to immigrate. Then, if immigration happens, new SIVs are equally-likely chosen from other islands. Using a matrix representation of islands and SIVs yields to the complexity of $O(g)$. To apply mutation, instead of using migration rates to compute new

probabilities for each species count, as presented in the original BBO work [8], just a simple comparison between random generated numbers and an upper bound mutation factor is performed to decide whether to mutate. The complexity of this new effortless approach also can't exceed $O(g)$. Likewise, the fitness function is evaluated during each iteration and its computational complexity is $O(g.n)$.

Summing up the computational burden of all these steps, the computational complexity of the algorithm would be $O(g) + O(g) + O(g.n)$. For large n and g it's simplified to $O(g.n)$. Not only this complexity order is significantly lower than the exhaustive search, it is also less than other suboptimal detection schemes mentioned above. In fact, GA has a comparable computational complexity with BBO; however, its performance falls behind BBO, as will be presented in the next section.

VI. SIMULATION RESULTS

The simulation results of the proposed BBO detection and its comparison with other detection techniques applied to a MD-STBC-MIMO system for performance comparison. The channels assumed to be quasi-static, and different channels in MIMO assumed to be independent. In all simulations, 4-QAM modulation scheme is used, and the number of timeslots in a space-time code block, T, is two. Moreover, some BBO parameters are kept constant through all simulations, such as $(I, E, m_{max}) = (1, 1, 0.05)$ and not mutating the best three solutions in each population in order to save them as elites.

The simulation results in figures 4 through 7 show the BER performance comparison between ZF, MMSE, V-BLAST, SDR[1], SD[2], GA and BBO detectors. The MIMO system configuration, $(K, N_T, N_R, M,)$, is $(4,2,6,4)$, $(5,2,8,4)$, $(6,2,10,4)$ and $(3,4,4,4)$ for figures 4, 5, 6 and 7, respectively. The Alamouti space-time coding is used for simulations results in figures 4, 5 and 6 [21], but for Fig. 8 a non-orthogonal four transmit antenna configuration is used for each mobile device. BBO parameters (iterations, islands per iteration) are $(60,100)$, $(100,100)$ $(100,120)$ and $(100,150)$, that has to search through search spaces of 4^8, 4^{10}, 4^{12} and 4^{12} possible solutions for figures 4, 5, 6 and 7, respectively. Moreover, the number of iterations and islands per iterations for BBO is the same as GA, in order to compare their results more precisely. The presented results are the average of 2500, 2500, 2000 and 1500 Monte Carlo trials for figures 4, 5, 6 and 7, respectively.

Figures 4, 5 and 6 indicate that in low SNRs in the MIMO system, BBO has significant better BER performance than ZF, MMSE, SDR, V-BLAST and GA, while it closely matches SD. In figure 6 for example, although BBO is just searching over approximately 10,000 individuals out of entire 4^{10} search space, its BER performance meets the near-optimal SD. According to this figure, BBO matches SD in low SNR, where GA requires about 0.6 dB more SNR than BBO to achieve BER of 10^{-2}. BBO perceptibly behaves as the best detection algorithm among other sub-optimal detection methods in all the four figures.

Figure 8 shows the results for the number of $K = 3$ users and $N_R = 8$ receive antennas. However, four transmit antenna is used for each user instead of Alamouti space-time code. Therefore, BBO searches through roughly 12,000 individuals out of entire 4^{12} search space, which is a reasonably small number of possible solutions. Similar to the previous figures, SD and BBO has the best SNR performance. In higher SNRs, GA's performance diminishes notably, while BBO pursues the near-optimal SD. BBO requires 2 dB less SNR than SD and BBO to achieve BER of 10^{-2}.

According to the computational complexity order of BBO and GA, more iterations until termination means more computation. Fig. 8 shows number of iterations required to achieve a desired BER. The plot shows that for about the first 27 iterations, GA has a better BER performance than BBO, but thereafter BBO per-forms better. After about 57 iterations, the curve becomes very close to SD and it matches SD for more than 87 iterations. This improved performance is consistently observed in simulations with other system configurations.
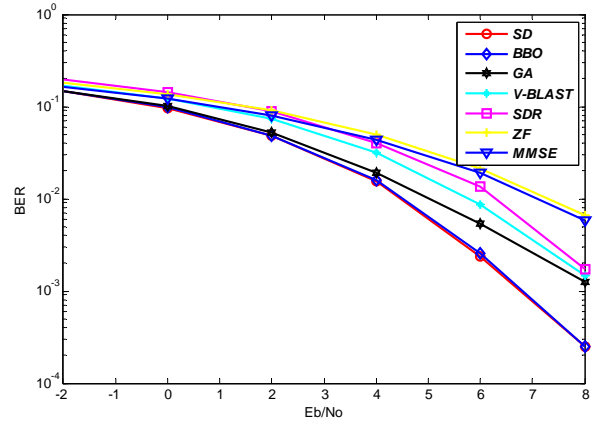
Figure 6. BER performance for $N_T=2$, $K=6$, 4-QAM with $N_R = 10$ antenna configuration over quasi-static fading channel. BBO with 100 iterations and population of 120 per iteration. Average BER of 2000 Monte-Carlo trials.
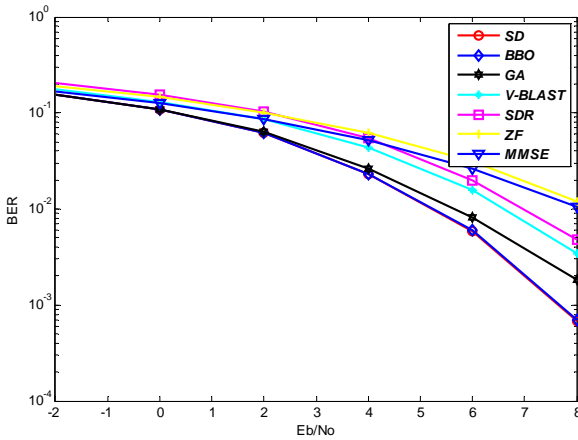
Figure 4. BER performance for $N_T=2$, $K=4$, 4-QAM with $N_R = 6$ antenna configuration over quasi-static fading channel. BBO with 60 iterations and population of 100 per iteration. Average BER of 2500 Monte-Carlo trials.
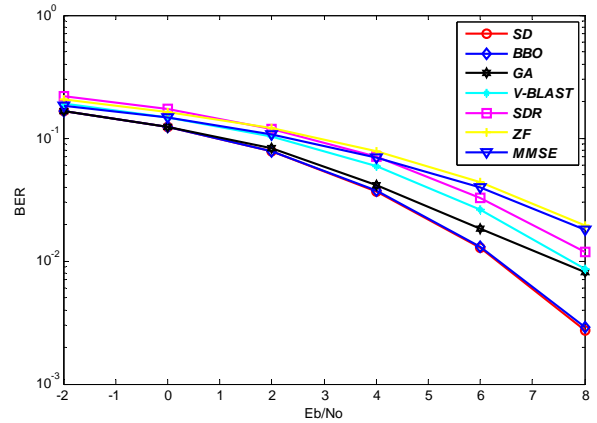
Figure 7. BER performance for $N_T=4$, $K=3$, 4-QAM with $N_R = 4$ antenna configuration over quasi-static fading channel. BBO with 100 iterations and population of 150 per iteration. Average BER of 1500 Monte-Carlo trials.
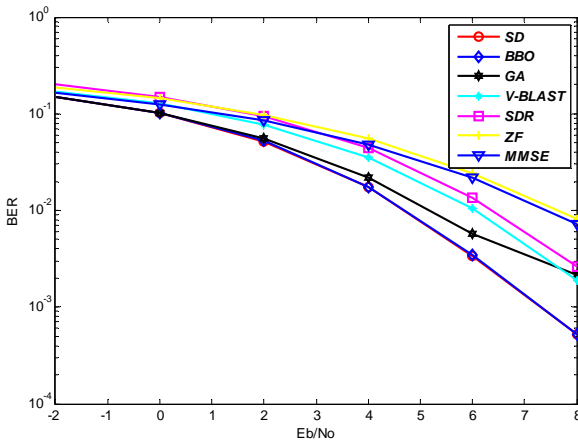
Figure 5. BER performance for $N_T=2$, $K=5$, 4-QAM with $N_R = 8$ antenna configuration over quasi-static fading channel. BBO with 100 iterations and population of 100 per iteration. Average BER of 2500 Monte-Carlo trials.
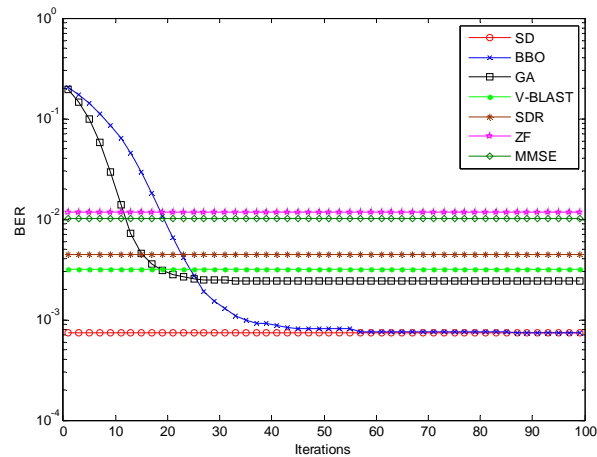
Figure 8. Average BER of 5000 Monte-Carlo trials. The MIMO system configuration, (K,NT,NR,M,), is (4,2,6,4). The population size is 100 and the minimum iteration is one.

VII. CONCLUSION AND FUTURE WORK

In this paper, we proposed the Biogeography-Based Optimization Algorithm for Multi-Device (MD) Space-Time Block Coded (STBC) Multi Input Multi Output (MIMO) Communication System. The complexity of these algorithms is very low as compared with optimal ML detector, so they are suitable for high-speed real-time communications. In addition, compare to other Evolutionary Algorithm like GA, BBO detection in MD-STBC-MIMO shows a significantly better performance, especially in high SNRs. The proposed algorithm has a consistently better performance-complexity trade-off at low SNRs, in comparison to existing algorithms. Even at high SNRs, this algorithm has relatively good performance-complexity tradeoff. Therefore, the proposed BBO algorithms are suitable for high-speed real-time communications. Also, the BBO-based MD-STBC-MIMO detectors are simple to implement, and also suitable for parallel implementation.

Because of a simplistic model, low implementation complexity, and convergence to a nearly optimal solution with a small number of iterations, BBO is a perfect solution to be applied to the same type of computationally complex problems in wireless communication. Moreover, further studies can be done in order to find ideal combinations of iteration/population in different SNRs, as well as working on other low-complexity algorithms for migration and mutation of BBO that can return more excelling result with the same level of complexity. Furthermore, applying a duplication check during each iteration with a low complexity level may be a good point of interest in order to improve the result by elevating the chance of searching through more unique individuals in the algorithm.

REFERENCES

[1] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Communications*, vol. 6, pp. 311-335, 1998.

[2] E. Telatar, "Capacity of Multi-antenna Gaussian Channels," *European Trans. on Telecomm*. vol. 10, pp 569-709, Nov. 1999.

[3] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communications: performance criterion and code construction," *IEEE Trans. Information Theory*, vol. 44, pp. 744-765, Mar. 1998.

[4] H. Vikalo and B. Hassibi, "On the sphere decoding algorithm: Part II, generalizations, second-order statistics and applications to communications," *IEEE Trans. on Signal Processing*, vol 53, no. 8, pp. 2819-2834, Aug 2005.

[5] J. Jalden and B. Ottersten, "On the complexity of sphere decoding in digital Communications," *IEEE Trans. on Signal Processing*, vol. 53, no 4, pp. 1474-14844, April 2005.

[6] S.Verdú, "Minimum Probability of Error for Asynchronous Gaussian Multiple Access Channels," *IEEE Trans. Information Theory*, vol. 32, pp. 85-96, Jan. 1986.

[7] D. Goldberg, Genetic Algorithms in Search, Optimization, and Machine Learning (Addison-Wesley, 1989).

[8] D. Simon, Biogeography-based optimization, IEEE Transactions on Evolutionary Computation (12) pp. 702-713, December 2008.

[9] R. Eberhart, Y. Shi, and J. Kennedy, "Swarm Intelligence" (Morgan Kaufmann, 2001).

[10] P.W. Wolniansky, G.J. Foschini, G.D. Golden, and R.A. Valenzuela, "V-BLAST: an architecture for realizing very high data rates over the rich-scattering wireless channel," *Proc. ISSSE-98*, Pisa, Italy, Sept. 29, 1998.

[11] Nekuii, M., Kisialiou, M., Davidson, T. N. and Luo, Z.-Q., "Efficient Soft Demodulation of MIMO QPSK via Semidefinite Relaxation," Proceedings of 2007 IEEE International Conference on Acoustics, Speech, and Signal Processing, pp. 2665-2668, April 2008.

[12] B. Hassibi and B.M. Hochwald, "High-rate codes that are linear in space and time," *IEEE Trans. Information Theory*, vol.48, no.7, Jul. 2002, pp. 1804−1824.

[13] D. Simon, "A probabilistic analysis of a simple biogeographybased optimization algorithm," unpublished.

[14] S. Verdú, "Computational Complexity of Optimum Multiuser Detection," *Algorithmica*, vol.4, no.3, pp. 303-312, 1989.

[15] B. Hassibi and H. Vikalo, "On the sphere decoding algorithm: Part I, the expected complexity," *IEEE Trans. Signal Processing*, vol. 53, no. 8, pp. 2806-2818, Aug 2005.

[16] O. Damen, A. Chkeif, and J. C. Belfiore, "Lattice code decoder for space-time codes," *IEEE Communications Lett*., vol. 4, no. 5, pp. 161–163, May 2000.

[17] C. Comaniciu, N.B. Mandayam, and H.V. Poor, *Wireless Networks: Multiuser Detection in Cross-Layer Design*, Springer, NY, May 2005.

[18] B. Hassibi, "An efficient square-root algorithm for BLAST, *Proceedings of IEEE ICASSP* 2000, pp. 737-740.

[19] M. Kisialiou and Z.-Q. Luo "Performance analysis of quasi-maximum-likelihood detector based on semi-definite programming," *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 3, 2005, p. 433.

[20] Wei, H.; Hanzo, L.; 2004. "Reduced-complexity near-optimum genetic algorithm assisted multiuser detection for synchronous multicarrier cdma," in *VTC 2004 - IEEE 59th Vehicular Technology Conference*, vol. 3, pp. 1717 – 1721.

[21] S. M. Alamouti, "A Simple Transmit Diversity Technique for Wireless Communications," *IEEE journal on select areas in communications*, vol. 16, no. 8, pp. 1451-1458, Oct. 1998.

[22] http://www.ece.umn.edu/~luozq/software/sw_about.html.

Saeed Ashrafinia received the B.S. degree in control engineering from Sharif University of Technology, Tehran, Iran, and currently he is a M.S. candidate in communication engineering in Simon Fraser University, Burnaby, Canada.

He has about three years work experience in telecommunication industry in Iran. These experience includes working at Fanamoj and Arg Telecomtwo of top-ranked Iranian telecommunication companies, as well as about two years studying, teaching and working for Advanced Information and Communiation Technology Center (AICTC) at Sharif University of Technology, and had all the responibilities of its labs. His industrial and academic experiences include computer networks, GSM mobile systems and 3G, VoIP, Session Initiation Protocol, Linux administration and security, Digital Video Broadcasting (DVB-T, DVB-h) and videoconferencing. In 2008, he moved to Canada to enhance his experience with the knowledge of great North American universities. His research areas includes wireless and mobile communications, optimization and evolutionary algorithms, queueing theory and networking.

IMECE2009-12878

A COMPARATIVE STUDY OF DIFFERENT CONTROL STRATEGIES FOR A MAGNETORHEOLOGICAL HAPTIC SYSTEM

Vahid Zakeri
Simon Fraser University
Surrey, BC, Canada

ABSTRACT

Magnetorheological (MR) dampers are a new generation of adjustable dampers that can generate large resistive forces against motion with low power. This feature is a unique characteristic that idealizes them for many applications including haptics. A haptic interface system is developed that includes a magnetorheological (MR) damper on the master side and a linear DC motor on the slave side. In this system, the slave motion follows the master and the force experienced by the slave is reflected to the master. The major challenge in this arrangement is control of the MR damper to create the desired resistive force. This paper investigates the different strategies for controlling the MR damper.

In haptics, rather than the perfect tracking of a desired force, it is more convenient to generate a force in the master that is 'similar' to the actual force sensed by the slave. In addition, in many applications of MR dampers such as suspension systems, seismic response reduction, etc., perfect tracking is unneeded. Considering all of these factors, this paper proposes the concept of 'similarity' between the desired force and the resistive force of the MR damper. 'Similarity' means whenever the desired force is large, the resistive force should also be large, and whenever the desired force is small, the resistive force should then be small. Based on this concept, three controllers are applied and tested in both open loop and closed loop schemes: a gain controller, a fuzzy logic controller, and a lookup table controller. The results of these controllers are evaluated and compared to each other with respect to smaller time delays and smaller and smoother control inputs. In the absence of major differences in the results, the closed loop gain controller appeared to be a suitable choice for control because of its simple scheme and its ease of application.
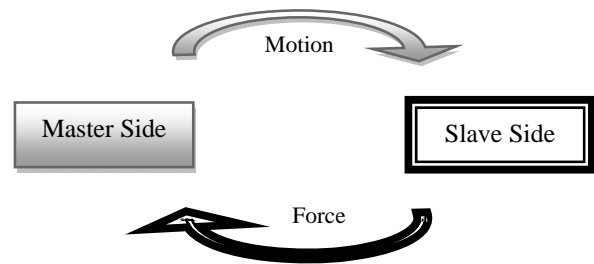
FIGURE 1: THE SCHEMATIC OF A HAPTIC TELEOPERATION SYSTEM

INTRODUCTION

Generally, haptics refers to a technology involving touch interactions between human and real/virtual objects. In this technology, the sense of touch can be created for the user by force, vibration, or motion. Teleoperation is one of the schemes that employs haptics. A 'Haptic-Teleoperative' scheme includes two sides: Master and Slave. Slaves are robotic tools usually located in a remote or dangerous environment. The slave should track all the motions of the master side, and the contact force in the slave side should be followed by the master. Different types of haptic interfaces for teleoperation are surveyed in [1]. Figure 1 displays the block diagram of a haptic teleoperation system, and indicates the direction of the motion/force between the master side and the slave side.

Haptic teleoperation can be applied in different fields. For example, Yoon, *et al.*, have developed an experimental haptic teleoperating system for space robotics to replace the skills of an astronaut, [2]. Yoon, *et al.*, has employed haptic interface to teleoperate a 6-DOF manipulator mounted on the Engineering Test Satellite VII (ETS-VII), [3]. Their results have indicated

the space robot is able to do surface-tracking and peg-in-hole tasks in a real space environment.

Another area that haptic technology can be used is Virtual Reality (VR). Virtual Reality is a technology, in which, a user can interacts with a computer-simulated environment, [4]. Most current VR technologies are based on visual experiences. However, haptics technology can modify them by adding the tactile information to the environment. Burdea, *et al.*, and Lecuyer, *et al.*, have surveyed the haptics and the VR technologies ([5] and [6]).

Haptic technology has a high potential for application in medical sciences such as rehabilitation. For example, haptic technology has been used to help visually impaired patients ([7] and [8]), to improve the performance of hand stroke patients ([9]), to assist with therapy tasks ([10]), etc. Surgery is another, where is useful. For instance, a joystick master with a force feedback mechanism can control slave robotics, [11]. In addition, minimally invasive surgery such as beating heart procedures has been undertaken by employing haptic feedback, [12]. This kind of surgery provides helpful guidance to the surgeon because it constrains the surgeon's hand motions, thereby protecting sensitive structures.

On the other hand, haptic technology can be combined with Magnetorheological Fluids (MR fluids). In order to define MR fluids, we first need to explain Smart fluids. Smart fluids are type of fluids, with properties that can be changed by applying electric or magnetic fields, [13]. An MR fluid is a smart fluid consisting of a suspension of micro-sized magnetic particles in a carrier fluid, [14]. This carrier fluid is usually a type of oil that can increase its apparent viscosity when subjected to a magnetic field. The apparent viscosity can increase to the point that the fluid becomes a viscoelastic solid.

MR fluids are traditionally used in shock absorbers to obtain a controllable damping force. This device is called an MR damper. MR dampers can generate high forces over a broad temperature range and with a simple structure, [15]. However, this force is a hysteretic force, ([15] and [16]). Therefore, MR dampers are considered as nonlinear devices. If all the physical characteristics of MR dampers are considered for modeling, the obtained formulae are too complicated, and thus, not suitable for engineering purposes. As a consequence, based on some physical facts of MR dampers, a simpler model is developed which is called a Semi-Physical or Phenomenological Model, [16]. One of the famous phenomenological models of MR dampers is the Bouc-Wen model, ([15] and [16]). Based on this structure, some modified models are developed, such as the Spencer Model (Spencer *et al.* [17]), and the reduced-parameter model (Ikhouane, [16]).

A combination of haptic and MR fluids has many applications. Ahmadkhanlou, *et al.*, have designed a 5-DOF MR-based robotic arm, that is used to control a remote 5-DOF robot (the slave), [18]. Haptic technology has been employed for transferring to the master those forces encountered in the slave. Ahmadkhanlou, *et al.*, have designed, and constructed a system which can be extended to the telerobotic surgery, [19]. They have also developed a microstructural and kinetic model of MR. This model includes the full range of tactile force exhibited from an MR damper.

In this paper, we have employed MR fluids and haptic technology to a teleoperating system, which consists of an MR damper in the master and a linear motor in the slave. Because this system is a haptic teleoperating system, the force experienced by the slave should be reflected to the master. Even considering the phenomenological model of MR dampers (i.e., Bouc-Wen), the system is still highly nonlinear. Thus, control of this device for creating a desired force is very difficult. This paper has focused on control of an MR damper to create the necessary force for haptic applications. Based on the Bouc-Wen model, the created force is dependent upon the velocity and the displacement of the MR damper as well as the value of the voltage. The control parameters are the voltage or the current.

In our haptic system, rather than the perfect tracking of a desired force, it is more convenient to generate a force in the master that is 'similar' to the actual force sensed by the slave. In addition, in many applications of MR dampers such as suspension systems [20], seismic response reduction (SRR) [21], etc., perfect tracking is unneeded. Considering all these facts, this paper has proposed the concept of 'similarity' between the desired force and the resistive force of the MR damper. Based on this concept, three controllers (i.e., gain, fuzzy logic and lookup table), are applied and tested in both open loop and closed loop schemes. The results of these controllers are evaluated, and compared to each other with respect to smaller time delays and smaller and smoother control inputs.

The rest of this paper is organized as follows: first, the system and experimental test bed are introduced. Then the problem is defined and its complexity is indicated using the Bouc-Wen model of MR dampers. Three controllers, including a gain controller, a fuzzy logic controller, and a lookup table controller are described in the next section of this paper. Finally, this paper concludes in the final section.

EXPERIMENTAL TEST BED

Our experimental setup is composed of a linear DC motor as the master and an MR damper as the slave. A linear motor moves the MR damper with a constant velocity:

Velocity = 4 mm/s

A Celesco SP2 string pot, which is mounted on the MR damper, measures the displacement. The displacement is sent to the data acquisition card (Sensory Model 626). The computer transfers the displacement to the linear motor via a PWM amplifier. This side of control is less challenging, because it is a position control of a linear motor. This position control was performed by a closed loop gain controller. Figure 3 displays the experimental result of the displacement tracking.

On the other hand, the slave is equipped with a force sensor (Sensotec Model 41). The output of the load cell is voltage. This voltage is monitored by the controller developed

by MATLAB/Simulink, and then is amplified and sent to the MR damper (LDC 500CMR). As opposed to the linear motor, developing a controller to reflect the force -sensed by the slave- to the master is a challenging task. This challenge and the proposed solution are explained in detail in the following sections.

PROBLEM DEFINITION

MR dampers can be modeled using Bouc-Wen equations [15]:

$$f(x_r, \dot{x}_r) = k_d x_r + c_d \dot{x}_r + \alpha z \qquad (1)$$

$$\dot{z} = A\dot{x}_r - \gamma|\dot{x}_r z^{n-1}|z - \beta\dot{x}_r|z|^n \qquad (2)$$

In these formulations f represents the MR damper force, and x_r and $\dot{x}_r$ are the displacement and the velocity of the MR damper, respectively. k_d, β, γ, A and n are constant parameters, while α and c_d are dependent upon the applied voltage of the MR damper.

Because the relationship between the voltage and the force in MR dampers are highly nonlinear, creating an exact and accurate desired force is difficult with voltage-control. However, in traditional applications of MR dampers, (Haptic interface system) and many other applications, such as suspension systems and SRR, producing the exact resistive force is not the goal. For those applications, the main concern is creating a 'similar' resistive force in the MR damper. Practically, the word 'similar' means that whenever the stimulating force (i.e., the base excitation in suspension systems or seismic force in SRR systems) is 'large', the resistive force of the MR damper should be 'large'; or whenever the stimulating force is 'small', the resistive force of the MR damper should be 'small'. Generally, this idea can be extended to cover the values 'medium', 'almost small', 'almost large', etc.
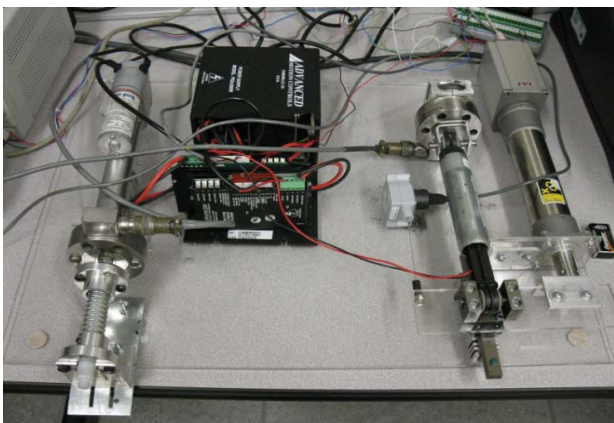
FIGURE 2: THE EXPERIMENTAL TEST BED
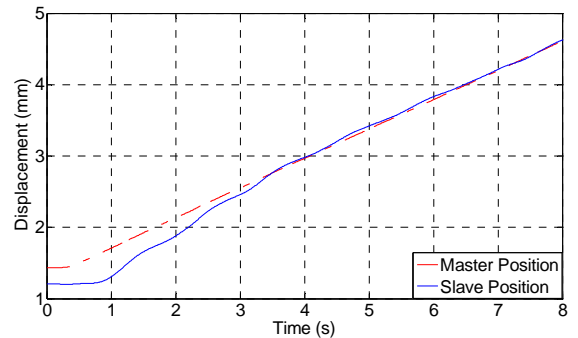
FIGURE 3: THE MASTER AND THE SLAVE DISPLACEMENT IN THE HAPTIC DEVICE
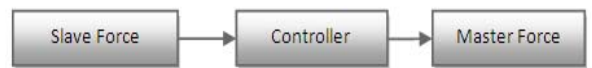
FIGURE 4: THE SCHEMATIC OF AN OPEN LOOP FORCE CONTROL OF THE MR DAMPER
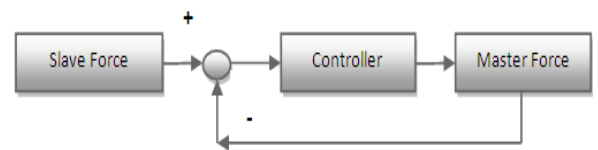
FIGURE 5: THE SCHEMATIC OF A CLOSED LOOP FORCE CONTROL OF THE MR DAMPER
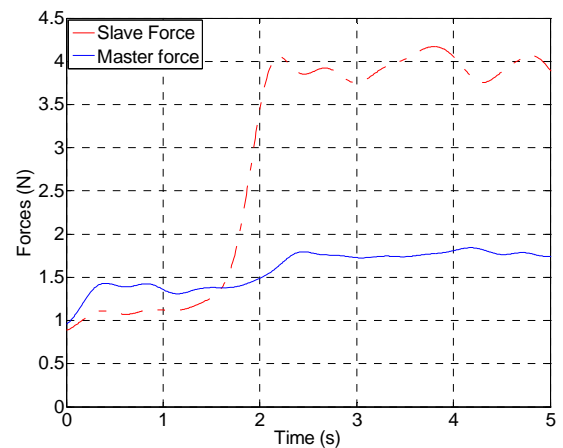
FIGURE 6: MASTER AND SLAVE FORCES IN THE HAPTIC DEVICE USING THE OPEN LOOP GAIN CONTROLLER

In this research, the concept of similarity is employed to generate a force in the master, based on the observed force in the slave. Three appropriate controllers are developed, and

tested: 1) Gain, 2) Fuzzy Logic, and 3) Lookup Table. In addition, all of the controllers have been used in both open loop and closed loop control schemes. Figure 4 and Figure 5 display the schematic of open loop and closed loop control methods, respectively:

GAIN CONTROLLER

The gain controller is the simplest control method, among the rest, for tracking of the force. In the open loop scheme, the force in the slave (Linear Motor) is measured, and then after multiplication by a gain is sent to the master (MR damper) to generate a similar force. A saturation block has been used in order to limit the input of the MR damper because the maximum voltage of the MR damper is limited. Figure 6 displays the master and slave forces. Figure 7 displays the control command of this controller.

For control commands greater than 10, the saturation block limits them to 10 (Figure 7). The time delay for this controller is approximately 0.4 s.

In the closed loop controller, indicated in Figure 5, the force in the slave is measured, and then is compared to the master-force for generating the error signal. Based on this error, the gain controller sends a command to the MR damper to create the appropriate force. Figure 8 and 9 display the result of using this controller:

By comparing Figures 7 and 9, the closed loop control scheme appears to provide a smoother and smaller control command. However, the time delay is approximately 0.6 s, which has a meaningful difference with the open loop case.
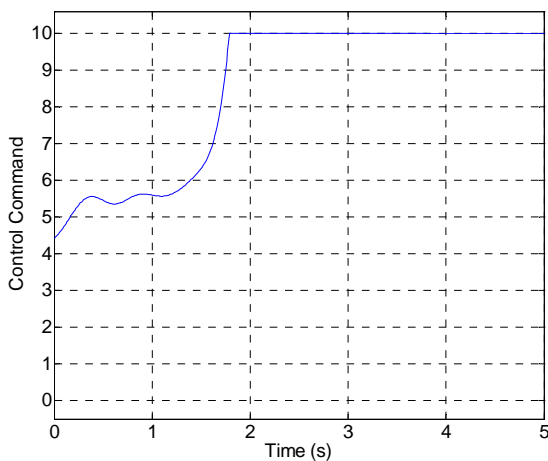
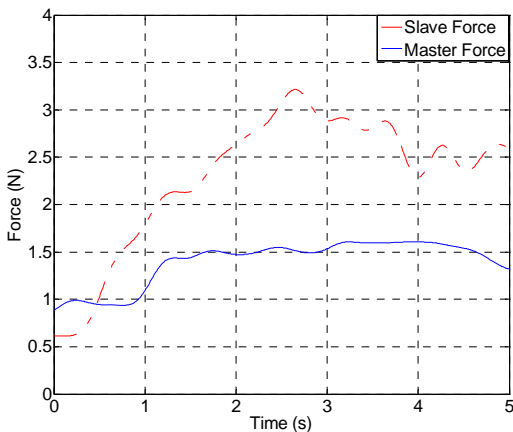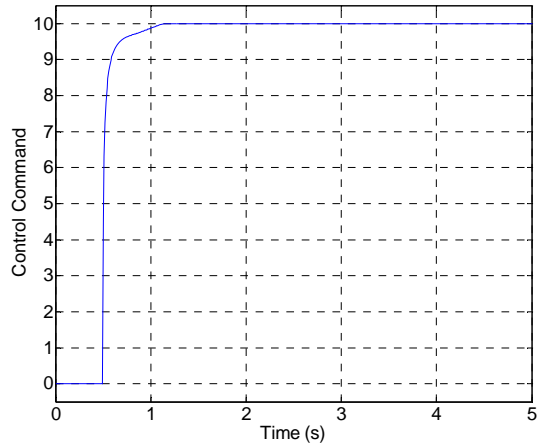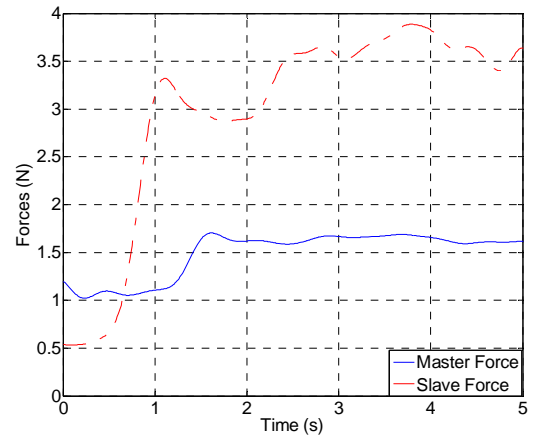**FIGURE 7: THE CONTROL COMMAND IN THE HAPTIC DEVICE USING THE OPEN LOOP GAIN CONTROLLER**
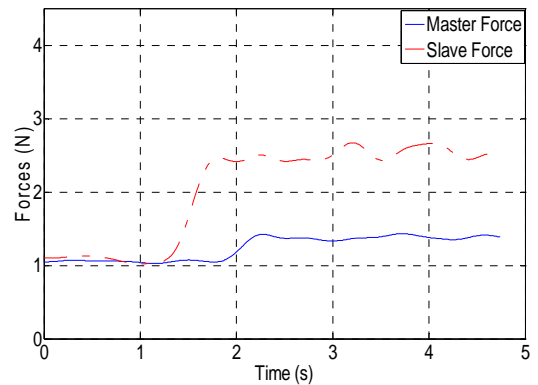
FIGURE 8: MASTER AND SLAVE FORCES IN THE HAPTIC DEVICE USING THE CLOSED LOOP GAIN CONTROLLER
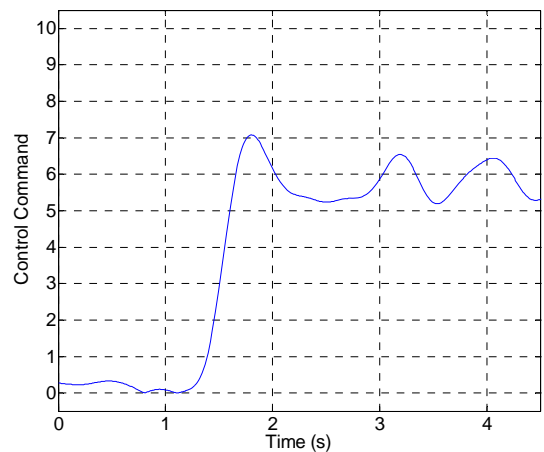
FIGURE 9: THE CONTROL COMMAND IN THE HAPTIC DEVICE USING THE CLOSED LOOP GAIN CONTROLLER
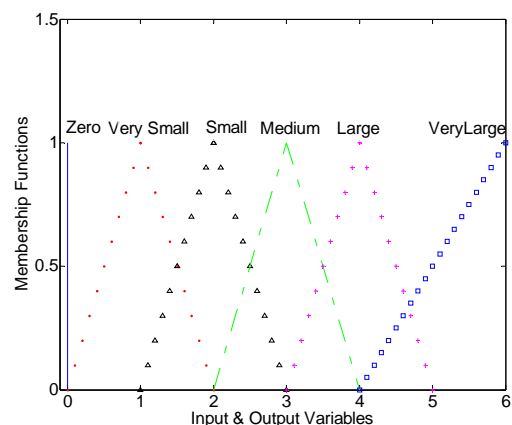
FIGURE 10: MEMBERSHIP FUNCTIONS FOR INPUT AND OUTPUT VARIABLES IN THE FUZZY CONTROLLER

FUZZY LOGIC CONTROLLER

The developed fuzzy controller is a single-input/single-output (SISO) controller. The input and output variables have been subdivided into a range of 'states': 'zero', 'very small', 'small', 'moderate', 'large', and 'very large'. The membership functions for the input and the output of the fuzzy controller are displayed in Figure 11.

The rules of this fuzzy controller are summarized in the table 1.

The measured force in the slave is the set point. A fuzzy logic controller calculates the appropriate voltage based on this set point, and then sends it to the master. The 'Mamdani' method is used for the fuzzification, and the 'centroid' method is employed for the defuzzification. Figure 11 displays the master/slave forces. Figure 12 displays the control command of this controller.

Figure 11 indicates that whenever the slave force is increased, the master force will also be increased. This sequence involves a time delay, because first, the force is measured in the slave, and then a force will be created in the master (based on the slave force). The time delay for this control scheme is approximately 0.6 s.

Figure 13 displays the master/slave forces. Figure 14 displays the control command of this controller.

Rules
If Input is Zero then output is Zero
If Input is Very Small then output is Very Small
If Input is Small then output is Small
If Input is Medium then output is Medium
If Input is Large then output is Large
If Input is Very Large then output is Very Large

TABLE 1: SET OF RULES FOR THE FUZZY LOGIC CONTROLLER

FIGURE 11: MASTER AND SLAVE FORCES IN THE HAPTIC DEVICE USING THE OPEN LOOP FUZZY LOGIC CONTROLLER

FIGURE 12: CONTROL COMMAND IN THE HAPTIC DEVICE USING THE OPEN LOOP FUZZY LOGIC CONTROLLER

FIGURE 13: MASTER AND SLAVE FORCES IN THE HAPTIC DEVICE USING THE CLOSED LOOP FUZZY LOGIC CONTROLLER
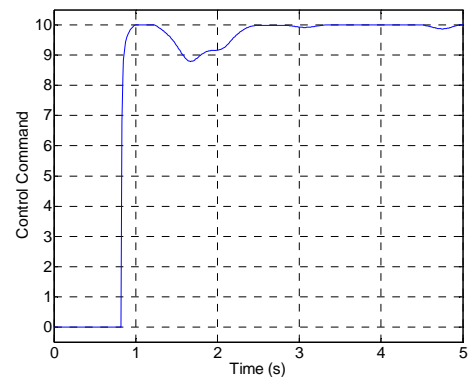

FIGURE 14: CONTROL COMMAND IN THE HAPTIC DEVICE USING THE CLOSED LOOP FUZZY LOGIC CONTROLLER

By comparing Figure 13 and Figure 14, the closed loop control scheme appears to provide more 'control levels' (0 9 and 10 in Figure 14), while open loop control has provided two 'control levels' (0 and 10 in Figure 12). Generally, whenever a controller can provide more levels of control, it can provide more levels of similarity. Therefore, the closed loop scheme has provided more levels of similarity. However, the time delay in this case is approximately 0.85 s, which is greater than the open loop scheme.

LOOKUP TABLE

Another suitable controller, for applying the 'similarity' idea, is the Lookup Table. This controller produces different constant values based on the input intervals. Table 2 summarizes this lookup table.

The results for the open loop controller are displayed in Figure 15 and Figure 16. Figure 17 and Figure 18 display the results for the same experiments using the closed loop controller.

By comparing Figure 16 and 18, we observe that both control schemes (open loop and closed loop) have almost the same 'control level' (0 5 and 10). However, the duration of level 5 in the closed loop (Figure 18) is more than the open loop (Figure 16). In addition, the time delay in the open loop case is approximately 0.6 s (while it was 0.77 s in closed loop case).

CONCLUSIONS

A haptic interface system was employed, using an MR damper in the master and a linear motor in the slave. The experimental results indicated the slave followed the master motion accurately. However, transferring the force from slave to the master was the major challenge because of the nonlinear nature of MR dampers. Three control methods, in both open loop and closed loop schemes were employed and compared to each other. Generally, closed loop schemes provided smoother and smaller control inputs while open loop schemes provided smaller time delays. Table 3 summarizes the time delays of these controllers.

Input Intervals	Output Constants
[0 1)	0
[1 2)	1
[2 3)	2
[3 4)	3
[4 5)	4
[5 6]	5

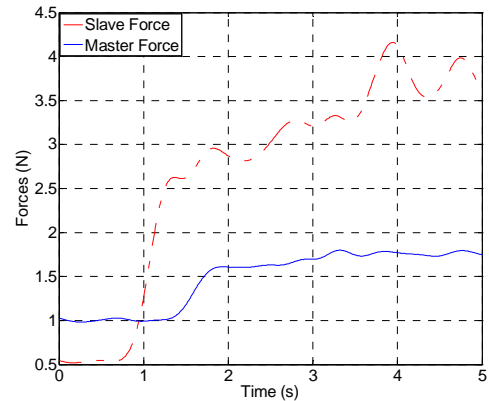**TABLE 2: INPUT INTERVALS AND OUTPUT CONSTANTS FOR THE LOOKUP TABLE**

FIGURE 15: MASTER AND SLAVE FORCES IN THE HAPTIC DEVICE USING THE OPEN LOOP LOOKUP TABLE CONTROLLER
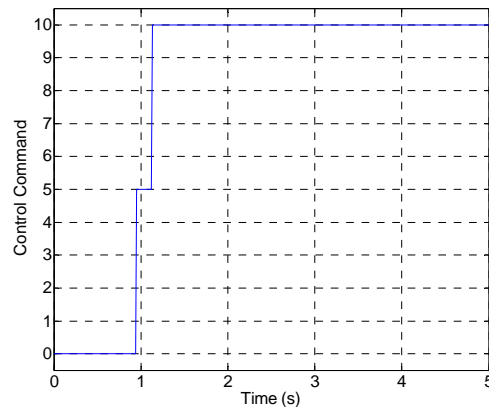

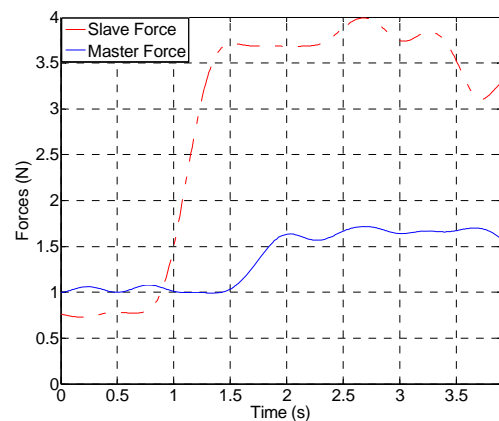**FIGURE 16: CONTROL COMMAND IN THE HAPTIC DEVICE USING THE OPEN LOOP LOOKUP TABLE CONTROLLER**

FIGURE 17: MASTER AND SLAVE FORCES IN THE HAPTIC DEVICE USING THE CLOSED LOOP LOOKUP TABLE CONTROLLER
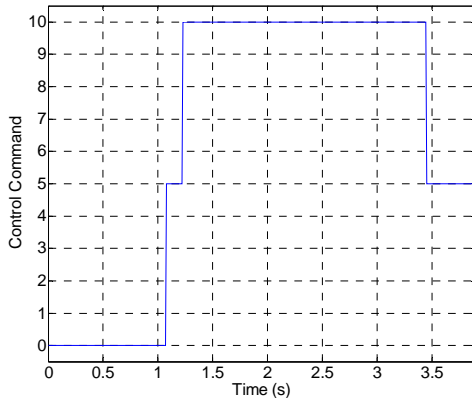
FIGURE 18: CONTROL COMMAND IN THE HAPTIC DEVICE USING THE CLOSED LOOP LOOKUP TABLE CONTROLLER

Control Method	Time Delay Value (s)
Open Loop Gain	0.4
Closed Loop Gain	0.6
Open Loop Fuzzy	0.6
Closed Loop Fuzzy	0.85
Open Loop Lookup Table	0.6
Closed Loop Lookup Table	0.77

TABLE 3: THE COMPARISON OF TIME DELAYS BETWEEN DIFFERENT CONTROLLERS

In the absence of major differences in the results, the closed loop gain controller appears to be a suitable choice for control because of its simple scheme and its ease of application.

ACKNOWLEDGMENTS

REFERENCES

[1] "Haptic interfaces for teleoperation and virtual environments", Hollerbach, J.M. (Dept. of Comput. Sci., Utah Univ., Salt Lake City, UT, USA); Jacobsen, S.C. Source: First Workshop on Simulation and Interaction in Virtual Environments, 1995, I-VI

[2] "Model-based robot teleoperation with haptic interface", Woo-Keun Yoon (Intelligent Syst. Inst., Japan); Tsumaki, Y.; Uchiyama, M., Industrial Robot, v 30, n 6, 2003, 584-91

[3] "Model-based teleoperation of a Space robot on ETS-VII using a haptic interface", Woo-Keun Yoon (Graduate Sch. of Eng., Tohoku Univ., Sendai, Japan); Goshozono, T.; Kawabe, H.; Kinami, M.; Tsumaki, Y.; Uchiyama, M.; Oda, M.; Doi, T., Proceedings 2001 ICRA, IEEE International Conference on Robotics and Automation (Cat. No.01CH37164), 2001, 407-12 vol.1

[4] "http://en.wikipedia.org/wiki/Virtual_reality"

[5] "Virtual reality technology - an introduction" ,Burdea, G. (Human-Machine Interface Lab., Rutgers Univ., New Brunswick, NJ, USA) Source: 2007 IEEE Virtual Reality Conference, 2007, 1 pp.

[6] "Integration of haptics in virtual environments - a perception-based approach", Lecuyer, A. (INRIA/IRISA, France); Harders, M.; Niemeyer, G.; Jones, L.A.; Otaduy, M.A.; Pai, D.K.

[7] "Haptic visualization", Oakley, I. (Dept. of Comput. Sci., Glasgow Univ., UK); Brewster, S.; Glendye, A.; Masters, M.M., Human-Computer Interaction INTERACT'99. IFIP TC.13 International Conference on Human-Computer Interaction, 1999, 97-8 vol.2

[8] "Haptic interactions", Oakley, I. (Dept. of Comput. Sci., Glasgow Univ., UK); Crosson, A.; Mcgee, M.; Wai Yu; Brewster, S.; Gray, P., 15th Annual Conference of the Human-Computer Interaction Group of the British Computer Society. IHM-HCI 2001, 2001, 255-6

[9] "Haptic rehabilitation exercises performance evaluation using fuzzy inference systems", Barghout, A. (Res. Lab., Multimedia Commun., Ottawa, ON, Canada); Alamri, A.; Eid, M.; El Saddik, A., 2008 IEEE International Workshop on Medical Measurement and Applications (MeMeA '08), 2008, 13-18, ISBN-10: 978-1-4244-1937-1

[10] "Telerobotic haptic system to assist the performance of occupational therapy tests by motion-impaired users", Pernalete, N. (Dept. of Electr. & Comput. Eng., Western Michigan Univ., Kalamazoo, MI, USA); Wentao Yu; Dubey, R.; Moreno, W.A., 2003 IEEE International Conference on Robotics and Automation (Cat. No.03CH37422), 2003, 1247-52 vol.1

[11] "Force Feedback Control of Robotic Forceps for Minimally Invasive Surgery", Ishii Chiharu, Kamei, Yusuke, AIP Conference Proceedings; 6/12/2008, Vol. 1019 Issue 1, p154-155

[12] "Dynamic 3-D Virtual Fixtures for Minimally Invasive Beating Heart Procedures", Jing Ren, Patel Rajni V., McIsaac Kenneth A., Guiraudon Gerard, Peters Terry M., IEEE Transactions on Medical Imaging; Aug2008, Vol. 27 Issue 8, p1061-1070, 10p, 4 charts, 11

[13] "http://en.wikipedia.org/wiki/Smart_fluid"

[14] "http://en.wikipedia.org/wiki/Magnetorheological_fluid"

[15] "Vehicle Suspension Vibration Control with Magnetorheological Dampers", Yu Shen, PhD Thesis, Waterloo, Ontario, Canada, 2005

[16] "Systems with hysteresis: analysis, identification and control using the Bouc-Wen model", Fayçal Ikhouane, José Rodellar, Chichester, England; Hoboken, NJ: John Wiley, c2007

[17] "Phenomenological Model of a Magnetorheological Damper", B. F. Spencer, Jr., S. J. Dyke, M. K. Sain and J. D. Carlson, Journal of Engineering Mechanics, ASCE, 123. pp. 230-238, 1997

[18] "The development of a five DOF magnetorheological fluid-based telerobotic haptic system", Ahmadkhanlou, F. (Ohio State Univ., Columbus, OH, USA); Washington, G.N.; Bechtel, S.E., Proceedings of the SPIE - The International Society for Optical Engineering, v 6926, 27 March 2008, 692604-1-10

[19] "The development of variably compliant haptic systems using magnetorheological fluids", Ahmadkhanlou, F. (Dept. of Mech. Eng, Ohio State Univ., Columbus, OH, USA); Washington, G.N.; Yingru Wang; Bechtel, S.E., Proceedings of the SPIE - The International Society for Optical Engineering, v 5757, n 1, 2005, 491-502

[20] "Semi-active Vibration Control Schemes for Suspension Systems Using Magnetorheological Dampers", Y. Shen, M.F. Golnaraghi, G. R. Heppler, Journal of Vibration and Control 2006; 12; 3

[21] "Semi-Active Control Strategies for MR Dampers: A Comparative Study", Laura M. Jansen, Shirley J. Dyke, ASCE Journal of Engineering Mechanics, Vol. 126, No. 8, pp.795–803

Vahid Zakeri received his B.Sc. and M.Sc. in control engineering from Shiraz University, Shiraz/Iran, in 2005 and 2008, respectively. He is currently a Ph.D. student with the School of Engineering Science, Simon Fraser University, Vancouver, BC, Canada. His current research interests include vibration analysis, haptics, dental drills, robotics and control theory.

On Evaluation of Multi-camera Calibration Methods

Xiaochen Dai

Abstract—**Camera calibration plays a critical role in building multi-camera tracking systems since the level of accuracy of calibration results determines the relationships between both qualitative and quantitative tracking results. In past many years, there has been a number of works proposed for camera calibration. In this paper, we compare and analyze the camera calibration methods in a multi-camera system using 2D checkerboard patterns and multiple spheres. The approach here is by putting the calibration objects in the common view of both cameras, and we design algorithms to capture cameras' intrinsic and extrinsic parameters. Thereafter, we use the computed extrinsic parameters and known geometries of multiple views to compute the transformation between camera frames. Experimental comparison and error analysis for different calibration methods are also discussed.**

Index: Camera calibration, Multi-camera, Intrinsic parameters, Extrinsic parameters, Checkerboard, Sphere

I. INTRODUCTION

As a modern technique of computer vision, multi-camera system has nowadays been widely used in various aspects of everyday life such as people surveillance, public security, medical monitoring and traffic commanding. Interesting topic has also been contrived to use PTZ (Pan, Tilt, and Zoom) cameras to get details of objects by using one camera with broad view to detect certain object in a passive multi-camera system. In addition, multiple cameras have also been used to track gestures and body language to trigger events such as turning on the lights.

In order to build a multi-camera visual tracking system, we must know the geometrical relations among all the cameras in the system, so that each one knows exactly where the others are. A typical multi-camera system consisting of two cameras is shown in Figure 1. Through a complete camera calibration, which includes recovery of both intrinsic and extrinsic parameters, we can obtain the transformation information of all camera frames with respect to each other, this enables the cameras in the tracking system to localize the objects accurately and follow it cooperatively.

In past decades, different calibration methods have been discussed in a large number of papers. Traditional methods, such as Tsai's two-stage calibration [1] and Zhang's planar pattern calibration [2], are using special patterns with known metric structure. Later on, calibration with spherical objects has been implemented in [3], [4], [5], [6] adopting

X. Dai is with Experimental Robotics Lab (ERL), School of Engineering Science, Simon Fraser University, 8888 University Drive, Burnaby, British Columbia, V5A 1S6, Canada xda6@sfu.ca
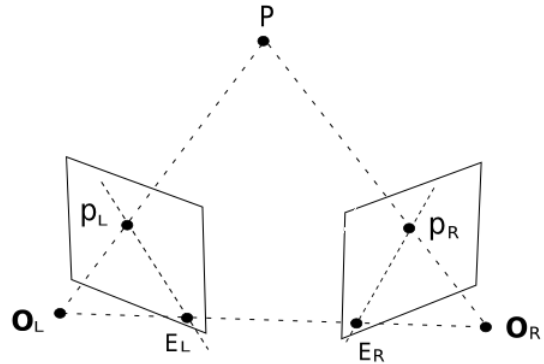

Figure 1. Geometry of a multi-camera system

either nonlinear approaches or linear approaches. In addition, some work has also been done for a multi-camera system in particular. [7] has introduced a dumbbell calibration object consisting of two spheres, for a multi-camera system and [8] has investigated the possibility of multi-camera calibration using 1D objects undertaking general rigid motion.

The advantages and disadvantages of all these methods have been an important topic among scholars over the world for long. A few papers have analyzed and compared various calibration methods. [9] has compared Faugeras-Toscani, Zhang's and a robust calibration algorithm while [10] has discussed the Tsai's, Zhang's and the typical linear calibration. However, these comparisons focus mainly on single camera calibration. Besides, few have included the calibration using spheres, which are supposed to be ideal objects for multi-cameras as spheres are visible in each view.

In this paper, we present the methods to calibrate a multi-camera system by using 2D pattern and multiple spheres as objects. Then we compare the results obtained with physical truth and discuss experimental errors and limitation of each method.

The rest of the paper is organized as follows. Section II introduces the camera model and the principles of calibration methods discussed in this paper. In Section III, we report the experimental results. Finally, in section IV we compare the results obtained with physical truth and analyze the errors.

II. THE CALIBRATION METHODS

A. *Pinhole Camera Model*

 A real world camera can be approximated by a pinhole camera model as shown in Figure 2, which is widely used in computer vision. With this model, the relationship between a

3D point with homogeneous coordinates $M = (X, Y, Z, 1)^T$
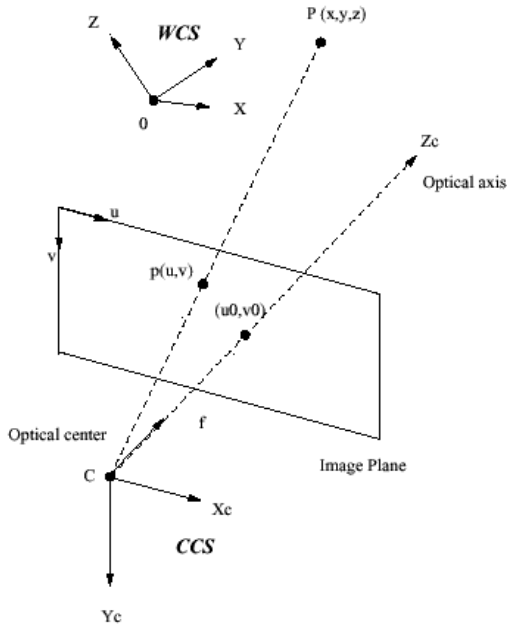
Figure 2. Model of a pinhole camera

and its 2D image point $m = (u, v, 1)^T$ is given by

$$m = K[R|t]M \text{ , with } K = \begin{bmatrix} f_x & s & u_0 \\ 0 & f_y & v_0 \\ 0 & 0 & 1 \end{bmatrix} \quad (1)$$

$P = K[R|t]$ is called the camera matrix; K is called the camera intrinsic matrix, with $f_x = fm_x$ and $f_y = fm_y$ represent the focal length of the camera in terms of pixel dimensions in the x and y direction respectively; m_x and m_y are numbers of pixels per unit distance in each direction; S denotes the scale factors in the image u and v axes; (u_0, v_0) is the principal point; $[R|t]$, called the extrinsic matrix, is the rotation and translation which relates the world coordinate frame to the camera coordinate frame.

B. Calibration with 2D Checkerboard

Zhang's method [2] described an algorithm that requires the camera to observe a planar pattern shown at a few (at least two) different views. The motion of the pattern need not be known. Thus, an object point M and its image m are related by a homograph $P = K[R|t] = \lambda K[r_1, r_2, r_3, t]$. In order to recover intrinsic and extrinsic parameters, firstly, given an image of the checkerboard plane, which is on $Z=0$ of the world coordinate frame, a homograph can be estimated. Then the maximum likelihood estimation of P is obtained by minimizing the following equation.

$$\begin{bmatrix} M^T & 0^T & -uM^T \\ 0^T & M^T & -vM^T \end{bmatrix} P = 0 \quad (2)$$

Secondly, let $P = \begin{bmatrix} p_1^T & p_2^T & p_3^T \end{bmatrix}^T$. Given one homograph,

the two basic constraints on the intrinsic parameters are as following:

$$p_1 K^{-T} K^{-1} p_2 = 0 \quad (3)$$

$$p_1 K^{-T} K^{-1} p_1 = p_2 K^{-T} K^{-1} p_2 , \quad (4)$$

Let,

$$B = K^{-T} K^{-1} . \quad (5)$$

Symmetric matrix B is defined by a 6D vector b.

$$b = \begin{bmatrix} B_{11} & B_{12} & B_{22} & B_{13} & B_{23} & B_{33} \end{bmatrix}^T \quad (6)$$

Matrix v is known from P and a unique solution b can be obtained according to the following equation.

$$\begin{bmatrix} v_{12}^T \\ (v_{11} - v_{22})^T \end{bmatrix} b = 0 \quad (7)$$

Once b is estimated, we can compute the intrinsic matrix K from Equation (5).

$$v_0 = (B_{12} B_{13} - B_{11} B_{23})/(B_{11} B_{22} - B_{12}^2) \quad (8)$$

$$\lambda = B_{33} - [B_{13}^2 + v_0(B_{12} B_{13} - B_{11} B_{23})]/B_{11} \quad (9)$$

$$f_x = \sqrt{\lambda / B_{11}} \quad (10)$$

$$f_y = \sqrt{\lambda B_{11} /(B_{11} B_{22} - B_{12}^2)} \quad (11)$$

$$s = -B_{12} f_x^2 f_y / \lambda \quad (12)$$

$$u_0 = s v_0 / f_x - B_{13} f_x^2 / \lambda \quad (13)$$

Once K is known, the extrinsic parameters for each image are easily computed. We have

$$r_1 = \lambda K^{-1} h_1 \quad (14)$$

$$r_2 = \lambda K^{-1} h_2 \quad (15)$$

$$r_3 = r_1 \times r_2 \quad (16)$$

$$t = \lambda K^{-1} h_3 \quad (17)$$

Suppose that our multi-camera system is composed of two cameras with relatively fixed camera coordinate frames. As shown in Figure 3, the problem is to solve the relative transformation between camera frames. Here we have introduced a world coordinate frame. Let R_1, T_1 and R_2, T_2 be the extrinsic parameters of camera one and camera two, respectively. A point in the space has coordinates V_w, V_1, and V_2 in the world frame, camera one frame, and camera two frame, respectively. As the extrinsic parameters determine the transformation between each camera frame with the same world coordinate frame, we have

$$V_1 = R_1 V_w + T_1$$

$$V_2 = R_2 V_w + T_2$$

Therefore, we can derive the geometrical relation between two camera frames as following equation.

$$R = R_1 R_2^{-1} \quad (18)$$

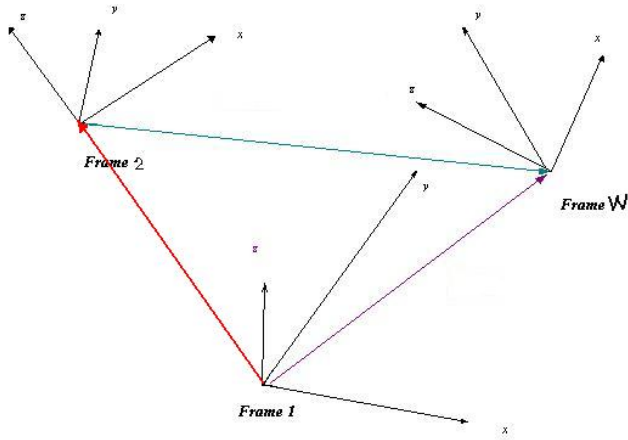$$T = T_1 - R_1 R_2^{-1} T_2 \quad (19)$$

Figure 3. Relative transformation between camera frames

C. Calibration with Multiple Spheres

The dumbbell calibration [7] proposed a novel method to calibrate multiple cameras using spheres. For each camera, three spheres are needed to calculate the intrinsic and extrinsic parameters. A complete camera calibration can be divided into three steps.

Firstly, we extract image conics from raw images. As shown in Figure 4, we set the origin of the world frame (red) coincide with the origin of the camera frame (blue). The surface of the smallest cone containing the sphere is a quadric Q (green). The intersection of the quadric Q with the image plane is an image conic C. As a ellipse, the image conic satisfies the equation

$$m^T C m = 0 \qquad (20)$$

Where C satisfies

$$C = \begin{bmatrix} a & b & d \\ b & c & f \\ d & f & g \end{bmatrix}$$

Conics are extracted on features of spheres such as color and shape. Six points are selected from each ellipse to determine the six unknowns in C.
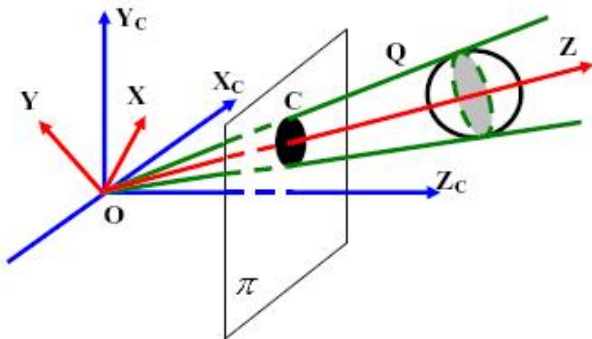
Figure 4. Projection of a sphere onto the image plane as an image conic C.

Secondly, we recover the intrinsic parameters. For Calibration using spheres, the origin of world coordinate frame coincides with the origin of the camera frame, so that (1) is changed to

$$m = KR[I|0]M = KRM$$

$$M = R^{-1}K^{-1}m \qquad (21)$$

Let us define the Image of the Absolute Conic (IAC) as

$$\omega = K^{-T}K^{-1} \qquad (22)$$

Let C_i be the conic extracted from the sphere i. We have

$$\kappa_i C_i = \omega - v_i v_i^T \qquad (23)$$

$$v_i = \sqrt{1 + \tan^2 \alpha_i}\, K^{-T} r_3 \qquad (24)$$

where κ_i is a scalar unknown; $2\alpha_i$ is the aperture of the cone and r_3 is the third column of R. From (23), we have

$$v_i v_i^T = \begin{bmatrix} \omega_{11} - \kappa_i a_i & \omega_{12} - \kappa_i b_i & \omega_{13} - \kappa_i d_i \\ \omega_{12} - \kappa_i b_i & \omega_{22} - \kappa_i c_i & \omega_{23} - \kappa_i f_i \\ \omega_{13} - \kappa_i d_i & \omega_{23} - \kappa_i f_i & \omega_{33} - \kappa_i g_i \end{bmatrix} \qquad (25)$$

The rank of matrix $v_i v_i^T$ is one. From the definition of rank, we know that all 2×2 submatrices of $v_i v_i^T$ are singular, which means their determinants are zero. We can obtain

$$(\omega_{11} - \kappa_i a_i)(\omega_{22} - \kappa_i c_i) = (\omega_{12} - \kappa_i b_i)^2$$
$$(\omega_{11} - \kappa_i a_i)(\omega_{33} - \kappa_i g_i) = (\omega_{13} - \kappa_i d_i)^2 \qquad (26)$$
$$(\omega_{22} - \kappa_i c_i)(\omega_{33} - \kappa_i g_i) = (\omega_{23} - \kappa_i f_i)^2$$

From each combination of two conics, we obtain linear equations by eliminating the second order components of ω in (26). Thus, a set of linear equations is obtained for ω, we solve the least square solution using singular value decomposition. Once we have ω, the intrinsic parameter K can be obtained by applying Cholesky decomposition.

The third step is to recover the extrinsic parameters. Once we solve ω, v_i can be computed from (23). Also, we got

$$\sqrt{1 + \tan^2 \alpha_i}\, r_3 = K^T v_i \qquad (27)$$

The rotation matrix R satisfies $RR^T = I$ and can be decomposed into three components: rotate ϕ about z axis, then rotate θ about the new y axis, and finally rotate ψ about the new z axis. Therefore,

$$r_3 = [\cos\phi_i \sin\theta_i \quad \sin\phi_i \sin\theta_i \quad \cos\theta_i]^T$$

The camera frames are fixed and the world coordinate frame differs from sphere to sphere. Three unknowns α_i, ϕ_i, and θ_i can be solved for each sphere, and ψ_i can be solved from $R_i R_i^T = I$. If the diameter of sphere is D_i, the sphere center in its world coordinate frame is

$$M_i = \begin{bmatrix} 0 & 0 & \sqrt{1 + \tan^2 \alpha_i}\, D_i / 2 & 1 \end{bmatrix}^T \qquad (28)$$

Thus, the coordinates of the sphere center in the camera frame are

$$M_{ci} = \begin{bmatrix} R_i & 0 \\ 0 & 1 \end{bmatrix} M_i \qquad (29)$$

Now we may obtain the coordinates of the sphere center in both camera frames. We denote R_c, t_c as the rotation and translation parameters between two camera frames. We have

$$M_{ci2} = \begin{bmatrix} R_c & t_c \\ 0 & 1 \end{bmatrix} M_{ci1} \qquad (30)$$

When cameras are placed on parallel planes, there is only one unknown rotational angle for R_c. Under this circumstance, it is easy to compute R_c and t_c.

III. EXPERIMENTAL STUDIES

The multi-camera system consists of two CCD cameras (Sony EVI D100), which are located on a horizontal planar white board with meshes of known dimension (Figure 5). The two cameras are positioned and their camera frames are fixed. Both cameras have the same focal length and their image resolution is 640×480.

Before we implement algorithms to calibrate the system, we measured the relative distance and orientation between two cameras. The left camera in Figure 5 is defined as camera one while the right one is defined as camera two. Based on the measurements, we have the following rotation matrix and translation vector between camera coordinate frames.

$$R_c = \begin{bmatrix} 0.707 & 0 & -0.707 \\ 0 & 1 & 0 \\ 0.707 & 0 & 0.707 \end{bmatrix}$$

$$t_c = \begin{bmatrix} 40cm \\ 0cm \\ 10cm \end{bmatrix}$$

Then we have designed algorithms based on the two calibration procedures introduced above to compute the intrinsic and extrinsic parameters of the cameras. Applying the calibration extrinsic parameters and the knowledge on multiple geometries, we, thereafter, compute the rotation matrix and translation vector between the camera coordinate frames. Finally, the accuracy of different methods can be analyzed by comparing computed results with respect to the physical ground truth.

A. 2D Planar Calibration Method

We made a planar checkerboard as our calibration object for the planar calibration. The checkerboard is composed of a pattern with 9×7 alternate black and white squares. There are totally 48 corners used as object points. All the squares have same dimension of $3cm \times 3cm$. We define the upper left corner of the checkerboard as the origin and the plane of board as $Z=0$ plane of the world reference frame. Following these criteria, we can obtain the coordinates of the corners in world frame.

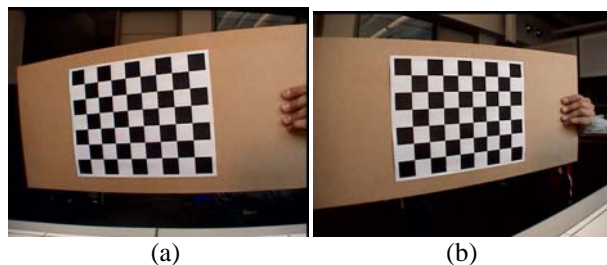Figure 5. View of the experimental set-up

(a)   (b)

Figure 6. Images captured by two cameras with the planar pattern checkerboard. In this case, actual angle between two views is $45°$, and $t_c = \begin{bmatrix} 40 & 0 & 10 \end{bmatrix}^T cm$. (a) Image captured by camera one. (b) Image captured by camera two.

During the calibration process, we use two cameras to capture images of the checkerboard from different views at an interval. The checkerboard can be freely moved in the space and the motion could remain unknown as long as the checkerboard remains within the common view of both cameras. Considering all the 48 corners clearly captured and shown on the images, the algorithm then extracts corner information such as coordinates in the image plane. We have used twenty images in our experiment with the checkerboard under different orientations. Figure 6 displays one of the twenty pairs of the images detected by the two cameras. Usually, the more images are used, the more accurate the result is obtained.

Table I. Intrinsic parameters for planar pattern calibration

	f_x	f_y	s	u_0	v_0
Cam 1	570.07	570.89	0.21	343.03	211.99
Cam 2	563.68	561.37	0.19	325.16	238.64
Ground	592.12	558.56	0.00	320.00	240.00

Table I shows the intrinsic parameters computed by using the planar checkerboard. Ground truth obtained from the camera specifications manual is shown for reference.

Table II. Extrinsic parameters for planar pattern calibration

	R(v)	t
Cam 1	[-2.01, -2.14, 0.67]	[-10.33, -7.18, 54.88]
Cam 2	[1.97, 1.99, 0.11]	[-7.14, -10.81, 51.13]

Table II shows the extrinsic parameters for planar pattern calibration. R(v) is a rotation vector which can be converted to the rotation matrix. t is the translation vector. According to (18) and (19), we compute

$$R_c = \begin{bmatrix} 0.71 & 0 & -0.69 \\ 0 & 1.06 & 0 \\ 0. & 0 & 0.76 \end{bmatrix}$$

$$t_c = \begin{bmatrix} 41.13 \\ 1.55 \\ 9.30 \end{bmatrix}$$

B. Multiple Spheres Calibration Method

We use a bright yellow sphere of diameter of 6cm as the calibration object. For every single camera, three images with the sphere at different locations are needed in order to recover the intrinsic and extrinsic parameters. The relative locations of cameras and spheres are shown in Figure 7. In Figure 8, we show the images captured by two cameras with the sphere at three different locations.

By using the linear approach, camera parameters are recovered. Table III shows the intrinsic parameters of both cameras with the actual ground truth given for reference. Table IV shows the coordinates of the sphere center in world frame and camera frame with respect to the two cameras. Here S_i represents the sphere in location i; M_i represents the world frame coordinate of sphere in location i and M_{ci} represents the camera frame coordinate of sphere in location i.

Table III. Intrinsic parameters for sphere calibration

	f_x	f_y	s	u_0	v_0
Cam 1	550.64	528.25	0.11	313.03	223.03
Cam 2	544.31	526.62	0.30	315.18	213.75
Ground	592.12	558.56	0.00	320.00	240.00

(a)   (b)   (c)
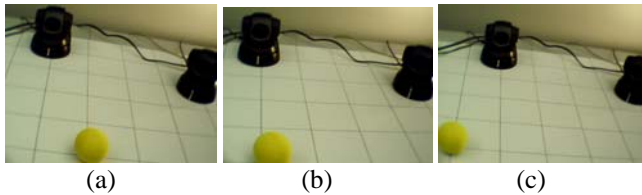
Figure 7. Images showing relative location of cameras and spheres in three cases. (a) sphere at location one. (b) sphere at location two. (c) sphere at location three

(a)   (b)

(c)   (d)

(e)   (f)
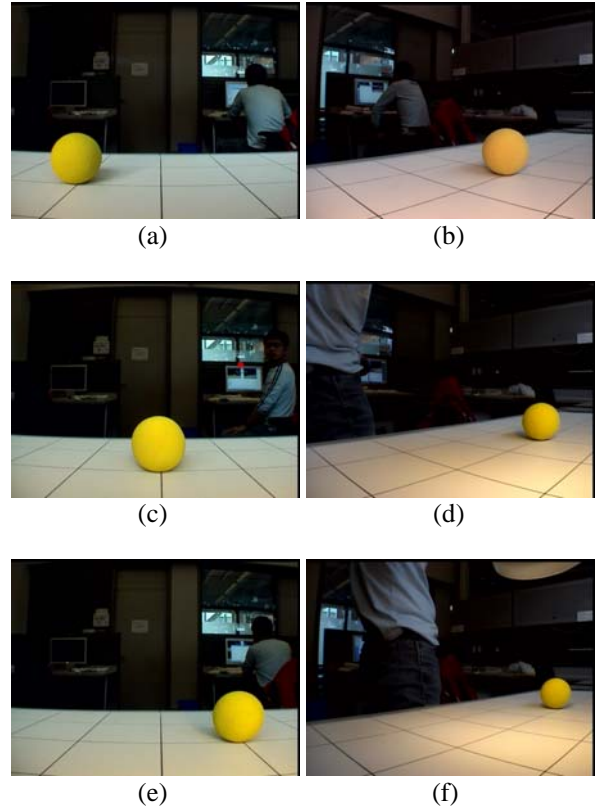
Figure 8. Images of spheres at different locations captured by two cameras. Actual angle between two views is $45°$, and $t_c = \begin{bmatrix} 40 & 0 & 10 \end{bmatrix}^T cm$. (a) sphere at location one captured by camera one. (b) sphere at location one captured by camera two. (c) sphere at location two captured by camera one. (d) sphere at location two capture by camera two. (e) sphere at location three captured by camera one. (f) sphere at location three captured by camera two.

Table IV. Extrinsic parameters for sphere calibration

	S_1	S_2	S_3
M_1	[0, 0, 42]	[0, 0, 41]	[0, 0, 41]
M_{c1}	[9, -5, 41]	[0, -4, 40]	[-11, -5, 42]
M_2	[0, 0, 45]	[0, 0, 51]	[0, 0, 62]
M_{c2}	[-12, -3, 39]	[-16, -4, 47]	[0, -22, 56]

According to (30), we compute

$$R_c = \begin{bmatrix} 0.68 & 0 & -0.75 \\ 0 & 0.96 & 0 \\ 0. & 0 & 0.69 \end{bmatrix}$$

$$t_c = \begin{bmatrix} 38.29 \\ 2.16 \\ 8.18 \end{bmatrix}$$

IV. DISCUSSIONS AND CONCLUSIONS

Through theoretical analysis and practical experiment, we present the 2D planar pattern calibration method and the multiple spheres calibration method for a multi-camera system. The planar pattern method uses nonlinear optimization to obtain camera intrinsic and extrinsic parameters; thus, reduces the error. The multiple spheres calibration recovers camera parameters by adopting a linear approach to solve constraints on IAC.

From the result obtained, we conclude that the planar pattern method has better practicability to recover camera parameters. However, for a multi-camera system, it certainly has some restrictions. For example, during our experiment, we found it not so easy to extract corner information from the images of both cameras simultaneously. So in the case that cameras in the system have very small or even no common view, the planar patter method does not work successfully.

In comparison, the multiple spheres method is theoretically considered to be ideal for systems consisting of multiple cameras because spheres remain a consistent appearance in all directions, but this method also has unsolved problems. In our experiment, we extract conic parameters applying a color and shape separation method, which has pretty high requirement for images taken. Illumination and hue are of great concern and thus should be taken into consideration. Otherwise, the conics extracted are not accurate enough to recover camera parameters and that leads to errors.

The results show that the errors remain close, although both calibration methods have limitations. Some future work may focus on calibrating multiple moving cameras with multiple objects fixed in the space. We will also carry on research on event tracking and motion recognition for a system containing both static and moving cameras.

REFERENCES

[1] R. Y. Tsai, "A Versatile Camera Calibration Technique for High-accuracy 3D Machine Vision Metrology Using Off-the-shelf TV Cameras and Lenses", IEEE Journal of Robotics and Automation, Vol. RA-3, No. 4, pp. 323-344, Aug, 1987

[2] Z. Zhang, "A Flexible New Technique for Camera Calibration", IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol. 22, No. 11, pp. 1330-1334, Nov 2000

[3] H. Teramoto, G. Xu, "Camera Calibration by a Single Image of Balls: From Conics to the Absolute Conic", The 5th Asian Conference on Computer Vision, pp. 499-506, Jan 2002

[4] M. Agrawal, L. S. Davis, "Camera Calibration Using Spheres: A Semi-definite Programming Approach", IEEE 9th International Conference on Computer Vision, pp. 782-791, Oct 2003

[5] H. Zhang, G. Zhang, K-Y. K. Wong, "Camera Calibration with Spheres: Linear Approaches", IEEE International Conference on Image Processing,, pp. 11-14, Sep 2005

[6] X. Ying, H, Zha, "Linear Approaches to Camera Calibration from Sphere Images or Active Intrinsic Calibration using Vanishing Points", IEEE 10th International Conference on Computer Vision, pp. 596-603, Oct 2005

[7] Y. Lu, S. Payandeh, "Dumbbell Calibration for a Multi-camera Tracking System", IEEE Canadian Conference on Electrical and Computer Engineering, 2007, pp. 1472-1475, Apr 2007

[8] L. Wang, F. C. Wu, Z. Y. Hu, "Multi-camera Calibration with One-dimensional Object under General Motions", IEEE 11th International Conference on Computer Vision, pp. 1-7, Oct 2007

[9] R. Guerchouche, F. Coldefy, "Camera Calibration Methods Evaluation Procedure for Images Rectification and 3D Reconstruction", The 16th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision, Feb 2008

[10] X. Feng, M. Cao, H. Wang, M. Collier, "The Comparison of Camera Calibration Methods Based on Structured-Light Measurement", IEEE International Congress on Image and Signal Processing, Vol. 2, pp. 155-160, May 2008

[11] R. Hartley, A. Zisserman, "Multiple View Geometry in Computer Vision", Cambridge University Press, 2000

[12] D. A. Forsyth, J. Ponce, "Computer Vision, A Modern Approach", Prentice Hall, 2002

Xiaochen Dai was born in Shanghai, China. He received the B. Eng. Degree in electrical engineering from Shanghai Jiao Tong University, Shanghai, China, in 2006. After graduation, he worked at ASE (Shanghai) Inc. as a software engineer. Since September 2008, he has been pursuing his M. A. Sc. Degree at the School of Engineering Science, Simon Fraser University, Burnaby, BC, Canada. His research interests include computer vision, camera calibration and visual tracking.

Multi-spots laser triangulation system for endoscopic imaging

Xiaochuan Sun and Shahram Payandeh

Simon Fraser University, Burnaby, BC, Canada

In minimally invasive surgery, lack of depth cues prevents surgeons from acquiring accurate information. As a result, a better visualization tool becomes vital in order to provide a 3D image of the inner cavity so that surgeons can know the exact locations the surgical tools are with respect to the target specimen. This paper proposes a design using triangulation-based laser scanner for depth acquisition. Solutions concerning this approach, such as, optical instrument selection, endoscope calibration and image processing are described in detail. Three increasingly complicated setups are introduced in this paper, evolving toward more realistic and general surgical circumstances, where typical triangulation cannot directly calculate the depth cues. A correlation table is later implemented to correct the errors caused by inconstant baseline distance. Finally, simulation results are presented to show the feasibility of the proposed method.

Key words—**Laser, Triangulation, Endoscopic imaging, 3D image, Depth extraction**

I. INTRODUCTION

MINIMALLY-INVASIVE surgery (MIS) provides a number of benefits to patients, including lower risk of infection and swifter recovery times. Minimal invasion is accomplished primarily through the use of a camera (endoscope) and other surgical instruments inserted into the abdominal cavity through small "keyhole" incisions. The surgeon navigates the process by viewing imagery on an external display. By moving the endoscope and instruments, complex surgical tasks can be efficiently accomplished. However, due to lack of depth cues, surgeons face significant visual and technical obstacles. Therefore, explicit depth information becomes crucial in MIS procedures to reconstruct a 3D image of the inner cavity.

3D surface reconstruction is one of the most challenging tasks in computer vision. This area has developed a variety of approaches to detect localization information.

Generally, the range sensing systems can be divided into contact and non-contact approach. Contact approach is widely used in vehicles and mobile device; while, non-contact sensors are applied in medical robots. In [1], author has presented a brief survey of the technologies currently available for range sensing.

Active triangulation is one of the first range imaging approaches used in robotics. For short range sensing (less than 10m), active triangulation methods were often used [1]. With laser based approaches, the 3D information becomes relative insensitive to background illumination and surface texture. In laparoscopic surgery the scanning range is far less than 10m. Thus, laser scanning based on active triangulation fits for our application in laparoscopic surgery.

Researchers in this field have proposed variety of contributions. [2] present a LARS system equipped with two high-speed galvanometers to steer the laser beam, allowing versatile and robust control. In [3], a portable digital 3D imaging system based on laser triangulation is designed which can be located in remote sites. [4] has built a three-dimensional imaging system based on the projection of structured light and active triangulation.

However, extra instruments will be added, making the whole system bulky and complicated.

This paper contributes a simple triangulation method adopted to obtain depth cues. Triangulation is the process of determining the location of a point by measuring angles to it from given points at either end of a fixed baseline, rather than measuring distances to the point directly.

The configuration consists of two sensors observing the item. In our set-up, one of the sensors is typically a digital camera device, and the other one is a light projector—a He-Ne laser. The projection centers of the sensors and the viewed point on the object's surface define a (spatial) triangle, as shown in Figure 1.

Within this triangle, the distance between the sensors is the base *L* and must be known. By determining the angles between the projection rays of the sensors and the basis, the intersection point, and thus the 3D coordinate, can be calculated from the triangular relations.

The calibration equation is [5]:

$$\text{RC} = \frac{AB \sin\alpha \sin\beta}{\sin(\alpha + \beta)}, \qquad (1)$$

In this formula, *B* is the optical center of camera, *A* is the laser source. In actual cases, however, *A* is not always aligned with *B*, which forms a non-typical triangulation relationship. Consequently, a correlation table will be introduced in section 3 in order to form more satisfactory scanning results.

Given the depth data, 3D reconstruction is operated in Matlab platform used for comparing to targeted surface. Our final result is not only a series of discrete range points, but a simulated mesh model, similar to actual surface. These simulation results will be presented in section 4.
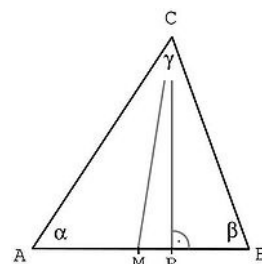
Fig. 1. Basic triangulation structure

II. SYSTEM DESCRIPTION

1) System overview

In this paper, we start with a basic configuration with no mirror; namely, a laser source aligns with the camera origin. A point laser is used in this simplest configuration.

The second configuration involves using one mirror, the center of which is also kept aligned with the optical center. The difference is that horizontal manual scanning is eliminated and replaced by the multi-spots laser system.

For the last and the most complicated set-up, we add another mirror, with the mirror and camera centers on different planes. A correlation table is introduced in this system to address the problem of non-typical triangulation.

A straightforward improvement of laser-spot-triangulation is accomplished by projecting a line or multi-spots instead of one point. In this system, we employed a multi-spot method which allows the acquisition of multiple range data simultaneously from a single video frame.

Figure 2 shows an overview of multi-spots system. A series of cylindrical lenses transform a laser point to a line; afterwards, a 'block paper' (paper with cut apertures) divides this line into multiple spots. In the real surgical operation, we attach the laser at the tip of surgical tools. When multiple spots are projected onto the testing surface, they can be observed using a CCD camera. From the captured images, x, y coordinates can be extracted after camera calibration. Finally, triangulation algorithm is applied to calculate the distance from the camera optical center to the object surface.

Figure 3 shows the typical triangulation scheme. Spot C, laser source, A and camera optical center, B, forms a unique triangle, in which α can be measured and β can be calculated.
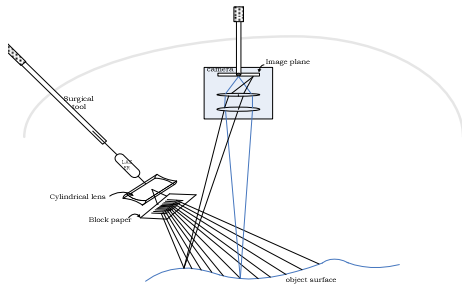
Fig. 2. The system overview
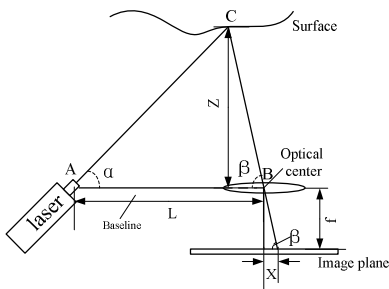
Fig. 3. Typical triangulation

The relationship between β and image location is as follows.

$$\beta = \tan\left(\frac{f}{x/s}\right)^{-1}, \tag{2}$$

X is the distance from targeted spot to image center in pixel numbers and s is the number of pixels in each millimeter. Having known α, β and L, depth z can be derived from the following equation

$$z = \frac{L\sin\alpha\sin\beta}{\sin(\alpha+\beta)}, \tag{3}$$

Finally, x, y, z coordinates can be plotted in 3-dimensional space for further reconstruction on computer.

2) Setup components

The optical system is the most important part for the setup. The basic equipment is as follows:

1. He-Ne laser source.
2. A series of cylindrical lens (diverging).
3. Block paper (9 apertures).
4. A prototype represents object surface.
5. Converging lens.
6. CCD camera

3) Design constraint

According to the camera's product specifications, the maximum horizontal distance of the image plane is 6.4mm,

As figure 4 shows above, the focal length is 35mm and the scanning surface is 15cm-20cm away from the optical center. Thus, the camera only allows 2.7cm-3.6cm horizontal space as targeted surface. However, due to loss of intensity by blocked paper, this space will shrink to approximately 2cm-3cm, as shown in Figure 5. Consequently, the calculated diverging angle is limited within the range of $5.7° \sim 11.4°$. This angle will act as a prerequisite for later lens selection.

The position of the block paper cannot be too close to cylindrical lens; otherwise it fails to form a line. Nor can it be too close to object surface, in which case the paper will block the beam path. Reasonable position would be 10-16cm away from lens. Hence, a block paper is designed within 0.6cm-1.5cm wide and 1.5cm-2.5cm high.
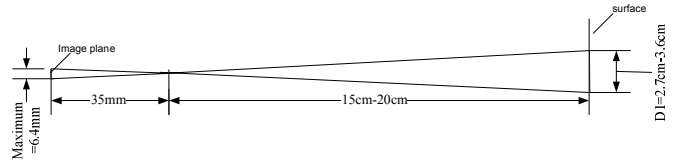

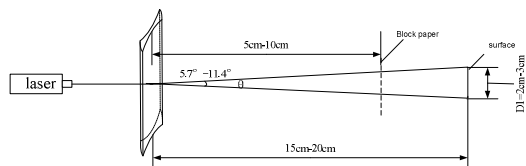Fig. 4. Distance constraint between the surface and image plane

Fig. 5. Distance constraint between lens, block paper and surface

4) Equipment selection

The following discusses the equipment selection according to technical and systematic constrains.

a) He-Ne laser source

He-Ne laser is the most common type of gas laser. It has the advantages in laser beam stability, laser modes, beam diameter, coherence length, output power, and most importantly, the lower price. Thus, we choose He-Ne for testament.

b) Converging lens

In order to force CCD camera focused on the image plane, we choose a converging lens with a focal length of 35mm. To test this lens, a point laser is shined orthogonal to a flat surface. A spot with radius of 0.5mm is captured, small enough to be employed in the following experiment.

c) Cylindrical lens

To spread light into a line, a Plano-convex cylindrical lens with negative focal length is chosen. As stated above in the part of design constraint, calculated diverging angle is $\theta = 5.7° \sim 11.4°$.

The relationship between focal length and converging angle is as follows.

$$f = \frac{\omega_0}{\tan\left(\dfrac{\theta}{2}\right)}, \qquad (4)$$

A typical He-Ne laser waist is 0.4mm, requiring focal length to be $f = -4.0 \sim -8.0mm$, which is far beyond regular lens capability. Hence, A series of lens are placed together to add up the powers (reciprocal of focal length). The relationship of total focal length is

$$\frac{1}{f_1} + \frac{1}{f_2} + \frac{1}{f_3} + \frac{1}{f_4} + \ldots = \frac{1}{f}, \qquad (5)$$

Finally, using two *FL*=50.8mm and three *FL*=25.4mm, we build up a lens system with f =6.35mm and $\theta = 7.2°$.

d) Block paper

According to design constrains, block paper is cut with 1cm*2.5cm size, 9 apertures, as Figure 6 shows.

5) Camera calibration

Range camera calibration is necessarily the first step for range data extraction. Defining the pixel-to-world mapping is known as camera calibration. Once the camera has been calibrated, one can transform the image pixel coordinates to their real-world equivalents. Basically, typical calibration includes calculating the real image center, focal length and scale factors.

In this setup, we approximate that our image center is the center of image plane, thus, center coordinate is (520, 696) in pixel unit. As to the second parameter, it has been discussed in section 3, part 3: design constrains. Finally, scale factor can be calculated according to camera specification, which is 217pixels/mm.

III. THREE TYPES OF SYSTEM SETUPS

In this section, three types of system setups are laid out in a sequence toward the most complicated type. They are separated in terms of laser pattern and triangulation character.
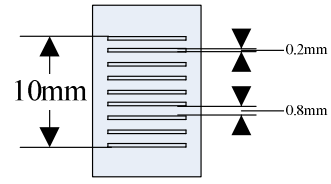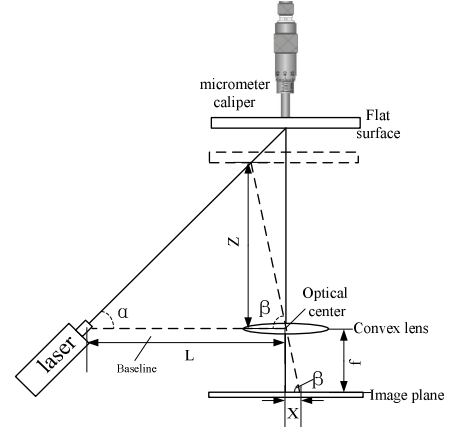

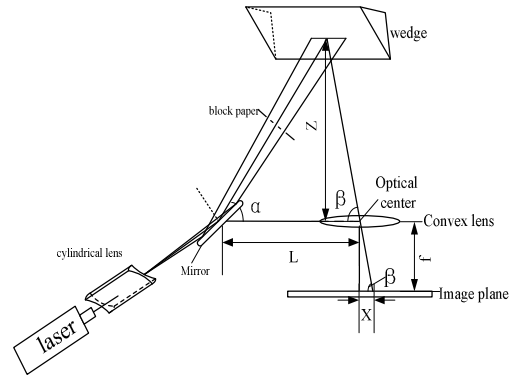Fig. 6. Design of block paper

Fig. 7. Point laser system

Fig. 8. One-mirror, multi-spots system

A. Laser point with typical triangulation

Typical triangulation form a triangle that laser source and optical center are on the same plane, thus, z can be calculated directly using Equation 3.

The basic set-up is shown in Figure 7. A micrometer caliper is used in this setup to move the surface in different distances.

Regarding its simplicity and inductivity, this type is act as a trial configuration to evaluate experiment feasibility.

Testing results are shown and discussed in the next section: experimental results.

B. One-mirror, multi-spots system

In order to scan a surface vertically, a mirror is added in the second configuration, as shown in Figure 8. Testing surface is a wedge with a slope angle of $45°$.
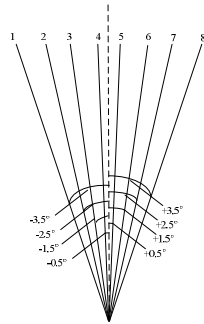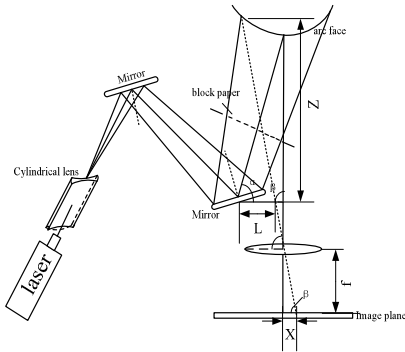
Fig. 9. Angle for each beam

Fig. 10. Two-mirror, multi-spots system

This setup is a typical triangulation pattern with constant baseline L=60mm and central angle $\alpha = 70°$. The total diverging angle is $7.2°$, which is separated to $1.03°$ for each beam. Figure 8 shows the angle relationship between each beam.

C. Two-mirror, multi-spots system

For more complicated situation, when optical center and mirror center are not on the same plane, or their line is not vertical to optical axis, the computing method is called non-typical triangulation.

This case is general in actual surgical operation because the surgical tools can be randomly positioned. In this case, laser spot moves accordingly, forcing inconstant L. To solve this problem, a correlation table is introduced to correct distorted error caused by baseline variance. In order to obtain a larger diverging angle, another mirror is added to this system, setup layout is shown in Fig.10.

In theory, distance z should be the same for every spot in a flat surface. However, actual experiment shows that computing distance changes as baseline varies. To extract the real distance, a simulated flat surface with a uniform depth is forced to correct data. The effect is several computed angle products with different values in different locations. Afterwards, these angles will be used to force necessary displacements for later curve testing surface, where a correlation angle table is produced, as shown in Appendix Table 2. Figure 11 shows the simulation comparison before and after correlation. Simulation results show great correction effect has been achieved after correlation table is applied.
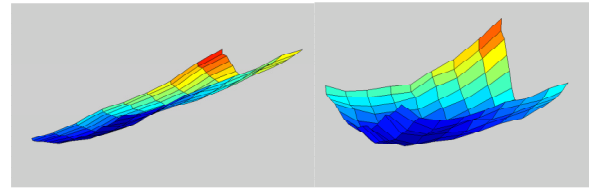

Fig. 11. a. Simulation before correlation
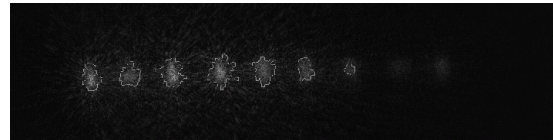b. Simulation after correlation

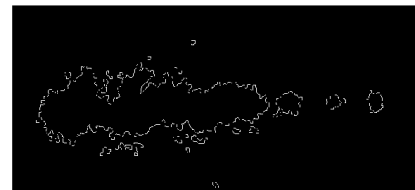Fig. 11. Overlapped image with segmented perimeter

Fig. 12. Segmentation with low threshold

The disadvantage of this method is that it only applies to a specific distance; when distance changes, correlation angle changes. However, only tiny difference ($\pm 5\%$) exists within a small range and the feasibility of this method has been verified in this preliminary design.

D. Image processing

The goal of image processing is to segment out the spots and calculate their center coordinates. The challenges involve big noises that coming from natural light, different surface material and unevenly cut paper, leading to undesired artifacts and randomly shaped spots. Moreover, material color affects imaging quality as well. If the material is black, it will absorb more light than white ones, reducing the necessary contrast and intensity for image processing. As a result, image processing becomes difficult in pursuit of noise reduction and coordinates acquisition.

The following steps are general processes for image preprocessing and spot segmentation using Matlab Imaging toolkit.

1. Read original image
2. Improve contrast
3. Binarize image
4. Fill image regions and holes
5. Morphologically open image (remove small particles)
6. Segment image

The result returns a binary image containing only the perimeter pixels of objects. Figure 11 shows segmented results overlapped with original image.

From the Figure 11 above, we notice that this process cannot segment out spots with lower intensity, such as, spot No.8 and 9. These two lost points can be retrieved back by changing threshold of binarizing steps. As a trade-off, the process will cost previous spots joint together, as Figure 12 shows below.
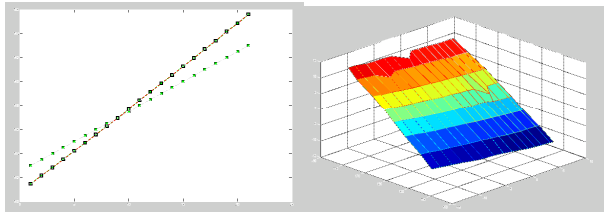
Fig. 13. Comparison simulation of point laser system
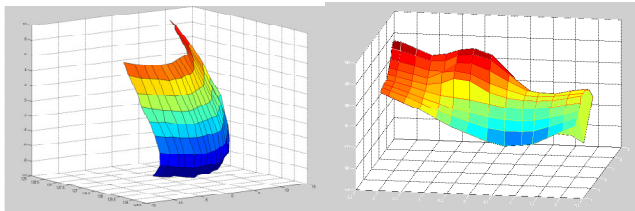Fig. 14. Simulation result of one-mirror, multi-spots system

Figure 15: a. Cup handle simulation
b. Barbie face simulation
some features in the simulation result, from left to the right they are
Barbie's mouth, nose, respectively; the shallow part is their eyes.

Regarding this problem, 5-10 thresholds are being selected, in general, to iterate segmentation process and average effective coordinates. The resulting shows error is within 10 pixels. The errors can be improved by using more advanced algorithm; however, 10 pixels error is small enough for preliminary test.

IV. EXPERIMENT RESULT

In this section, each configuration result is displayed to evaluate the proposed method. They are point laser with typical triangulation system, one-mirror-multi-spot system and two-mirror-multi-spots system, respectively.

A. Point laser with typical triangulation system

The depth information can be obtained by using Equation 3. The calculated system error is within 3mm.

Figure 13 shows the comparison simulated in Matlab, Green line is computing result and red line is measured distance.

The errors come from variety of sources, such as, image processing error, inaccuracy of physically measured distance and angle inaccuracy, etc.

B. One-mirror, multi-spots system

The second configuration uses two-mirror system to scan a wedge. The simulated result is shown below in figure 14.

To evaluate the computing results, we calculate the slope angle of wedge.

The real object has an inclination angle of $45°$; while, the measured angle is $42.3°$, showing acceptable accuracy.

Notice that in Figure 14, some of the points are obviously separated from the plane. This error is caused by low intensity spots and imperfect spot shape. Details will be discussed in part D.

C. Two-mirror, multi-spots system

The following figures are simulation results for testing surface as a cup handle and a Barbie face, respectively.
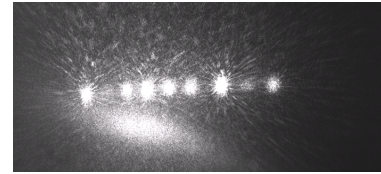

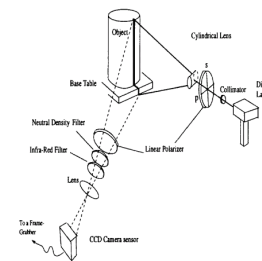Fig. 16. Image on a metal surface

Fig. 17. Possible setup to remove artifacts caused by shiny surface

Since scanning area is only about 5mm*5mm, limited area is detected.

D. Material difference analysis

Different scanning material can produce different images. As to low-noise images, when laser shines on a white paper, for example, segmentation is possible to operate. Otherwise, if the surface is shiny or highly specularly reflective, such as metal, it can result in undesired artifacts.

When these artifacts mix with targeted spots on the image, it is difficult for image processing tool to do analysis. Figure 16 is an example when light shines on a metal cup.

A good way to separate these artifacts and noises from images are using two polarizers, designed by J. Clark et al [6]. The possible setup is shown in Figure 17.

V. CONCLUSIONS

This paper presents a laser-scanning system to extract depth information. The general idea is to obtain 2D image from a camera and combine it with depth cues obtained by using a triangulation method to reconstruct the 3D figure. One of the characteristics in this paper is that an evolved setup scheme is demonstrated. By using a correlation table, we managed to improve data accuracy when we encountered the problem of non-typical triangulation. Another feature is that we divide one point laser into several spots in order to simultaneously extract depth, avoiding horizontal scan.

The simulation result is close to that of the actual surface; however error always exists due to inaccurate physical setup, material difference, image processing, and spot size.

The result of this research can be applied to surgical operations in the future. However, it has some drawbacks before being used in real situations. The first problem is that the image processing work and reconstruction is not a real-time simulation. Future work can focus on using an integrated software system to capture images, process data and simulate an interface at the same time. Another problem is that the third configuration requires a correlation table for each distance, making this method time-consuming and tedious in real surgical operation. It suggests that further investigation of the

geometry of the system is needed. Finally, since this system is not accurate enough, the errors are large, including an unevenly spaced block paper, lens distortion, camera distortion, and spot size. These issues can be addressed with more accurate equipments and improved image processing methods.

APPENDIX

TABLE 1
COMPARISON OF MEASURED AND CALCULATED DISTANCE

MEASURED DISTANCE (mm)	CALCULATED DISTANCE (mm)	CALCULATED DISTANCE (mm)
153.0	151.4	1.6
153.5	152.1	1.4
154.0	152.8	1.2
154.5	153.5	1
155.0	154.2	0.8
155.5	154.8	0.7
156.0	155.5	0.5
156.5	156.3	0.2
157.0	157.0	0
157.5	157.7	-0.2
158.0	158.4	-0.4
158.5	159.1	-0.6
159.0	159.9	-0.9
159.5	160.6	-1.1
160.0	161.3	-1.3
160.5	162.0	-1.5
161.0	162.7	-1.7
161.5	163.4	-1.9
162.0	164.2	-2.2
162.5	164.9	-2.4
163.0	165.7	-2.7

THE CALCULATED SYSTEM ERROR IS WITHIN 3MM

TABLE 2
DATA BEFORE AND AFTER CORRELATION FOR A FLAT SURFACE

THEORET-ICAL ANGLE (degree)	DISTANCE BEFORE CORRELATION (mm)	DISTANCE AFTER CORRELATION (mm)	CALCULATED CORRELATE ANGLE (degree)
-4	124.4419	126.8341	-5.6
-3	125.0529	126.8471	-4.2
-2	125.6609	126.8573	-2.8
-1	126.3681	126.8169	-1.3
0	126.7818	126.7818	0
1	127.1484	126.8498	1.2
2	127.5938	126.8479	2.5
3	128.0838	126.7430	3.9
4	128.4850	126.8461	5.1

REFERENCES

[1] M. Hebert, "Active and Passive Range Sensing for Robotics", Proceedings of the 2000 IEEE International Conference on Robotics and Automation, pp.102 – 110.
[2] D. Chapman, S. E. Hakim, "Short-Baseline Active Triangulation for CAD Reconstruction of Room Sized Industrial Environments", Proceedings of the XIX Isprs Congress, July 2000, pp. 122-129
[3] J.A. Beraldin, F. Blais, L. Cournoyer, M. Rioux, F. Bernire and N. Harrison, "Portable Digital 3-D Imaging System for Remote Sites",
[4] G. Sanson, M. Caroci. S. Lazzari and R. Rodella, "A three-dimensional imaging system for industrial applications with improved flexibility and robustness", J. Opt. A: Pure Appl. Opt. 1(1999) pp. 83-93.
[5] Zhang, Xiaoli. (2001). Application of Image Tracking and 3D Image Reconstruction For Laparoscopic Surgery. Simon Fraser University.
[6] J. Clark, E. Trucco and H-F. Cheung. (1998). Improving Laser Triangulation Sensors Using Polarization Department of Computing and Electrical Engineering, Heriot- Watt University, Edinburgh. 1995.

Xiaochuan Sun received the B.Sc degree in 2008 from Anhui University of Technology. Since the winter of 2009, she has been pursuing her MASc in Simon Fraser University and is currently associated with the Experimental Robotics Laboratory at the School of Engineering Science. Her research interests are medical imaging, computer vision, and human interactive system in minimal invasive surgery.

Shahram Payandeh received the B.Sc. degree in 1981 from the University of Akron, and the MASc and PhD degrees in 1986 and 1990, respectively, from the University of Toronto. He is serving as a professor in the School of Engineering Science at Simon Fraser University. He has several publications to his credit. His research interests are in the general area of mechanics, modeling, and control of robotic (physical/virtual) mechanisms.

Preliminary Analysis of a Legged Robot Designed to Climb Vertical Surfaces

Yasong Li

Abstract—**This paper presents the kinematic analysis of a hexapod climbing robot relying on the use of dry adhesion. Kinematics equations are validated in both multi-body software simulation and robotic platform test. A particular trajectory of the legs, conceived to minimize force required to detach the robotic feet from a vertical wall, is proposed and tested. Further study is performed based on the data acquired by the forces exerted on the tip of each robotic leg during locomotion. Experimental results proved the correctness of kinematic analysis and its potential use for optimizing gait and adhesion features during wall climbing.**

Index Terms— *climbing robot; kinematics; hexapod; peeling; dry adhesion; legged robot.*

I. INTRODUCTION

International interest on climbing mechanisms is growing rapidly within the scientific and industrial communities. The research and development efforts in this field can be categorized in four key areas: (1) servicing (e.g., maintenance of skyscrapers, ships' hulls, nuclear plants, etc.); (2) rescue (e.g., during fire, earthquakes, landslides, etc.); (3) security (e.g., surveillance, inspection, and military operations in buildings, extreme natural environments, etc.); and (4) space (e.g., planetary exploration, Intra-Vehicular Activities, Extra-Vehicular Activities, etc.). Our study is mainly focused on the synthesis of a robust and energy efficient climbing robotic mechanism based on a spider-inspired adhesion system. All components of the mechanism, the mechanical design, sensing, and actuation systems, will ultimately be spider-inspired.

A successful climbing robot should be able to handle tasks which conventional wheeled mobile robots could not achieve. Randall [1] specified several properties a robot should have: (1) highly adaptive to environment, (2) light, (3) reliable and robust, (4) high force to weight ratio, (5) navigate using 3D path planning, (6) possibly 'learn from experience', (7) have reliable and robust control system to drive various actuators,

(8) be able to sense its environment to achieve tasks efficiently, (9) design functions in module in order to 'plug and play' in specific tasks. For a wall-climbing robot, additional requirements are required to achieve the particular climbing feature. The robot should be able to: (A) obtain transition between horizontal and vertical surface in all directions, (B) negotiate obstacles and ledges, (C) prevent machine falling for occasionally adhesion failure, and (D) maximize adhesion efficiency.

Most of the climbing robots that have been developed in recent years [2-6] are implemented with suction or vacuum cups, magnets, or PSAs (Pressure Sensitive Adhesions). Since the confirmation that Van der Waals Forces highly contribute to the impressive adhesion ability of several natural species such as gecko and jumping spiders [7], researchers started to replicate the adhesion mechanisms of animals to develop novel bio-inspired climbing robots [8-15]. Three key criteria are described in [13] for reliable climbing robot using dry adhesion: (1) maximize attachment area, (2) apply a preload between the vehicle and the surface to increase the attachment force, (3) use a peeling mode to detach the dry adhesion. Aimed at those criteria, the design of a climbing robot should maximize the adhesion efficiency by adapting its unique characteristics.

This paper presents the outcomes resulted by a preliminary study aimed at developing a climbing system relying on dry adhesion (namely adhesion based on Van der Waals Forces). The kinematic analysis and the design of the robot are proposed and preliminary results of performed simulations and tests carried out with a robotic platform are discussed. Long-term goal of this research is to develop a bio-inspired and miniaturized climbing system which is capable to climb on any kind of surface and autonomously navigate in unstructured environments.

II. BIO-INSPIRED DESIGN OF THE CLIMBING SYSTEM

We are developing a climbing robotic platform [16], Abigaille-I, inspired by spiders (see Fig. 1). The platform consists of six legs – the use of multiple legs could potentially guarantee adhesion reliability, flexibility to agilely avoid obstacles, and climb on rough terrains. Instead of using eight legs as in spiders, our system has six legs to decrease the mass of the robot. By reducing the number of legs, complexity is also decreased as the number of actuators, sensors, and

electronic components are reduced significantly. Each leg has 6 degrees of freedom (DOF), three passive (wrist) and three active joints. Active joints are controlled by motors directly. Two of the joints, hip and shoulder, which are located at the interface between the robot's body and the leg, have axes perpendicular to each other; an additional joint, acting as elbow joint, has axis parallel to the body plane, and the axis of shoulder joint.
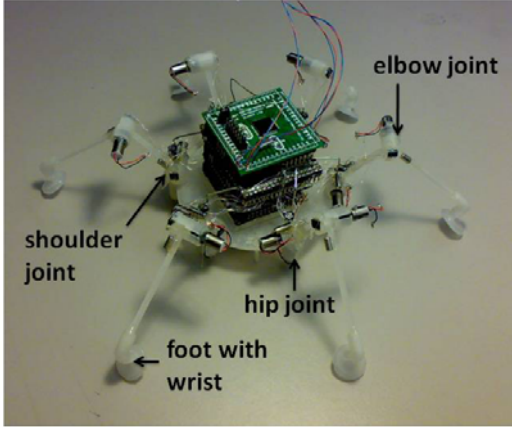

Fig. 1. Abigaille-I, a preliminary prototype of climbing robot [16].

Differently from most of the hexapods with rectangular body, the designed robot consists of a circular body, with six robotic legs symmetrically arranged around. Such bio-inspired configuration allows higher dexterity especially while transferring between surfaces at different angles [17].

III. KINEMATIC ANALYSIS

A. Kinematics of a Leg

The proposed legged robot can be considered as a parallel platform when its legs are in contact to a surface. Each leg can however be considered as a serial manipulator when it is not in contact to the ground. In this section, we consider and discuss this second instance.

Fig. 2 shows a schematic representation of a leg. According to the parameters represented in Fig. 2, the Denavit-Hartenberg transformation matrix in direct kinematics is:

$$T_0^3(\theta_1\theta_2\theta_3) = \begin{bmatrix} c_1c_{23} & -c_1s_{23} & s_1 & c_1(a_2c_2 + a_3c_{23}) \\ s_1c_{23} & -s_1s_{23} & -c_1 & s_1(a_2c_2 + a_3c_{23}) \\ s_{23} & c_{23} & 0 & a_2s_2 + a_3s_{23} \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (1)$$

where c_i and s_i denotes cosine and sine θ_i and s_{ij} denotes sine θ_{i+j}; a_2 and a_3 denote the second and third link length.
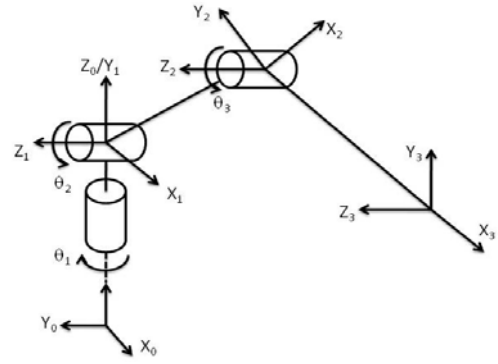

Fig. 2. Frame attachment of one leg.

The last column of the transformation matrix is the position vector of the end-effector, which is the end tip of the leg, responding to a set of joint variables. In inverse kinematics, given the end-effector position (p_x, p_y, p_z), the joint variables can be calculated as

$$\theta_1 = arctan\frac{p_y}{p_x} \quad (2)$$

$$\theta_2 = \pm arccos\frac{p_z^2+p_x^2+a_2^2-a_3^2}{2a_2\sqrt{p_x^2+p_z^2}} + arctan\frac{p_z}{p_x} \quad (3)$$

$$\theta_3 = \mp arccos\frac{p_z^2+p_x^2-a_2^2-a_3^2}{2a_2a_3} \quad (4)$$

B. Kinematics of the Robot Platform

The legged robot could be analyzed as a parallel manipulator when its legs are in contact to the surface. The main body of the robot could be considered the end-effector of the platform in this case. By applying Kutzbach criterion, the DOF of the robot is always six, no matter how many legs in contact to the ground. Fig. 3 shows a schematic representation of the parallel robot. Equations for inverse kinematics are:
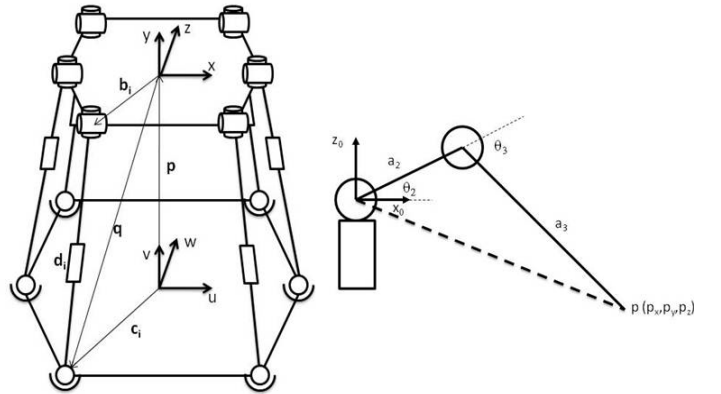

Fig. 3. Frame attachment of simplified parallel platform.

$$R_B^C = \begin{bmatrix} u_x & v_x & w_x \\ u_y & v_y & w_y \\ u_z & v_z & w_z \end{bmatrix}$$

$$= \begin{bmatrix} c\alpha c\beta c\gamma - s\alpha s\gamma & -c\alpha c\beta s\gamma - s\alpha c\gamma & c\alpha s\beta \\ s\alpha c\beta c\gamma + c\alpha s\gamma & -s\alpha c\beta s\gamma + c\alpha c\gamma & s\alpha s\beta \\ -s\beta c\gamma & s\beta s\gamma & c\beta \end{bmatrix} \quad (5)$$

$$q_i = R_C^B (c_i - p) = R_B^{C^T}(c_i - p) \quad (6)$$

$$d_i = q_i - b_i \quad (7)$$

$$\theta_{1i} = \arctan\frac{d_{iy}}{d_{ix}} - \arctan\frac{b_{iy}}{b_{ix}} \quad (8)$$

$$\theta_{3i} = \mp arccos\frac{|d_i|^2 - a_2^2 - a_3^2}{2a_2 a_3} \quad (9)$$

$$\theta_{2i} = \pm arccos\frac{|d_i|^2 + a_2^2 - a_3^2}{2a_2|d_i|} + arctan\frac{d_{iz}}{d_{ix}} \quad (10)$$

where α, β and γ denote the Euler angles with respect to the ground coordinate; frame B and C locate at the robot body plane and ground plane.

IV. KINEMATIC VALIDATION

A commercial platform, Lynxmotion AH3-R (see Fig. 4 (a)) is used to validate the kinematic equations. This platform has a similar mechanical design of the prototype shown in Fig. 1.

Pentapedal gait, which is a gait that moves one leg forward at a time before finally moving the body forward, is considered to move the robot; the robot body is kept parallel to the ground. The pentapedal gait is one of the desirable gaits for a climbing robot, as it keeps as many legs as possible in contact to the surface during the step cycle. By utilizing kinematic equations in the pentapedal gait, the trajectories of each joint within a step cycle were generated and stored in a data table.

A robot with the same shape and weight of the AH3-R was built in a multi-body simulator based on Open Dynamic Engine – the simulated robot (see Fig. 4 (b)) was able to move its 18 actuators according to the computed trajectory. Simulated results were compared to experiments performed by using the AH3-R. Fig. 5 shows both the simulated and tested displacement of the body of the robot. The gait was generated to move the robot 1 cm per step cycle. After 39 step cycles, the robot was supposed to reach 39 cm but, due to the accumulated error, the displacement in the test was 36.5 cm. There was about 1mm of error in each step (see Fig. 5). The error was due mainly to the mechanical clearances of the 18 revolute joints of the robot. Test results therefore validated both the kinematic equations and multi-body simulator.
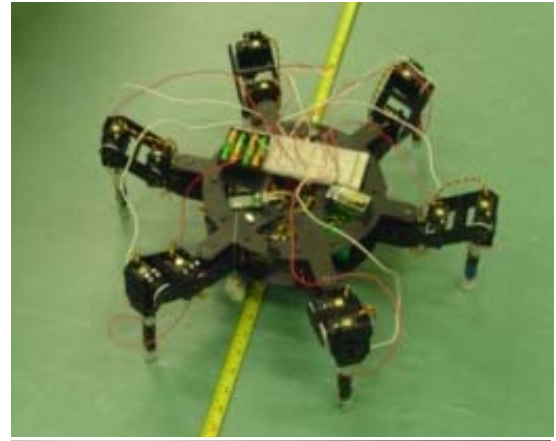
(a)
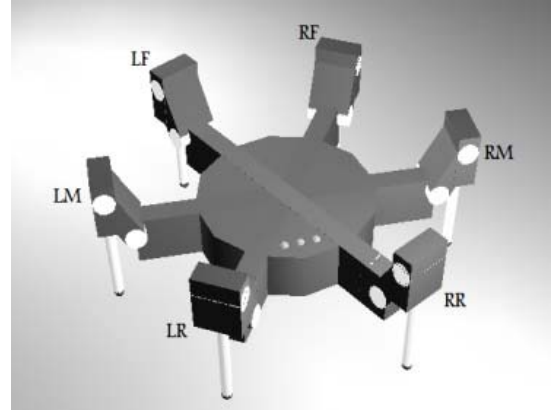
(b)

Fig. 4. (a) Test platform Lynxmotion AH3-R. (b) Robotic platform in the multi-body simulator.
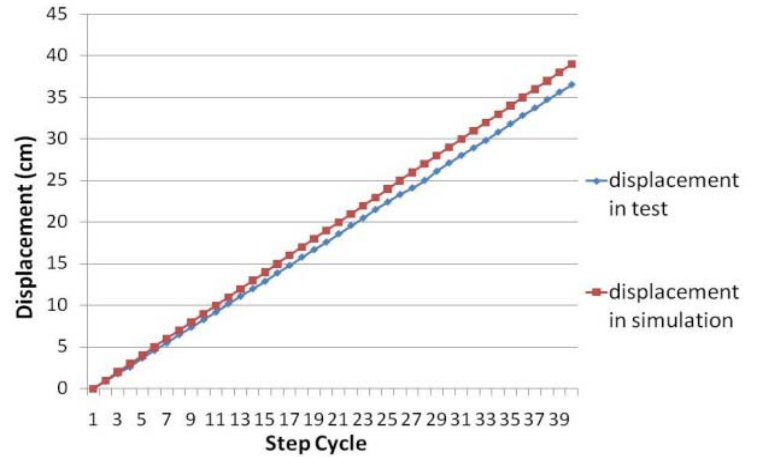

Fig. 5. Displacement test of robot moving forward with pentapedal gait.

V. OPTIMAL MOTION OF RELEASING ADHESION

A. Theory of Peeling Elastic Thin Film

The designed robotic foot shown in Fig. 1 is composed of a thin layer of Polydimethylsiloxane (PDMS). Although dry adhesion is achieved by combining compliant structures both in macro and micro scale, it worth validating the theory of peeling thin films first in macro scale before implementing micro or nano structures to the adhesive.

It is easier to release a dry adhesive when peeled it off. Several studies discussed about efficiently peeling an elastic

thin film and micro-fibrilla, such as the gecko seta, in different directions [18-23]. Kendall [24] proposed a simplified theory of peeling an elastic thin film from a rigid substrate and identified the relationship between peeling force and peeling angle, which can be summarized by the following equation:

$$\left(\frac{F}{b}\right)^2 \frac{1}{2dE} + \left(\frac{F}{b}\right)(1 - cos\theta) - R = 0 \qquad (11)$$

where b, d and E denote the width, thickness and Young's modulus of the elastic thin film. θ denotes the peeling angle between thin film and substrate. F represents the peeling force in the same angle θ with respect to the substrate. R denotes the adhesive energy, which is the experimental energy required to fracture unit area of the interface. Equation (11) could be simplified as follow:

$$\frac{F}{b}(1 - cos\theta) = R_c \qquad (12)$$

where R_c denotes the adhesive energy when the crack speed is constant. Equation (12) yields that the minimum force required to peel an adhesive is obtained by pulling its edge at 90 degree respect to the substrate.

B. Experimental Setup

In order to experimentally verify the simplified theory of peeling an adhesive and assess its suitability for our application, a simple experimental setup was prepared as shown in Fig. 6. Same Hall Effect rotary sensors (HMC1512) and revolute motors (GM15) used in [16] were employed in this setup to produce data relevant to improve the design of Abigaille-I.

Fig. 7 shows the desired (blue line) and the experimental (red line) trajectories of the shoulder and elbow joints of the leg prototype. A PID controller was implemented in LabVIEW to control the leg movement. Fig. 7 shows a small jittering during its motion – this is due to the simultaneous movement of the elbow. Trajectories represented in Fig. 7 are samples of the trajectories used during peeling tests.
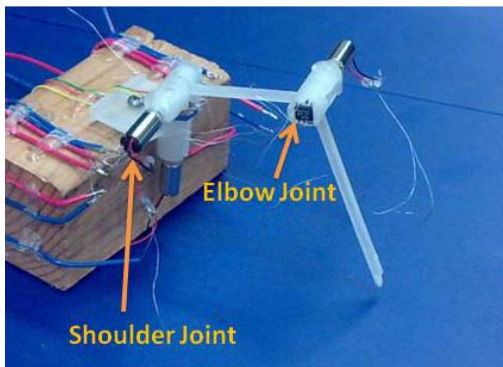

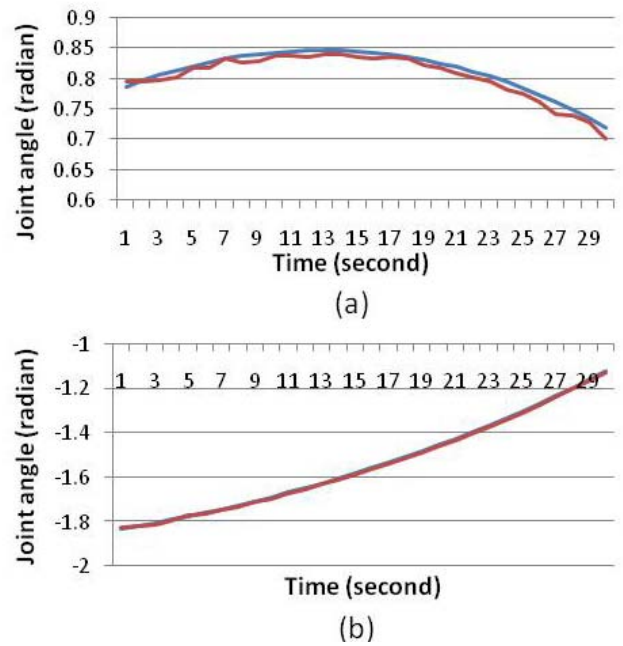Fig. 6. Leg prototype used in releasing adhesion test.

Fig. 7. (a) Position control of the elbow joint in one robotic leg. (b) Position control of the elbow joint in one robotic leg. In both graphs, red line: measured data; blue line: desired trajectory.

C. Power Consumption Test with Various Peeling Angle

According to the results obtained in section A, peeling off an elastic thin film at a constant angle of 90 degrees should provide power saving. Tests using the setup shown in Fig. 6 were performed for peeling an elastic thin layer at different angles.

The trajectory which guarantees a fixed peeling angle, for instance 90 degrees, is shown in Fig. 8; the dashed lines denote the unchanged length of the already released part of the adhesive. In Fig. 8 the peeling force is normal to the substrate and the end-effector of the robotic leg is attached to the left end of the adhesive. Geometrical observations based on Fig. 8 yield that the tip of the leg should follow a trajectory at 45 degrees with respect to the substrate in order to detach the adhesive at a constant angle of 90 degrees. Trajectories of the end-effector would always follow straight lines to peel the adhesive at different constant angles; for example, peeling angle of 30 and 60 degrees are respectively accomplished by end-effector straight trajectories at 75 and 60 degrees with respect to the substrate.
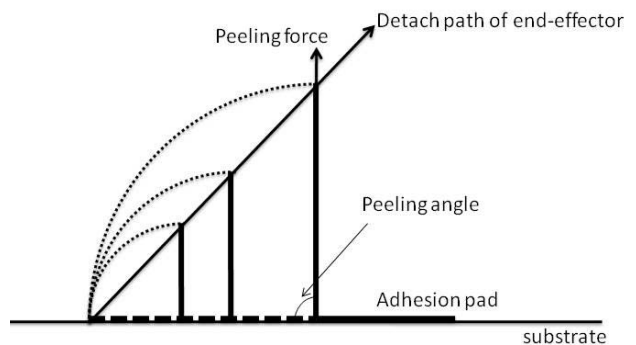

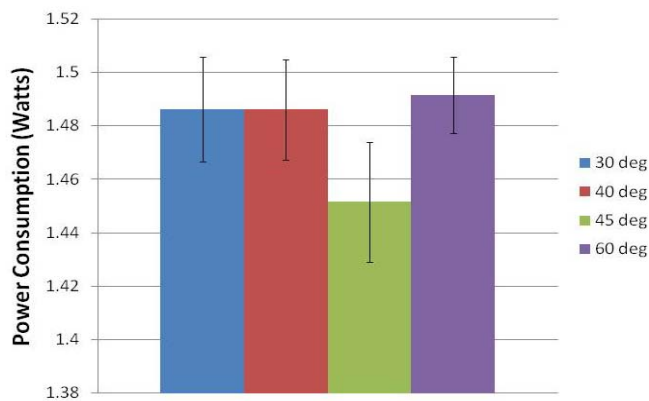Fig. 8. Trajectory with Constant Peeling Angle

Fig. 9. Trajectory with Constant Peeling Angle.

Power consumption measurements were performed during peeling at different angles. A thin layer of silicone strip was attached to the tip of the leg shown in XFig. 6. Kinematic equations derived in previous sections were used to control the joints of the leg prototype. A digital multimeter was connected between the power supply and motor power pin of H-bridge chip to measure and record the current provided to the actuation system. As the voltage supplied to the motors was constantly 3V, the peeling force was proportional to the current flowing to the motors. Four sets of straight trajectories at 30, 40, 45 and 60 degrees to the substrate were tested. As shown in Fig. 9, trajectories at 45 degrees, which are associated to 90 degree peeling angle, consumed the least power thus confirming both theoretical predictions and applicability to our robotic climbing platform. Notice that results obtained for trajectories at 30 and 60 degree produced similar power consumption and that there is a dramatic variation between the ideal 45 degrees trajectory and the 40 degrees trajectory – good position control should therefore be used to guarantee the lowest power-consumption trajectory.

VI. FORCE DISTRIBUTION OF LEGS DURING ONE STEP CYCLE

We investigated the force distribution in each leg during a designed gait, in order to predict and further optimize the adhesion performance. First, we simulated a pentapedal gait and estimated reaction forces at the tip of each leg by using our multi-body simulator. Afterwards, force distribution measurements were obtained by using the AH3-R, which we modified by integrating Force Sensing Resistors (FSRs) at the tip of each leg. Forces were recorded while the robot moved using a pentapedal gait. Fig. 10 (a) and (b) respectively show the simulated and measured forces of the Left Front Leg (LFL) during a full walking cycle.

In our designed gait, the LFL is the first leg taking action in the sequence (see Fig. 10). As seen in Fig. 10, during Phase A no force is applied to the end tip of the LFL as it is in swing phase. Phase B starts when the LFL contacts the ground and the Right Front Leg (RFL) is lifted. Since the RFL is not supporting any body weight, the LFL and Right Middle Leg

(RML) should undergo the highest force as they share the load of the moving leg. Phase C starts with the RML moving up. The LFL is not adjacent or opposite to the moving leg (RML) and has the same load as all legs are in contact to the ground. Data in Phase D shows that the force decreases when opposite leg is in swing phase. Phase E and F have the similar effect of Phase C and B. Phase G represents the robot body moving forward with all legs attached to the ground. Since the FSRs are not suitable to show the absolute value of the force, both the simulated and experimental data are normalized to better compare the trends of interested forces. We can observe that (a) the moving leg bears no force and its opposite leg bears minor force, (b) two adjacent legs undergo the highest load during the movement, (c) forces at the non-adjacent and non-opposite legs are identical to the stance phase. Test results showed that the simulation is consistent to the realistic force distribution of the robot; therefore future analyses predicting loads during vertical climbing or upside-down locomotion could accurately be performed.
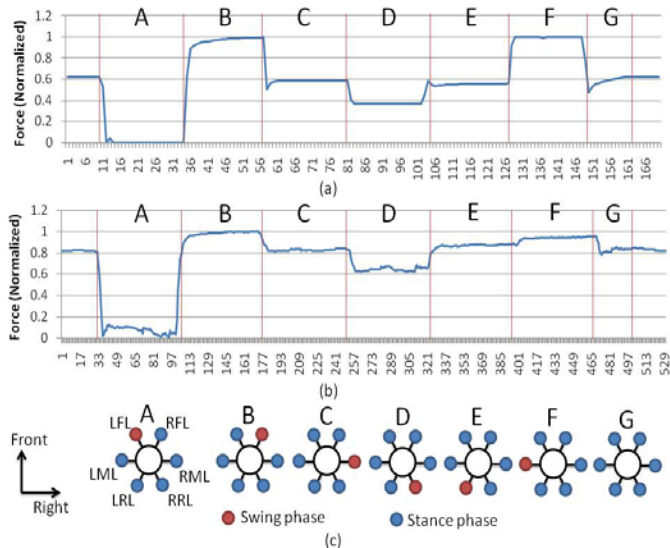

Fig. 10. Force measurement targeted at Left Front Leg (LFL) within one step cycle. (a) simulation data; (b) measurement from real robot; (c) devided phases during the robot movement. Looking from top, RFL stands for Right Front Leg, RML stands for Right Middle Leg, RRL stands for Right Rear Leg respectively.

VII. CONCLUSION

This paper presented a kinematic analysis of a hexapod robot designed to climb vertical walls. Multi-body software was used to simulate the robot's locomotion. Both kinematic equations and software were tested by comparing displacement and force predictions to experimental measurements. In addition, it was proven that peeling motion of the tip of the robot's leg, which follows a straight line at 45 degrees with respect to a flat surface, is the most efficient trajectory to detach a thin adhesive foot.

VIII. ACKNOWLEDGEMENT

simulation result in multi-body simulator; Claire Wu in recording measurement values on power consumption test; Dr. Carlo Menon in providing guidance on the work.

REFERENCES

[1] M. J. Randall, "Adaptive neural control of walking robots," London: Professional Engineering, 2001, pp.2.

[2] B. L. Luk, A. A. Collie, and J. Billingsley, "Robug II: An intelligent wall climbing robot," Proceedings of the 1991 IEEE International Conference on Robotics and Automation, pp. 2342-2347, 1991.

[3] L. Briones, P. Bustamante, and M. A. Serna, "Wall-climbing robot for inspection in nuclear power plants," Proceedings of the 1994 IEEE International Conference on Robotics and Automation, pp. 1409-1414, 1994.

[4] R. T. Pack, J. L. Christopher, and K. Kawamura, "A rubbertuator-based structure-climbing inspection robot," Proceedings of the 1997 IEEE International Conference on Robotics and Automation, pp. 1869-1874, 1997.

[5] K. A. Daltorio, A. D. Horchler, S. Gorb, R. e. Ritzmann, and R. D. Quinn, "A small wall-walking robot with compliant, adhesive feet," IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 3648-3653, 2005.

[6] M. Murphy, W. Tso, M. Tanzini, and M. Sitti, "Waalbot: an agile small-scale wall climbing robot utilizing pressure sensitive adhesives," IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 3411-3416, 2006.

[7] K. Autumn, M. Sitti, Y. A. Liang, A. M. Peattle, W. R. Hansen, S. Sponberg, T. W. Kenny, R. Fearing, J. N. Israelachvili, and R. J. Full, "Evidence for van der waals adhesion in gecko setae," PNAS, vol. 99, pp. 12252-12256, 2002.

[8] M. Greuter, G. Shah, G. Caprari, F. Tache, R. Siegwart, and M. Sitti, "Toward micro wall-climbing robots using biomimetic fibrillar adhesives," presented at Proceedings of the 3rd International Symposium on Autonomous Mini Robots for Research and Edutainment, Awara-Spa, Fukui, 2005.

[9] K. A. Daltorio, S. Gorb, A. Peressadko, A. D. Horchler, R. E. Ritzmann, and R. D. Quinn, "A robot that climbs walls using microstructured polymer feet," Proceedings of the Eighth International Conference on Climbing and Walking Robots and the Support Technologies for Mobile Machines, pp. 131-138, 2005.

[10] O. Unver, A. Uneri, A. Aydemir, and M. Sitti, "Geckobot: a gecko inspired climbing robot using elastomer adhesives," Proceedings of the International Conference on Robotics and Automation, pp. 2329-2335, 2006.

[11] S. Kim, M. Spenko, S. Trujillo, B. Heyneman, M. V, and M. Cutkosky, "Whole body adhesion: heirarchical, directional and distributed control of adhesion forces for a climbing robot," IEEE Conference on Robotics and Automation, pp. 1268-1273, 2007.

[12] M. Murphy, M. Sitti, "Waalbot: an agile small-scale wall climbing robot ustilizing dry elastomer adhesives", IEEE/ASME transactions on Mechatronics, vol. 12 (3), pp. 330-338, 2007.

[13] C. Menon, M. Murphy, and M. Sitti, "Gecko inspired surface climbing robots," presented at IEEE International Conference on Robotics and Biomimetics, Shenyang, China, 2004.

[14] Menon, C., Murphy, M., Sitti, M., Lan, N., "Space exploration - towards new advanced bioinspired climbing robots, in (M.K. Habib Ed.) Bioinspiration and Robotics - Walking and Climbing Robots", Advanced Robotic Systems International and I-Tech, pp544, 2007.

[15] Menon, C., Sitti, M., "A Biomimetic Climbing Robot Based on the Gecko", Journal of Bionic Engineering, Elsevier, Vol. 3, No. 03, pp. 115-125. 2006.

[16] C. Menon, Y. LI., D. Sameoto, and C. Martens, "Abigaille-I: towards the development of a spider-inspired climbing robot for space use," IEEE RAS / EMBS International Conference on Biomedical Robotics and Biomechatronics, Scottsdale, Arizona, USA, 2008.

[17] Z. Wang, X. Ding, A. Rovetta, "Structure Design and Locomotion Analysis of a Novel Robot for Lunar Exploration," 12th IFToMM World Congress, Besancon, Fance, June18-21, 2007.

[18] K. Autumn, Y. A. Liang, S. T. Hsieh, W. Zesch, W. P. Chan, T. W. Kenny, R. Fearing, and R. J. Full, "Adhesive force of a single gecko hair," Nature, vol. 405, pp. 681-685, 2000.

[19] S. Kim and M. Sitti, "Biologically inspired polymer microfibers with spatulate tips as repeatable fibrillar adhesives," Applied Physics Letters, vol. 89, pp. 261911 1-3, 2006.

[20] H. Yao, H, Gao, "Mechanics of robust and releasable adhesion in biology: Bottom-up designed hierarchical structures of gecko," Journal of the Mechanics and Physics of Solids, vol 54 (6), pp. 1120-1146, 2006.

[21] Menon, C., Lan, N., Sameoto, D., "Towards a methodical approach to implement biomimetic paradigms in the design of robotic systems for space applications", Applied Bionics and Biomechanics, pp.1–13, 2009

[22] Sameoto, D., Li, Y., and Menon, C., "Multi-scale compliant foot designs and fabrication for use with a spider-inspired climbing robot", Journal of Bionic Engineering, Vol. 05, No. 03, pp. 189-19, 2008.

[23] Sameoto, D., Li, Y., Menon, C., "Micromask Generation for Polymer Morphology Control: Nanohair Fabrication for Synthetic Dry Adhesives", Advances in Science and Technology, Trans Tech Publications, Vol. 54 (2008) pp 439-444.

[24] K. Kendall, "Thin-film peeling – the elastic term," Journal of Physics D (Applied Physics), pp. 1449-1452, 1975.

Yasong Li received her Bachelor of Engineering in South China Agricultural University (Guangzhou, Guangdong, China) in 2006 and Master of Applied Science in Simon Fraser University (Burnaby, BC, Canada) in 2009. She is currently a Phd candidate in MENRVA Group at School of Engineering Science, Simon Fraser University. Her research interests are focus on intelligent system, biomimetics and robotics.

TOWARD UNIFICATION OF CONSTRAINED MECHANICS AND VIRTUAL FIXTURES IN HAPTIC RENDERING

Yifan Yang

Department of Engineering Science, Simon Fraser University, Canada

Abstract: **Haptic virtual fixtures are software-generated force and position signals applied to human operators in order to improve the safety, accuracy, and speed of robot-assisted manipulation tasks. Virtual fixtures are effective and intuitive because they capitalize on both the accuracy of robotic systems and the intelligence of human operators. In the context of human-machine collaborative system, the term virtual fixtures is most often used to refer to a task dependent aid that guides the user's motion along desired directions while preventing motion in undesired directions or regions of the workspace. In this paper, we will show how to apply the computational mechanics for solving constrained system to haptic rendering of stiff-constrained environment. We intend to present the modeling and simulation of constrained multi-body systems based on the Phantom Omni haptic device and Windows platform with advanced graphics card.**

Keywords: constrained mechanics, virtual fixtures (VF), computational mechanics, haptic rendering

1. Introduction

The concept of virtual fixtures was firstly introduced in (Rosenberg, 1993) as an overlay of abstract sensory information on a workspace in order to improve the tele-presence in a tele-manipulation task. This kind of technology is widely used in the remote control systems where the tip of the manipulator needs very complex constraint guidance forces which are not accessible from the physical world. For example, in the area of tele-operation, we need either the guiding virtual fixtures or forbidden regions virtual fixtures in order to enhance the precision of the operation.

Some challenges must be considered when implementing VF. These can be: a)design of VF as a assembly of basic elements (Kuang et al., 2004); b) definition of force fields associated with the VF; c) stable computational models of constraining and guiding forces; and d) development of stable hybrid control laws where the user can move between interconnected VFs. This paper is based on the modeling of constrained multi-body systems. We present the simulation and the interaction with a class of VF in a haptic rendering environment.

This paper is organized as follows: section 2 presents some details of main results; section 3 presents the stable formulation for constrained computational mechanics and section 4 and section 5 present discussions and some future work.

2. The modeling and the simulation of constrained mechanics

1) The model of the constrained surfaces

Let us consider a multi-body system characterized by n-dimensional generalized coordinate vector q, and n × n-dimension square symmetric mass matrix M(q) and a n-generalized internal force vector(Coriolis, centrifugal or elastic forces) $f_{int}(q,\dot{q})$. If this system is subjected to the m holonomic constraints defined as:

$$\Phi(\mathbf{q}) = 0, \quad (1)$$

Then the governing equations given by application of fundamental principle of mechanics (Newton-Euler, Lagrange equations, virtual power, ...) are:

$$\mathrm{M}\ddot{q} = f_q + f_{co}, \quad (2)$$

if all the holonomic constraints are independent, the dimension of H is n − m.

   Instead of thinking about constraints in terms of hard surfaces, rigid links, slip-less rolling contacts, etc., a legitimate and old treatment, motivated then by Arnold [10], is to replace the 'stiff' constraints by a strong force field in a neighborhood of the manifold $\phi(q) = 0$, directed towards the manifold. The effect can be interpreted as the action of an elastic forces pointed in the constrained directions towards the respective constraints with stiffness k and proportional to the constraint violations (Figure 1).
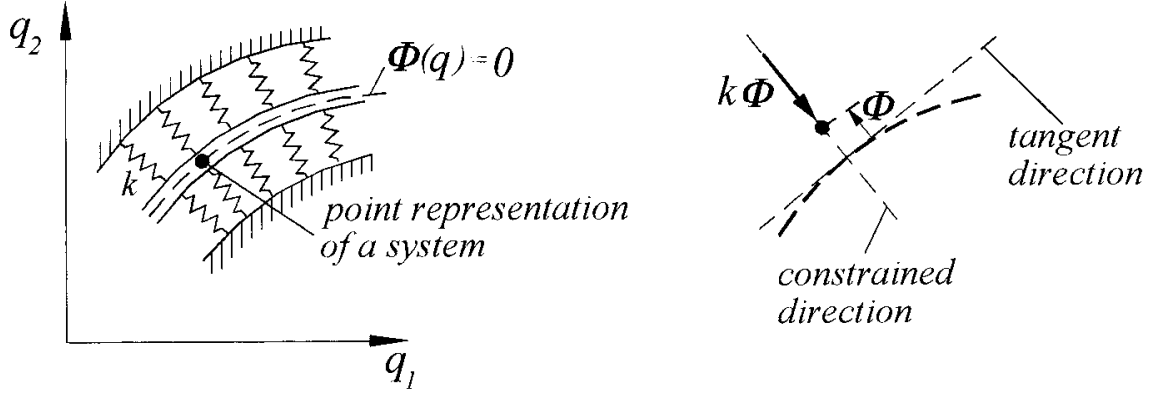

Figure 1.The physical interpretation of Arnold's/penalty formulation.

   In order to realize a stable modeling of the constrained curves, we need to add the damping element $c\,\dot{\Phi}(q)$ and an inertial force $\mu\ddot{\Phi}(q)$. Thus, the effect of the constrained forces can be rewritten in the familiar form $\mu(\ddot{\Phi} + 2\eta\,\dot{\Phi} + \omega^2\Phi)$ where $c = 2\eta/\mu$ and k=$\omega^2 * \mu$.

   On the other hand, we can easily derive the following equations:
$$\Phi(q) = 0$$
$$\dot{\Phi}(q) = C_q\,\dot{q} = 0 \qquad\qquad (3)$$
$$\ddot{\Phi}(q) = \dot{C}_q \cdot \dot{q} + C_q \cdot \ddot{q}.$$
Where $C_q = \partial\Phi/\partial q$

Considering the additive mass of the virtual fixture, we can get the equation of constraints in the form:
$$(1 + M_\phi/\mu)\ddot{\Phi} + 2^\eta\,\dot{\Phi} + \omega^2\Phi = M_\phi\ddot{\Phi}/\mu, \text{ with } M_\phi = (C_q M^{-1} C_q^T)^{-1}. \qquad (4)$$

   The parameter M_ϕ here can be explained as the dynamical component of constrains. If the parameter $\mu \to \infty$, the nonlinear feature of the system decreases. This formulation is well known as a penalty formulation and classic value of μ found in the literature are between 1E5 and 1E7.

   So the work is emphasized on finding the incremental motion Δy perpendicular to the constrained surfaces. With the position displacement of Δy in each servo loop, we can get the velocity and the acceleration of the tip. The values of μ, η, ω are to be tuned in the experiment.

2) The displacement

    This part of the work is critical and usually we can use the geometry way to solve the problem. Here we will discuss a simple condition of geometric constraints: a constraining line in 3-D space.

    Assume there is a point P(x, y, z) to a line with a unit vector n. The displacement of D can be calculated in following steps:

    1) Find a random point on the line, we mark it Q(x0, y0).

    2) Find the projection of PQ to the line, that is, PQ $\bullet$ $\vec{n}$.

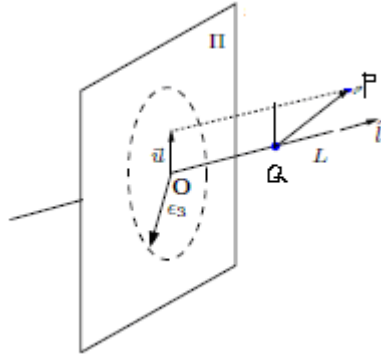   3) The vector of displacement $\vec{L}$ can be easily found by PQ- PQ $\bullet$ $\vec{n}$.

Figure 2.Moving along a line

3. Main results of the simulation

    In our experiment, we use the Phantom Omni device and P4 2GHz PC platform with advanced graphics card to do the simulation work. I build a model of 3-DOF robot arm in the graphical environment (see Figure 3). The small sphere in the figure represents the tip of Phantom device, and the green line represents the constrain line in 3D space. With the GLUI interface, users can change the parameter of the constraining line to get a different guiding force.

    Figure 4 shows another visualization example of constraining surface in a form of parabolic curve. Figure 5 is the error along the direction perpendicular to the linear constraining path. Forces of constrain are generated to guide the user to maintain close proximately to the VF.
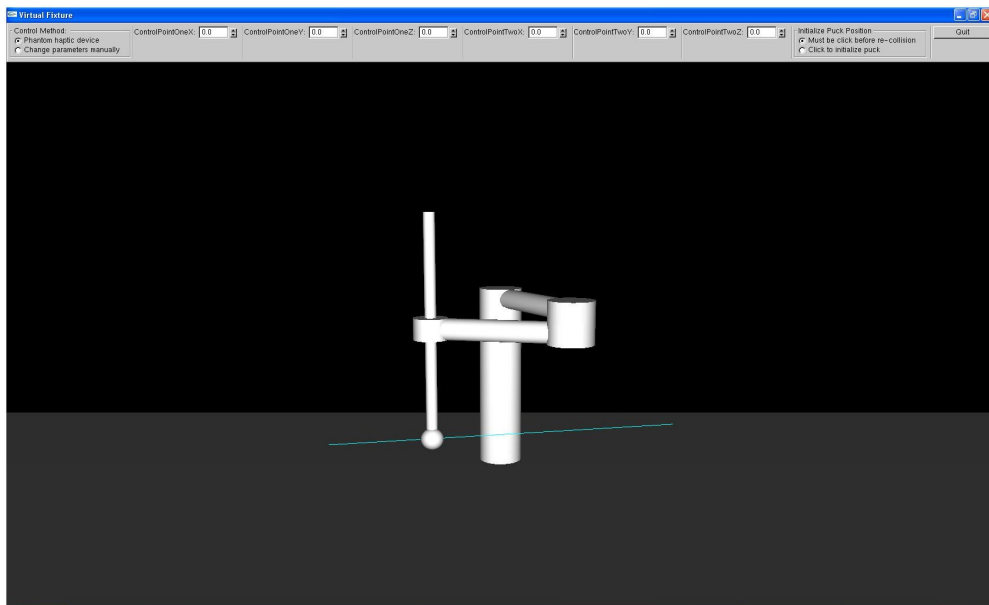
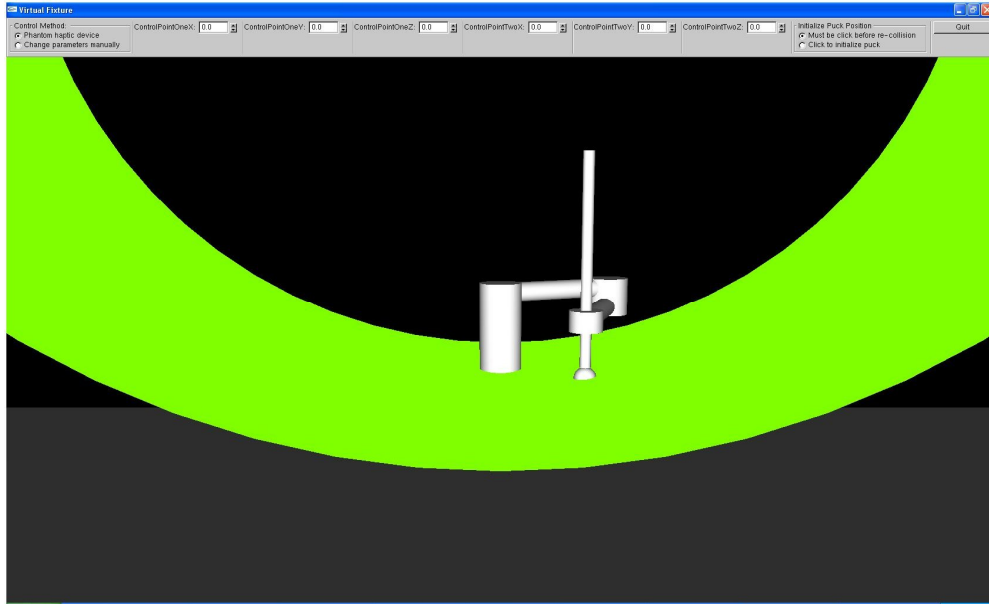Figure 3　Visualization of a 3-DOF robot arm with a constraining line

Figure 4   Visualization of a 3-DOF robot arm with parabolic constraining surface
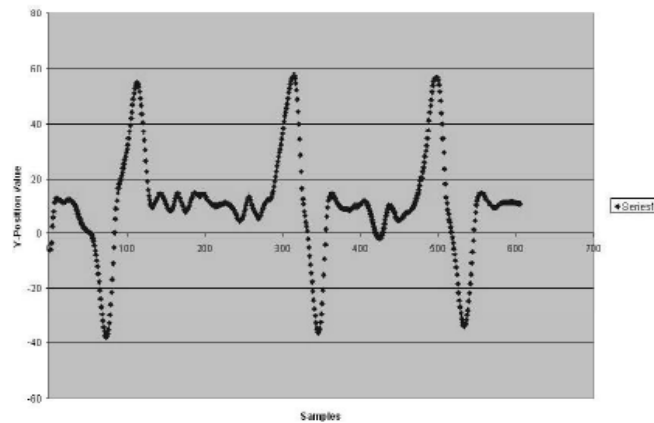

Figure 5 A y-direction position error

4. Discussions and the future work

   In this paper, we integrate the Lagrange Mechanics to the virtual environment in order to get a stable, robust constraining force. For this initial studies, we have not add the addictive mass in the simulation and the constraining force here is simply a spring with damping force and a static inertial force. The effect of the constraining condition is only discussed theoretically. Meanwhile, the simulation work in this paper is limited on the one degree freedom sample of virtual fixture. In our future work, we will intend to build a model of multiple constraining surfaces which need more complex research on their integration.

5. References

[1]. Seguy, N., P. Joli, Z. Q. Feng and M. Pascal(2003). A modular dynamic model for multi-body system adapted to interactive simulation. ASME Design Engineering Technical Conference.

[2]. Li, M., A. Kapoor and R. H. Taylor (2005). A constrained optimization approach to virtual fixtures. Proceedings of IEEE/RSJ   International Conference on Intelligent Robots and Systems pp. 2914–2929.

[3]. Aghili, F. and J. Piedeceuf (2003). Simulation of motion of constrained multi-body systems based on projection operator. Multi-body Systems Dynamics 10, 3–16.

[4]. Blajer, W. (2002). Augmented lagrangian formulation: Geometrical initerpretation and application to system with singularities and redundancy. Multi-body Systems Dynamics 8, 141–159.

[5]. Payandeh, S. Joli, P. Toward unifications of constrained mechanics and virtual fixtures in haptic rendring. 2004 research project activity report on Image Synthesis, Animation, Modeling and Simulation.

[6]. Oliver D. Johns, (2005) Analytical Mechanics for Relativity and Quantum Mechanics, Great Britain: Oxford university press.