Body and Gesture Tracking, Representation and Analysis: Implications for HCI, Human Motor Function & Virtual Human Action.

1. Research Team

Project Leader:	Prof. Albert Rizzo, IMSC and Gerontology
Other Faculty:	Prof. Isaac Cohen, Computer Science
	Prof. Gerry Kim, IMSC
	Prof. Chris Kyriakakis, Electrical Engineering
	Prof. Gloria Mark, UC-Irvine
	Prof. Shrikanth Narayanan, Electrical Engineering
	Prof. Cyrus Shahabi, Computer Science
	Sharon Carnicke, USC School of Theatre
	Stacey Marsalas, USC/ICT
	Jon Gratch, USC/ICT
Post Doc(s):	Tim Marsh
Graduate Students:	Yeh Shih-Ching, Jane Hwang, Wei Peng, Kiyoung Yang, Baharak Zali
Undergraduate Students:	Alnur Ali

2. Statement of Project Goals

This User Centered Sciences (UCS) area has multiple projects with unique goals. These involve:

Gestural HCI – This project aims to design, develop and evaluate a hand gesture based methods for enhancing human computer interaction in collaboration with the IMSC 2020 Classroom Research Vision. The research in this area will also have impact on user interface components across a wide spectrum of IMSC research areas. Comparison of hand gestures for object selection and manipulation with mouse usage is being focused on along with the development of a 3D User Interface benchmarking scenario.

Vision-Based Tracking – This project aims to create a vision-based tracking and sensing system in collaboration with Isaac Cohen and Shrikanth Narayanan to be used for tracking, representing and quantifying human motor performance within the IMSC Communication Vision. Advances in vision-based tracking technology could form the basis for the capture and recognition of human action required for the creation of multimodal HCI options that utilize gestural behavior. This technology could also have significant impact on applications that target gait analysis and motor rehabilitation and we have developed three VR scenarios for this purpose.

Body Gesture and the Expression of Emotional State in Virtual Human Representations This project aims to investigate human capacity to decode the communication delivered via body action in virtual human representations (avatars/agents). Previously, IMSC UCS research has examined the factors that contribute to successful detection, by human users, of emotional facial expressions conveyed by avatars actuated using Performance Driven Facial Animation [1]. The current work has expanded to investigate similar issues for full body non-verbal expression of emotion.

The knowledge gained from these projects will drive the integration of UCS methodologies in the design, development and evaluation of IMS that incorporates gestural interaction with both standard magnetic tracking and vision-based systems. This work could also have wider generalizable value for creating the enabling technology required for development of multimodal and perceptual user interfaces (PUIs). The goal here is to supplement current human-computer input modalities, such as mouse, keyboard, or cumbersome tracking devices and data gloves, with multimodal interaction and PUI approaches to produce an immersive, coupled visual and audio environment. The ultimate interface may be one which leverages human natural abilities, as well as our tendency to interact with technology in a social manner, in order to model human-computer interaction after human-human interaction. Recent discussions on the incremental value of PUIs over traditional GUIs suggest that more sophisticated forms of bi-directional interaction between a computational device and the human user has the potential to produce a more naturalistic engagement between these two complex "systems". This enhanced engagement is anticipated to lead to better HCI for a wider range of users across age, gender, ability levels and across media types.

3. Project Role in Support of IMSC Strategic Plan

As the underlying enabling EE/CS technologies continue to evolve and allow us to create more useful and usable Integrated Media Systems (IMSs), a continuing challenge involves the design of innovative and more naturalistic user interface methods that are more effective and efficient than traditional methods for certain tasks. As well, the development of computational devices that have the capacity to sense human user facial, voice and gestural actions (from which user state inferences are made) and use of this information to support better interaction and engagement between a user and IMS scenarios is a desirable goal that is in line with the IMSC mission. Advances in our capacity to track, represent and analyze human motor action will also serve to drive IMSC application development that could have significant positive impact on society via applications designed to address motor training and rehabilitation. Further, the "populating" of these environments with believable virtual human representations is a desirable goal for enhancing user "presence" for a variety of IMS interaction and communication purposes. Whether virtual humans represent real humans in real time or are programmed entities with some level of artificial intelligence (agents), the incorporation of "believable" virtual human representations into IMS, that both look and *act* like real people, is a vital research direction across the IMSC strategic plan.

4. Discussion of Methodology Used

Multiple methodologies have been employed across the projects in this area. For the Gestural HCI project, in the previous year, we conducted an informal evaluation that involved the observation of naturalistic gesture commands emitted by 38 users. The sample that was observed comprised of 18 male and 20 female participants with ages ranging from 19 to 68. Users were presented with 31 stimulus slides that contained screen captures across 4 different media types and requested users to "Use what you feel is the best hand gesture for signaling" a certain common command (Figure 1). Twenty-one common commands were selected to represent HCI activities that are typically performed in normal practice using a mouse and keyboard. User gestures were noted and following the standard administration of the stimuli, users went through the commands with the investigator and discussed the rationale for their chosen gestures. The goal was to define the best gestures that would allow a user to effectively and efficiently interact with 3D objects in a biological learning scenario as part of the 2020 Classroom project. In that scenario, users are required to select and manipulate human body parts and organs as part of a guided discovery game-based learning approach to education. More recently we have examined a number of existing computer-based games that were designed to teach anatomy and biological concepts in a digital gaming format to better understand how the mouse and keyboard were used to interact with the content. We explored the content in such game-based learning tools as Electronic Arts "Emergency Room" and the Encyclopedia Britannica teaching game, "My Body, Myself". It was noted that the interaction was limited by mouse and keyboard interaction that was simply lacking in the area of learnability and in the type of natural interaction that would readily support intuitive learning of the content and engage the learner. The challenge here is to make interaction and navigation within digital educational environments as seamless, intuitive and naturalistic as possible. This challenge is particularly relevant since in order for users to be in a position to ultimately benefit from an IMS application (usefulness), they must be capable of learning how to navigate and interact with objects and processes within the IMS environment. Even if a user is capable of using an interface system at a primitive level, the extra non-automatic effort required to navigate and interact can serve as a distraction and limit success in achieving the targeted educational objective. In this regard, Psotka [2] hypothesizes that facilitation of a "single egocenter" found in highly immersive interfaces serves to reduce "cognitive overhead" and thereby enhance information access and learning.



Figure 1. Examples of Stimuli for Naturalistic Gesture Study

These initial investigations provided a wealth of material that served to inform our thinking about what types of gestures might be effective for the universal 3D interaction tasks of object selection, object manipulation, navigation and system control. Armed with this information we

have begun to address the issue of selection and manipulation of objects as part of the 2020 Classroom project. To do this properly, we decided that we needed a benchmarking scenario that would allow us to make comparisons between gesture interaction and traditional mouse/keyboard methods. Based on this need we have created a block configuration methodology that was derived from some of our previous work [3] with Shepard and Metzler [4] mental rotation stimuli (Figure 2).



Figure 2. Shepard & Metzler Block Configurations

These block stimuli have many features that make them attractive as 3D interaction benchmarking stimuli. There is a rich history of research examining human performance with these stimuli, particularly in the area of researching spatial abilities across age and gender [5]. The stimuli are easily capable of being hierarchically presented from very simple renderings too much more complex configurations and in the types of rotations required to perform tasks with them. The influence of prior learning and experience on interaction performance measures with these stimuli is lessened when compared with using realistic objects. And the influence of a person's innate level of spatial ability can be statistically parceled out of the user's hands-on interaction performance with the blocks, by factoring out results from user performance on the highly reliable Vandenberg and Kuse [6] Mental Rotations Test (Figure 3).

Mental Rotations Test (MRT)



Vandenberg and Kuse, 1978

Figure 3. Vandenberg & Kuse Mental Rotations Test

We have now designed a series of progressively more complex block configurations that can be rendered in either mono or stereo mode on a common PC monitor (shutter glasses are required for stereo mode). The task presents pairs of identical block stimuli in orientations (that are

readily controllable by the experimenter) and the user is instructed to select and manipulate one of the blocks and to superimpose it upon the target blocks using a designated input device. When the user completes the trial, data is automatically collected on the speed of successful superimposition and the efficiency of the actual movement as measured by the ratio of the shortest path of travel to the actual path. We are now able to configure the blocks to be actuated by any input device. With our performance metrics, we are able to compare speed and efficiency across whatever input devices that we chose to compare.

Our first test includes a 20-item interface training task that presents very basic block configuration in very simple orientation and depth rotations for interface training purposes. We then follow this training with a more complex 24-item task. Subjects are presented with the tests where the selection and manipulation of the blocks is done with both a 3-button mouse and with simple hand based gesture selection and grasping of the blocks, in counterbalanced fashion. Interestingly, our previous work with 3D Stereo versions of this test found that while females perform more poorly than males on the 2D paper and pencil test, male and female performance was equivalent on this type of hands-on spatial rotation test [3]. As well, by administering the Vandenberg & Kuse [6] MRT, we can statistically parcel out the variability due pre-existing "mental" rotation performance from the actual hands-on task and thereby get a more specific measure of user interface effectiveness and efficiency. This system is currently being used to compare mouse with gesture interaction and will soon test users with a variety of 3D user interface devices that are commercially available. This testing system will provide data to compare interactional methods to determine the best options for interaction in the 2020 classroom and could possibly serve more the general purpose of providing a standardized format for testing new interfacing methods. Such research would allow comparisons to be made with archived normative data to study such factors as: learnability, throughput, fatigue, preference, scale of display, command structures, age/gender and the relative value of gesture behavior compared to standard mouse/keyboard tools and voice interaction for certain commands. The programming of the gestural commands (using 5DT gloves with 6DF magnetic tracking) being used in this work is currently being done in collaboration with Cyrus Shahabi's laboratory.

The UCS component of the **Vision-Based Tracking** project (in collaboration with Isaac Cohen and Ram Nevatia) began last year with the submission of an NSF ITR proposal and with the building of a proof of concept virtual environment that used magnetic tracking of movement for concept demonstration and for collecting baseline data on existing technology for comparison purposes with a future vision-based system. The increasing use of video sensors in daily life environments is motivated by the importance of visual sensing of human activity. While security systems have been the major driving application, vast spectrums of new topics have emerged, such as computer-aided training, multimodal interaction and post-trauma rehabilitation.

Although the ITR proposal was not funded, we have continued the research program to use visual sensing of 3D body motion and its analysis for improving the rehabilitation processes for motor impairments. An essential part of the rehabilitation process for physical dysfunction is the remediation of motor deficits in order to improve the functional ability of the patient and to enable him or her to live as independently as possible. Conventional therapy focuses on muscle strengthening, increasing joint range of motion and improving balance reactions. One of the major challenges facing clinicians in rehabilitation is in the identification of intervention tools

that are effective, motivating, and that promote transfer of functional ability to the "real" world. However, traditional therapies employ intervention tools that tend to be tedious, monotonous and provide little opportunity for systematically grading task difficulty levels. Recent efforts that use single camera "fixed-plane" vision-based approaches have appeared in the literature in this area and have shown promise. These applications produce a representation of the user embedded within a two-dimensional flat screen display environment where they can interact with graphical objects (Figure 4)



Figure 4. Illustration of a single camera vision-based virtual environment used for posttrauma rehabilitation by our collaborators at the University of Haifa. On the left, a stroke patient uses his intact left arm to attain full passive range of motion at the wrist, elbow and shoulder joints of his impaired right arm. On the right, a therapist is providing proximal counter resistance and support to his left shoulder thereby enabling the patient to initiate active movement of the impaired upper extremity.

In existing applications, the user interacts with graphic objects by way of a mirror image of their body (see Figure 4). Such systems have significant limitations; quantifying and understanding 3D body motion from a single visual system is inaccurate since only 2D projection of the body motion is captured by the camera. Moreover, approximately one-third of the body joints are nearly unobservable due to motion ambiguities and self-occlusion. Multiple views are therefore required to quantify, disambiguate and identify the 3D human body motion. Additionally, existing systems do not support the first person perspective that would be more valuable for functional rehabilitation. Within this context, our goal is to use multiple video sensors to accurately detect and track 3D body motion, identify body postures and quantify motor performance within a "first person" virtual environment.

The advantage of visual sensing, compared to magnetic trackers, is that it allows the patient to move freely during sessions with a physical and/or occupational therapist. This provides a better understanding of the patient's range of motion, movement speed, muscle strength, endurance, dexterity and accuracy. The User Centered Sciences group has now spearheaded collaboration with the USC Department of Physical Therapy and Biokinesiology to begin to study these issues. We have also created a network of collaborators within three major centers of physical/occupational therapy (i.e., Haifa University, University of Ottawa, Kessler Medical Rehabilitation Research Center). These centers currently use the 2D vision-based approach, and are limited by these systems with regards to providing accurate 3D body motion measurements. These collaborative centers are poised to conduct the clinical evaluation of the vision-based body motion tracking system that we develop, and will integrate the 3D body measurements obtained during this research into efficient and systematic motor training applications. Thus far, we have

created three proof of concept virtual environments that use magnetic tracking of user movement and have begun user testing to explore issues for scenario development (Figure 5). We are also currently running comparison testing with an established magnetic tracked "object grasping" test that the USC School of PT and Biokinesiology has been using for the last year. We have concurrently captured user movement within the existing magnetic system using high-speed cameras and are comparing performance measurements using this system with the existing magnetic tracking capture.



Figure 5. 3D Virtual Environments to support first person interaction and tracking of performance for motor rehabilitation.

Collaboration with our clinical research partners will allow us to develop the vision-based system with insight derived from the clinical experience of our collaborators on the limitations of existing tools. Additionally, working under the design requirements needed for populations with motor impairments can drive the enabling technology development which should have generalizability to system development with unimpaired populations for general 3D user interface applications.

Another **Vision-Based Tracking** project is embedded in the Communication Vision (in collaboration with Isaac Cohen and Shrikanth Narayanan) that is detailed in the Volume One report on this Vision. In this project we have begun collecting data on human performance on a computer card replacement task. We are testing users first under standard conditions using a hard copy of a manual to guide performance on a 10-step task to replace a computer card on the mock industrial apparatus (Figure 6). Speed and successful completion of each step is recorded from video transcripts of each user trial. This is designed to serve as a baseline condition for

comparison purposes with interactive vision-based sensing systems that observe the actions and state of the user and provide assistance for completion of the steps in the task via an animated virtual character. The animated character serves to represent a remote expert that can observe motor and speech behavior and infer the state of the user in order to modulate the type and manner of instruction and feedback provided to the user.



Figure 6. Scenes from computer card replacement baseline task

Finally, the Body Gesture and the Expression of Emotional State in Virtual Human **Representations** project aims to use the Vicon motion tracking system at the USC Zemeckis Center to capture trained actors who are expressing dynamic body postures and gestures that implicitly communicate emotional states. The body actions captured in this system will be rendered as "faceless" animated characters using 3D projection technology and will be presented to research subjects who attempt to decode the nature of the emotional expressions. No verbal or facial cues will be presented to subjects in an effort to isolate specific body gesture communication efficacy. Actors have been trained to perform the actions described above via collaboration with Sharon Carnicke, Assistant Dean of the USC School of Theatre. We anticipate that this methodology, expanded from the early non-verbal communication literature [7], and the subsequent stimuli generated, will serve to drive a series of studies that will later involve factorial combinations of body, face and verbal cues presented to subjects for judgments. We recently conducted one perceptual judgment study, using simple videotapes of actors under a range of interactional conditions. Subjects observed 10-20 second video clips of actors that were captured from positions where the face was partially occluded and with no sound. The subjects were asked to make open ended, forced choice and rating judgments of the emotional content perceived in the body gesture behavior of the actors. From this data we were able get an initial assessment of the saliency of certain gestures that will be trained in actors working with the Vicon capture system. This research is expected to advance knowledge on the relative communicative value of multiple inputs for creating believable virtual human representations. As well, data on human judgment capacity for this type of communication can serve as the basis for the longer-term goal of creating PUI systems that can perform similar functions in an automated fashion.

5. Short Description of Achievements in Previous Years

Previous achievements by the groups we are collaborating with on the **Gestural HCI** and the **Vision Based Tracking** projects are available on their specific technical project reports found elsewhere in this volume. For the **Gestural HCI** project, initial user testing was conducted with results feeding into the design of hand actions for gesture interaction that we are creating in

support of the 2020Classroom project. Creation of gestures, user evaluation with tracked gloves and iterative redesign of gestures based on user results and technology limitations has been ongoing. For the **Vision-Based Tracking** project, a proof of concept magnetically tracked hand movement VR scenario was created and initial user evaluation of the usability of the system commenced. An NSF ITR was submitted to support the technical development of the system, but was not funded. However, we continued to nurture collaborations with leading international rehabilitation centers and tasks/requirements analysis were conducted with these groups to guide our continued efforts at system design. Our initial work on performance driven facial animation (PDFA) decoding in humans was a precursor for our UCS interest in the **Body Gesture and Expression of Emotional State** area. That work resulted in published research on our methodology and results from the PDFA project [1]. As well, an IMSC website includes the paper and the actual facial action stimuli [8]. This website was created for better dissemination and comprehension of our findings since the uniqueness and complexity of the data presented in the paper are difficult to interpret by a reader without visual inspection of the video and animation stimuli (linked to results) on which it is based.

5a. Detail of Accomplishments During the Past Year

Gestural HCI –Initial user testing is being conducted comparing mouse selection and manipulation of 3D virtual objects with gestural action. To do this, a benchmarking system was created using block rotation stimuli that could serve as standardized stimuli in the future for conducting 3D user interface comparison trials across different devices. We have spent considerable time with the 2020 Classroom researchers using paper prototyping methods to specify the interactional needs and requirements for this application.

Vision-Based Tracking – We continued to expand our collaborators to include the USC Department of Physical Therapy and Biokinesiology. Within this department, we gained access to a standard grasping test and are now running comparison trials between the existing magnetic tracking system and a vision-based system. We also have continued to build new virtual environment testbeds that use magnetic tracking for initial user studies with the aim to replace our magnetic system with vision-based capture. As well, baseline data has been collected on human performance during the card replacement task within the Communication Vision.

Body Gesture and the Expression of Emotional State in Virtual Human Representations – The methodology for this project has been formalized after analysis of the results of our initial videotape gestural perception study. Based on these results, actors from the USC School of Theatre are being trained on the key gestures for use in the motion capture study. Vicon motion capture of these actions will commence once training is complete. Rendering and post-production of the animated characters to be used in a formal decoding study planned to commence during the summer of 2004.

6. Other Relevant Work Being Conducted and How this Project is Different

Gestural HCI – While the field of multimodal interaction with computers has been quite active over the years, we were unable to locate an existing gesture-based language that met our requirements for the 2020 Classroom project. Our requirements are based on developing a one-handed set of commands that have fast learnibility (for novice users) while promoting efficient throughput (efficiency with expert users) that will scale across a wide range of media types and

display systems. Also our methodology is designed for evolving this work in the future to evaluate the comparative value of hand gestures relative to other input devices via our new 3D User benchmarking application.

Vision-Based Tracking – recent efforts that incorporate simple single camera "fixed-plane" vision-based approaches have appeared in the literature in this area, and have shown promise [9-12]. These applications have used a single camera vision-based tracking system that produces a representation of the user embedded within a two-dimensional flat screen display environment. The users can interact with graphical objects in this environment as depicted in Figure 4. The origins of this work can be traced back to Krueger's [13] seminal "Videoplace" application in the early 1970s, where it was observed that humans were compelled to interact with graphic objects displayed in this format. In the early 1990s, the Vivid Group [14] designed and marketed a series of single camera vision-based applications as location-based arcade entertainment systems. Known as the Gesture Xtreme VR System, it uses a blue backdrop and a chroma key to separate the user's image from the background. This system has now come to be embraced by motor rehabilitation specialists as a research and clinical tool for the treatment of motor impairments. Our aim is to improve on this system by applying vision-based capture to support 3D movement tracking and quantification as well as first person, 3D interaction.

Body Gesture and the Expression of Emotional State in Virtual Human Representations – The uniqueness of this research is based on its use of previous decoding methodology developed at IMSC [1], its use of the Vicon tracking system and on the integration of extensively trained actors to produce the stimuli that will be presented to human judges.

7. Plan for the Next Year

Gestural HCI – We will continue user testing needed to iteratively evolve the 3D user interface benchmarking stimulus system. Evaluation research will compare value of hand gestures relative to other input devices to produce effective interaction with the game-based learning content created with the 2020 Classroom project. We will also explore vision-based tracking options to replace the current tracked glove device that we are currently testing.

Vision-Based Tracking – We will focus on developing and evaluating the performances of the vision-based algorithms within specific motor rehabilitation VE and real object grasping scenarios. The evaluation of the accuracy of the vision-based body motion estimation will continue to be performed using magnetic sensors and optical markers for comparison studies. Our collaborators will provide input for the design of scenarios of interest for the rehabilitation applications.

Body Gesture and the Expression of Emotional State in Virtual Human Representations We intend to complete the first round of decoding studies with human raters. Knowledge gained in this round will be incorporated into the next study to expand our library of gestural behaviors that underlie non-verbal expression of emotion. This knowledge will also support our development of PUI technology.

8. Expected Milestones and Deliverables

We plan to continue our system development in this area and produce:

User studies that build a database for standardizing our 3D user interface benchmarking system for the Gestural HCI project.

Continue integration of gesture interaction, where demonstrated to provide added value over existing input devices, into selected 2020 Classroom media applications and begin user testing across a range of display systems.

Continue user studies with magnetic tracked interaction scenarios for range of motion testing for the Vision-Based tracking project.

Create a number of scenarios with collaborative input that integrates gaming features into the interactional format for future testing of vision-based system.

Complete 1st Body Gesture and the Expression of Emotional State in Virtual Human Representations decoding study.

Publish and present results from this research.

9. Member Company Benefits

If found to be usable and useful, the knowledge generated by these tools and methods will offer better ways for users to interact with information technology, whether operating on basic tasks or in more complex immersive environments, and could be of value for IMSC corporate sponsors.

10. References

[1] A.A. Rizzo, U. Neumann, R. Enciso, D. Fidaleo, J.Y. Noh, Performance Driven Facial Animation: Basic research on human judgments of emotional state in facial avatars, CyberPsychology and Behavior, 4:4, 471-487, 2001.

[2] J. Psotka, Immersive training systems: Virtual reality and education and training, Instructional Science, 23, 405-431, 1995.

[3] Rizzo, A.A., Buckwalter, J.G., Bowerly, T., McGee, J., van Rooyen, A., van der Zaag, C., Neumann, U., Thiebaux, M., Kim, L., Pair, J. & Chua, C. (2001). Virtual Environments for Assessing and Rehabilitating Cognitive/Functional Performance: A Review of Project's at the USC Integrated Media Systems Center. *Presence: Teleoperators and Virtual Environments*.Vol. 10 (4), 359-374.

[4] Shepard, R.N. & Metzler J. (1971). Mental rotation of three-dimensional objects. *Science*. 171, 701-703.

[5] Voyer, D., Voyer, S., & Bryden, M. P. (1995). Magnitude of sex differences in spatial abilities: A meta-analysis and consideration of critical variables. *Psychological Bulletin*, *117*, 250-270.

[6] Vandenberg, S. G., & Kuse, A. R. (1978). Mental rotations, a group test of threedimensional spatial visualization. *Perceptual and Motor Skills*, 47, 599-604.

[7] H.G. Wallbott, Bodily expression of emotion, European Journal of Social Psychology, 28, 879-896, 1998.

[8] <u>www.USCAvatars.com/MMVR</u>

[9] D.L. Jaffe, Use of virtual reality techniques to train elderly people to step over obstacles, Paper presented at the Technology and Persons with Disabilities Conference, Los Angeles, CA.,1998.

[10] D. Cunningham, M. Krishack, Virtual reality promotes visual and cognitive function in rehabilitation. CyberPsychology & Behavior, 2, 19-23,1999.

[11] R. Kizony, N. Katz, H. Weingarden, P.L. Weiss, Immersion without encumbrance: adapting a virtual reality system for the rehabilitation of individuals with stroke and spinal cord injury. In Sharkey, P., Lanyi, C.S. & Standen, P. (Eds.), Proceedings of the 4th International Conference on Disability, Virtual Reality, and Associated Technology. Reading, UK: University of Reading, 55-62, 2002.

[12] Rand, D., Kizony, R., Feintuch, U., Katz, N., Josman, N., **Rizzo, A.A.** and Weiss, P.L. (in press). Comparison of two VR platforms for rehabilitation: Video capture versus HMD. *Presence: Teleoperators and Virtual Environments*.

[13] M. Krueger, Environmental technology: Making the real world virtual, Communications of the ACM, 36, 36-37, 1993.

[14] Vivid Group Website: <u>http://www.vividgroup.com/</u>