Interaction Tomorrow
SIGGRAPH 2007, Course Notes #34

Co-Organizer

Michael Haller
Upper Austria University of Applied Sciences
haller@fh-hagenberg.at

Chia Shen
Mitsubishi Electric Research Laboratories (MERL)
shen@merl.com

Lecturers

Gerald Morrison
Smart Technologies Inc.
gerald.mirroson@smarttech.com

Bruce H. Thomas
University of Southern Australia
bruce.thomas@unisa.edu.au

Andy Wilson
Microsoft Research
awilson@microsoft.com

Course Web Page
http://www.interactiontomorrow.org
Summary statement
This course provides a comprehensive overview to user interface technologies on the newly emerging interactive tabletops and large wall displays. The course will cover input devices, interface metaphors, modality of interaction, sensing technologies, applications, and future directions. Materials will be drawn from both commercial systems and research prototypes.

Course abstract
Interactive tables and walls are becoming increasingly popular. Large augmented surfaces are already part of our physical environment. These newly emerging form factors require novel human-computer interaction techniques. Although movies such as Minority Report and the Island popularized the idea of novel, off-the-desktop gesture-based human-computer interaction and direct manipulation-based interfaces, in reality, making the interactions with a digital user interface disappears into and becomes a part of the human-to-human interaction and conversation on these large interactive surfaces is still a challenge. Conventional metaphors and underlying interface infrastructure for single-user desktop systems have been traditionally geared towards single mouse and keyboard-based WIMP interface design, while people usually meet around a table, facing each other. A table/wall setting provides a large interactive visual surface for groups to interact together. It encourages collaboration, coordination, as well as simultaneous and parallel problem solving among multiple people.

In this course, we will describe particular challenges and solutions for the design of direct-touch tabletop and interactive wall environments. The participants will learn how to design a non-traditional user interface for large horizontal and vertical displays. Topics include physical setups (e.g. output displays), tracking, sensing, input devices, output displays, pen-based interfaces, direct multi-touch interactions, tangible UI, interaction techniques, application domains, current commercial systems, and future research.

Prerequisites
This course is self-contained and does not assume prior knowledge of interactive environments. Attendees should be familiar with the basics of traditional human-computer interaction, computer graphics, image processing, and interactive media.

Intended audience
This course is intended for students, researchers, and industrial developers who are either involved in, or are interested in gaining an understanding of the design and implementation of interactive environments. People with general interest in HCI, or in alternative interface design, tangible interfaces, gestural input, and interactive environments will find this course informative.
Syllabus

Wednesday, 08/08/07

Half-Day, 1:45 - 5:30 pm

Level: Beginning

<table>
<thead>
<tr>
<th>Time</th>
<th>Topic</th>
<th>Presenter</th>
</tr>
</thead>
<tbody>
<tr>
<td>1:45</td>
<td>Welcome and Introduction</td>
<td>Michael</td>
</tr>
<tr>
<td>2:00</td>
<td>Input devices and sensing technologies</td>
<td>Andy</td>
</tr>
<tr>
<td>2:15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2:30</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2:45</td>
<td>Interactive tabletops: user Interface design, metaphors and gestures</td>
<td>Chia</td>
</tr>
<tr>
<td>3:00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3:15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3:30</td>
<td>Break</td>
<td></td>
</tr>
<tr>
<td>3:45</td>
<td>Interactive wall displays: Interaction techniques and commercial</td>
<td>Gerald</td>
</tr>
<tr>
<td>4:00</td>
<td>applications</td>
<td></td>
</tr>
<tr>
<td>4:15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4:30</td>
<td>Pen-based Interfaces: commercial systems, new research usage</td>
<td>Michael</td>
</tr>
<tr>
<td>4:45</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5:00</td>
<td>What’s next? Outlook to the future</td>
<td>Bruce</td>
</tr>
<tr>
<td>5:15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5:30</td>
<td>Summary and Conclusions</td>
<td>Chia</td>
</tr>
</tbody>
</table>

1. Welcome and Introduction
2. Input devices and sensing technologies
   - Motivation
   - Sensing technology
     - computer vision
     - accelerometers, and
     - others enable novel interaction techniques beyond the desktop
   - New Opportunities and challenges
     - Applications
     - How to handle uncertain input.
   - Reality-based computing
3. Interactive tabletops: user Interface design, metaphors and gestures
   - Tabletop Fundamentals: Why
     - Why touch and multi-touch
     - Horizontality vs vertical displays
     - Social – Collaborative vs single users
   - Tabletop Gestures: How
     - From clicks to touches
     - From postures to gestures
     - From one-hand to bimanual
   - Tabletop Applications: What
     - Tabletop design from microscopic documents to astronomical galaxies
     - DiamondSpin
     - UbiTable
     - DiamondSpace
   - Tabletop Integration: When
     - In which domains are tabletops appropriate?
     - Integrating multiple display configurations
     - Direct vs. indirect input
4. Interactive wall displays: Interaction techniques and commercial applications
   - Introduction
   - Large interactive displays in the past and present
   - A quick review of touch technologies
   - From small size to wall size, a versatile touch technology
   - Why touch walls?
   - Touch walls in research - the future
   - Commercial applications
   - Conclusion

5. Pen-based Interfaces
   - Motivation
   - Technology
     - 3d approach
     - 2d approach
   - Applications
     - ButterflyNet
     - ModelCraft
     - Shared Design Space, Interchange of Ideas

6. What's next? Outlook to the future
   - Wearable Computing Interfaces
     - Information on the go
     - Garment integrated technologies
   - Augmented/Mixed Reality Interactions
     - What is augmented/mixed reality?
     - Indoor application examples
     - Outdoor application examples
     - Through walls collaboration

7. Summary and Conclusions
Lecturer Biography

**Michael Haller**

Michael Haller is senior lecturer at the department of Digital Media of the Upper Austria University of Applied Sciences (Hagenberg, Austria). He obtained his MS and PhD at the Johannes Kepler University of Linz. Michael is active in several research areas, including interactive computer graphics, augmented and virtual reality, and human computer interfaces. His current focus is on innovative interaction techniques and interfaces for next generation working environments. In 2004, he received the Erwin Schrödinger fellowship award presented by the Austrian Science Fund for his stay at the HITLabNZ, University of Canterbury (New Zealand), and the IMSC, University of Southern California (USA).

http://www.fh-hagenberg.at/staff/haller

**Gerald Morrison**

Dr. Morrison is an Electrical Engineer who has worked for SMART Technologies Inc. for more than 11 years. During that time he has held many roles in development, research and management. All of his roles have related to the development of new interactive touch technologies and products. He is currently the External Research Manager and is actively working with individuals and organizations developing interactive and collaborative technologies. Dr. Morrison has published several papers in conferences and journal publications and is listed as inventor or co-inventor on several patents and patent applications.

http://smarttech.com/

**Chia Shen**

Chia Shen (Ph.D 1992), a Senior Research Scientist at Mitsubishi Electric Research Labs in Cambridge, Massachusetts USA, has published over 40 papers and given numerous keynotes and lectures on the design and development of newly emerging multi-touch, multi-user tables and interactive spaces. She is leading one of the major research groups in this area (diamondspace.merl.com) that has extensively and systematically studied human–computer interactions and interfaces for direct-touch surfaces. Prior to working in HCI, she had over ten years of experience in parallel and distributed real-time systems. She is ACM UIST 2007 Conference Chair and is on the Editorial Board of ACM Computers in Entertainment.

http://www.merl.com/people/shen

**Bruce H. Thomas**
I am the current Co-Director of the Wearable Computer Laboratory at the University of South Australia. I am currently a NICTA Fellow, CTO A-Rage Pty Ltd, and visiting Scholar with the Human Interaction Technology Laboratory, University of Washington. I am the inventor of the first outdoor augmented reality game ARQuake. My current research interests include: wearable computers, user interfaces, augmented reality, virtual reality, CSCW, and tabletop display interfaces. My academic qualifications include the following: 1) B.A. in Physics, George Washington University; 2) M.S. in Computer Science, University of Virginia with a thesis titled: Pipeline Pyramids in Dynamic Scenes; and 3) Ph.D. in Computer Science, Flinders University with a thesis titled: Animating Direct Manipulation in Human Computer Interfaces My experience includes working at the School of Computer and Information Science, University of South Australia since 1990. I have run my own computer consultancy company. I was a Computer Scientist at the National Institute of Standards and Technology (A major US government laboratory for the Department of Commerce.), and a software engineer for the Computer Sciences Corporation and the General Electric Company.


**Andy Wilson**

Andy Wilson is a member of the Adaptive Systems and Interaction group at Microsoft Research. His current areas of interest include applying sensing techniques to enable new styles of human-computer interaction, as well as machine learning, gesture-based interfaces, inertial sensing and display technologies. Before joining Microsoft, Andy obtained his BA at Cornell University, and MS and PhD at the MIT Media Laboratory.

http://research.microsoft.com/~awilson.
Welcome

Michael Haller

A digital world with 1000 interfaces
Goals

- Get an overview of current devices
- Possibilities we have now
- Explore alternative methods/devices
- See new applications in research and industry
Lecturers

- Michael Haller
  - Upper Austria University of Applied Sciences
- Gerald Morrison
  - SMART Technologies
- Chia Shen
  - MERL
- Bruce H. Thomas
  - UNISA
- Andy Wilson
  - Microsoft Research

Survey

- Background
  - Computer Graphics, VR/AR/MR, HCI
- Work
  - How many of you are involved in development?
  - How many of you are involved in design of interfaces?
- Field of work
Overview

- Input devices & sensing technologies  Andy
- Interactive tabletops  Chia
- Interactive wall displays  Gerald
- Pen-based interfaces  Michael
- What's next  Bruce

WWW.INTERACTIONTOMORROW.ORG
Input devices & sensing technologies

Andy Wilson
Adaptive Systems and Interaction

In the “future”...

Sensing technology can enable a wide variety of new interactions

As hardware approaches “free”, we can afford a diversity of form factors
  already have: phones, tablet, TV, car, console game
  will have: walls, tables, rooms, ?

Not every device will be used to do email!

Devices can and should be pleasing to use, as well as useful.
TouchLight

an *imaging touch screen* with some unique capabilities

Two webcams +
DNP Holoscreen +
IR illuminant

[video](#)

DNP HoloScreen
Image Processing
Edge Maps

Potential sensing capabilities

**On surface**
- Visual barcodes, object recognition
- Document scanning (helps to have a transparent surface!)
- Gesture-based manipulation of onscreen objects
- IR stylus
- Multiple hands

**Off surface**
- Face detection/recognition, gaze, person-tracking, awareness
- Stylus + hand combination, hands in space
- IR remote/pointing device, tracking of devices in 3D

...essentially all vision-based perceptual user interface techniques...
Volumetric Imaging Touch Screen

TouchLight has no transition cost from on surface to off-surface

Removes the ‘blind spot’ between ‘on the surface’ and 2 feet away

Applications

Post WIMP, direct manipulation, Minority-report gesture-based interfaces
Eye to eye videoconferencing
ClearBoard (Ishii) redux
Visible light surface scanning
2.5D interfaces
Spatial displays
Magic mirror
Augmented reality
Tables and other direct manipulation form factors
Really, What’s the Killer App?

What is the whiteboard’s killer app?
PlayAnywhere: A Compact Tabletop Computer Vision System

“Lunchbox” interactive vision system
PlayAnywhere

Short-throw projector, very wide angle lens on the camera
Lens distortion & projective transform correction for camera-projector alignment
Off-axis IR LED illuminant

Very few assumptions about the appearance of the surface
All calibration is done at the “factory”

Front Projection Vision Systems

Ceiling installation of projector is difficult, dangerous
Not easily moved
Vibrations in the building are a problem
User’s head and hands occlude the system

Digital Desk (Wellner), EnhancedDesk (Koike), Augmented Surfaces (Rekimoto), I/O Bulb (Underkoffler), Visual Touchpad (Malik)

Also see SenseTable (Patten), DiamondTouch (Dietz), SmartSkin (Rekimoto)
Rear Projection Systems

Self-contained device
Leg room and screen size are difficult to balance
Housing can be large, heavy

MetaDesk (Ullmer), Perceptive Workbench (Leibe),
Designer’s Outpost (Klemmer), HoloWall
(Matsushita)

PlayAnywhere

Portable
As long as sitting on the same plane, no need to calibrate after moving
Occlusion by hands, not heads
Decently large projection
Allows legs under the table, but
One side of the table is effectively blocked
Detecting touch is tricky
Vision for Sensing

High computational cost
Low frame rate, high latency
Precision, noise
Calibration

But...

Extreme flexibility

Hands, fingers, visual tags, pages, tangrams, dice, textures, object recognition, OCR ...

Image Processing

Input  Lens distortion, projective distortion removed
Shadow-based touch

Shadow as second projection

Shadow-based hover
Paper tracking

30Hz, based on finding strong lines

Page tracking

• Sobel edge detection is based on gradients in the horizontal and vertical directions.
  
• The strength of an edge at \((x, y)\) is

\[
\begin{bmatrix}
-2 & 0 & +2 \\
-1 & 0 & +1 \\
0 & +1 & 0 \\
0 & +2 & 0 \\
1 & 0 & +1
\end{bmatrix}
\]

• Orientation of the edge is
Page tracking

• At each pixel location (x,y) we have |G| and theta

• Transform (x,y, theta) to (r, theta), where r is shortest distance from the origin to the line

• Histogram |G| over (r, theta)

Page tracking

To find an 8” x 11” rectangle, look for specific pattern of peaks in the histogram
Fast visual codes

Read edge orientation, rotate and read 12 bits blindly, compare against list of known bit patterns
Hough transform for circles, computed from edge image

Rotating, Scaling, Translating Objects

Existing approaches to freeform manipulation of objects, e.g. photos

Decorate the object with widgets
  Visually cumbersome
  Requires training
  Reduces the immediacy of tangible UIs

Track distinct objects (contacts) and compute movement
  Assumes good tracking/correspondence frame to frame
  Sometimes difficult to define an object in this scheme
  Gross manipulation with the hand is often limited to translation
Tracking is Hard

We fall for it because we have such a strong notion of the *cursor*

See kids interacting with tables!

Tracking rarely allows for graceful failure

Tracking reduces the richness of human motion to that of a gnat

Flow Move

Summarize optical flow field as simultaneous translation, rotation, scaling

[PlayAnywhere VE](#)
Flow Move

Translation  Rotation  Scaling

Can solve for simultaneous rotation, translation, scaling via least squares

Surface Computing Challenges

Technical

- Projection: somewhat doable now, affordable tomorrow
- Sensing: still research

Interaction Design is still an open question

- How does it work?
- How to break out of how things work today?
  - WIMP isn’t a great fit
  - The key may be diversity of UI
  - What is it really good for? Intuition only gets you so far.
**Bluetooth photo synch**

How does it work?

Detect phone-shaped object?
Connect to each Bluetooth device:
  - Is it advertising our software service?
  - Command it to blink the IR port on the phone
  - Did vision system detect IR port blinking?
  - Start synch over Bluetooth

---

**PlayTogether**

[video]
IR Laser Pointer Tracking

Track shaped laser pointer (hologram)
Theoretically 6 degree of freedom
Today, 4: position, depth, roll

Mini PlayAnywhere
Future Devices

Canesta, VKB, Virtual Devices

Symbol laser projector
A common problem in vision-based HCI

Tracking the hands, but... how to “drop” it?

How to get to Buxton's 3-state model?

null tracking selection pen up
hands off moving click pen down
Pinching: touching thumb and forefinger

Unambiguous to the user
Discrete signal maps to discrete input
Stable transition in and out

Discrete sensing for discrete state

- **button state**
  - not clicked
  - clicked
- **hand size**
  - closed hand
  - open hand
- **pinch state**
  - not pinched
  - pinched
Pinching: touching thumb and forefinger

Ergonomics

Natural analogues:
- Tugging on a piece of fabric
- Using a stylus
- Picking up a small object

Some previous pinching work

VideoDesk (Myron Krueger)  Fakespace Pinch Gloves
Visual TouchPad (Malik and Laszlo)  Sato et al (demo, this conference)
Above the keyboard vision

Quek & Mysliwiec, *FingerMouse*

Kjeldsen & Kender

Wilson & Cutrell, *FlowMouse*

Recognition problem
The technique

One shape
Not pinching

Two shapes
Pinching

Connected components
Consider an image as an undirected graph where each node corresponds to a pixel, and each node has edges to neighboring nodes (pixels) of the same value.

A set of pixels is a connected component if for every pair of pixels $u$ and $v$ there is a path from $u$ to $v$.

A connected component can often correspond to a distinct object.
Image processing

Background subtraction
Connected components analysis
Count the number of components, pick the smallest

Cursor control

Pinching is a natural “clutch”
“tap and a half” for click
Open, close, open
Dragging
Open, close quickly
Free transforms

Translation: change in position

Rotation: change in orientation of ellipse

Scale: change size of ellipse

Two hands
Limitations

Only as robust as the segmentation
Dependent on line of sight
Not a full 3-state interaction model
Tracked position is not the finger tip
  Implications for direct manipulation framework
  Motion is (mostly) relative, like the mouse

The Orb Platform

XWand, CHI 2003
The Orb Platform

New hardware, designed in coop with Steve Bathiche (MS Hardware), and Mike Sinclair (MSR)

- 3 magnetometers
- 2 MEMS gyros
- 3 MEMS accelerometers
- Bluetooth support

Vast improvement in orientation sensing over original XWand

Applications in Spotlight, VIBE wall large display, PAN with SmartPhone

More of a platform approach
- Layout/PCB is easy, software is not, & people are picky about form factor
- Serve a variety of needs around MSR

Orientation with Magnetometers & Accelerometers

3 mags or 3 accels alone doesn’t cut it, but combination does

Take cross product of mags and accels

Caveats
- Only correct when still
- Magnetic north wanders indoors
- This formulation gives priority to one of the sensors
- N and g must not be colinear
Coffee Compass

with Raman Sarin

Most interfaces attempt to do too much
   And become unusable as a result

Coffee compass is familiar,
   kitschy, easy-to-use,
   delightful, humanizing

“where’s the nearest Starbucks?”

Questions
Interactive tabletops

User Interface, Metaphors and Gestures

Chia Shen

Why Tabletops?

What's wrong with this picture?
- “Single” point of control in “group” interaction
Why Tabletops?

- Touch and multi-touch
- Horizontality vs vertical displays
- Social – Collaborative vs single users

12 Challenges of direct-touch tabletops

<table>
<thead>
<tr>
<th>What: Physical Platform?</th>
<th>How: Social?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Digital Tabletops</td>
<td>Multiple people</td>
</tr>
<tr>
<td>Interactive Rooms</td>
<td>Walk-up usage</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Input resolution</td>
<td>Touching</td>
</tr>
<tr>
<td>Orientation</td>
<td>Perceiving</td>
</tr>
<tr>
<td>Occlusion</td>
<td>Collaborating</td>
</tr>
<tr>
<td>Bimanual (2 hands)</td>
<td></td>
</tr>
</tbody>
</table>
Tables are not desks:

**Multi-user:**
Face-to-face and simultaneous operations

**Arbitrary orientation of objects:**
Multiple rotating documents.
Anti-alias continuously, especially text

**Rotation of the entire virtual tabletop:**
Mega-pixel operation
rotation-sensitive items.

Direct-touch interfaces have two inherent characteristics:

1. UI targets can be occluded by the fingers, and
2. the large touch-area of a fingertip needs to be mapped to a screen-point.

System defined UI-controls, such as window manager decorations (min/max), sliders, buttons, menus, etc., need to be reengineered to allow for direct-touch input. Changes include the need to make these controls larger, to compensate for imprecise input, and reposition them to minimize occlusion of window content while in use.
When working on a tabletop, there are many fundamental differences in the user input paradigm that challenge conventional WIMP assumptions:

1. The input area of any given user/point is larger, and a different shape, than the single pixel defined by a mouse pointer.
2. Windows/widgets may be rendered off-axis (rotated).
3. A single user may be touching multiple UI-controls/screen points simultaneously.
4. Multiple users may be touching the same UI-controls/screen points simultaneously.
5. Multiple users may be touching different UI-controls/screen points simultaneously.
6. Multiple users may be entering text from different keyboards (soft or hardware) simultaneously.

Example: A multi-user multi-touch tabletop
Example solution for “fat fingers”:

*Interactive Corner Handle Feedback*

**Before touch**

![Before touch image]

**Upon touch**

![Upon touch image]

**Multiplicity of Tools and Hands When Working on a Table**

![Tools and hands image]
Postures vs. Gestures

- Postures – Static hand marks
- Gestures – Time and space varying postures
  - E.g., sliding two hands apart to “zoom out”

A Posture Engine

A possible posture recognition process:
- Mean value and covariance matrix for each pre-recorded posture
- Dynamic touch data is converted into with pre-recorded posture 26 unique features
- Use univariate Gaussian algorithm to compare
- Posture detection accuracy ~ 80%
- A set of heuristics are added to improve accuracy up to ~95%
Multi-User Multi-Hand Gestural Input

How to support both conventional point-based interaction and freehand gestures within the same environment? How to multiplex tools? And multimodal interaction…?

A fundamental gesture design framework
- Gesture Registration
- Gesture Reuse
- Gesture Relaxation

- Supported by a generic gesture engine

Gesture Registration

- Problem: Multiple interaction styles and tools can be present – pointers, cursors, stylus, free-hand gestures.
- Solution: Assign a “start” gesture to facilitate seamless and fluid mode change with a simple registration phase:
  - Transitioning a pen or a finger between being a pointing device and a writing device
  - Start of a particular gestural interaction with a particular context
Gesture Relaxation

• Problems:
  (1) Maintain fairly precise hand postures requires muscular tension.
  (2) High variability of sensed data due to physical affordances of a touch tabletop: height, size and reach, location, user position.

• Solution:
  Gestural interface allows relaxed hand postures and movements.

Gesture Reuse

• Problem: A large set of gesture primitives both burdens the users in memorizing gestures, and the system in having to recognize many different patterns.

• Solution: Employing the same gesture, including hand postures, finger touches or stylus, to accomplish different tasks within different context.
ExpressiveTouch Gesture Video

Gesture State Transition Diagrams

- **Annotate**
  - Asymmetric bimanual

- **Wipe**
  - Unimanual

- **Cut/Copy-n-Paste**
  - Asymmetric bimanual continuous compound

- **Pile-n-Browse**
  - Symmetric bimanual
DiamondSpin – A Java Tabletop Toolkit

**Multi-user:**
Face-to-face and simultaneous operations

**Arbitrary orientation of objects:**
Multiple rotating documents.
Anti-alias continuously, especially text

**Rotation of the entire tabletop:**
Mega-pixel operation rotation-sensitive items.

---

Tabletop Space Management 1/2

Black-hole
Tabletop Space Management 2/2

Fisheye

Draggable Popup menus (CoRDs)

Movable to anywhere around the table, rooted at the context.

Multi-user, or privileged.

Contextual: for Element or for Background
Using DiamondSpin: Many Polygonal Shapes of Tabletops

Using DiamondSpin: Subdivision of Tabletop into Regions

- Dividing tabletop area into separate semantic regions
  - Regions can have both distinctive visual layout and semantics.
  - Policy issue – As a document is moved across a region, how should the document’s properties change?
UbiTab - Walk-Up Interaction

The UbiTable provides a large shared area for interaction:

- The table supports the connection of multiple devices
  - Laptop, USB device, Camera, PDA
- We use personal spaces to denote a user’s area of control
- Multiple devices can be put in each personal space
- Blue portals are used to copy content and transfer+access files

DiamondSpace

DiamondSpace
ACM UIST 2006
Example Interactive Tabletop Projects

DiamondSpace
ACM UIST 2006

PDH Table
CSCW 2002

Under-the-Table
ACM UIST 2006

UbiTable
UbiComp 2003

DT-Lens
ACM UIST 2005

DiamondSpin
CHI 2004

Citations

http://www.diamondspace.merl.com/publications.php


Shen, C., Everitt, K., and Ryall, K. UbiTable: Impromptu Face-to-Face Collaboration on Horizontal Interactive Surfaces. in Proceedings of the 5th International Conference on Ubiquitous computing (Seattle, WA, USA, October 12-15, 2003), Springer, 281-288.
Interactive wall displays: Interaction techniques and commercial applications.

SIGGRAPH 2007
San Diego
August, 2006

Presented by
Dr. Gerald Morrison, Ph.D.
External Research Manager
SMART Technologies Inc.

Interactive Whiteboards

• An electronic version of a traditional dry erase marker whiteboard
• Used for remote education, distance collaboration, training, whiteboard capture, etc.
• SMART Technologies Inc. is the largest player in interactive whiteboard market.
• Building Interactive Whiteboards for over 15 years.
• We have lot’s of experience in what works and what the customer wants.
The Blackboard as Technology

“… in the winter of 1813 & ‘14, during my first College vacations, I attended a mathematical school kept in Boston by the Rev. Francis Xavier Brosius . . . On entering his room, we were struck at the appearance of an ample Black Board suspended on the wall, with lumps of chalk on a ledge below, and cloths hanging at either side. I had never heard of such a thing before. There it was—forty-two years ago—that I first saw what now I trust is considered indispensable in every school—the Black Board—and there that I first witnessed the process of analytical and inductive teaching.” [May 1855]

Introduction of the Blackboard

- James Pillans, Headmaster of the Old High School of Edinburgh, Scotland, is widely credited for inventing the blackboard but there is some uncertainty as to who really created it (and colored chalk).
- An instructor at West Point Military Academy named George Baron, is considered to be the first American instructor to incorporate the use of a large black chalk board into the presentation of his math lessons in 1801. However, it’s probable that a few other schools had access to it, also.
- By 1856, 72% of schools in Canada were using Blackboards. Five years later (1861), 83% of schools had them and in another five years, over 90% of schools were using them.
Before the Blackboard

- Prior to the introduction of the blackboard, students and teachers used handheld slates.
- Teachers would then have to go from student to student copying, for example, a math problem onto each student's individual slate.
- When they could not afford slates, teachers would write on the back of students hands.
- The blackboard changed the way people were educated and it became the single most important educational tool in the 1800's and for most of the 1900's.
- More than 200 years after it’s introduction, it is still in use!

Large Area Displays

- The Blackboard is a large area interactive information display.
- A lifespan of over two centuries speaks to the enduring need for this technology.
- Information has significantly changed and we are now in the “digital” age where information is easily, shared, stored, transmitted, disseminated, etc.
- But the need for large area interactive displays goes on …
What it Was Like Then…

The little one-room school at High Street’s Granard Flats series is reminiscent of a time when each school served as a centre for schools, church halls, and community centres.

And Where We Want to Go …
Today

Many Interactive Displays are Available Today

- Electronic White/Black Boards.
- Tablet PC’s.
- Touch overlays for flat panel displays.
- Portable and desktop systems.
- etc.
A Quick Side Tour on a Specific Touch Technology

- **Machine Vision Touch** – use CMOS cameras to look across the display to detect the presence of a pen or finger.

Integration of low cost CMOS sensors and powerful DSP’s allow for Smart Camera construction.

Plasma Display with a Machine Vision Interface

Smart Cameras analyze the scene and send pointer information to a master controller that reports the pointer position to a computer.
What the Camera Sees

Image processing is done on the Smart Cameras. A real-time image is captured and then analyzed for important information. Below is an example of a finger approaching the display surface.

After processing, characteristics such as finger tip position, can be identified. The image below shows an example.
Collaboration with Academic Research

Touch Wall – University of Toronto

• Movie
Touch Wall Projector Matrix (3x6)

• Movie

Georgia Tech BigBoard

The BigBoard is a touch sensitive SmartBoard measuring 17.5 wide and 4.5 high. Using Virtual Rear Projection technology for computer output, we are able to provide a large interactive surface without using rear projection.
Georgia Tech Project

- June 2003 - December 2003
- Research project to adapt DViT to touch enable a 4.8’ x 17.11’ front projection touch surface for Georgia Tech
- Display: “Virtual Rear Projection”, 6 projectors but effectively 3
- Each at approx. 1024 x 768

- Movie

Touch Table – University of Calgary

- Movie
Other Touch Walls

• Several other systems in both academia and industry are either under construction or being proposed.
• More to follow …
A CMOS Camera-Based Man-Machine Input Device for Large-Format Interactive Displays

Dr. Gerald D. Morrison*
SMART Technologies Inc., 1207 – 11 Ave. SW, Suite 300, Calgary, AB, Canada T3C 0M5

ABSTRACT

Human-computer interaction using large-format displays is an active area of research that focuses on how humans can better work with computers or other machines. In order for this to happen, there must be an enabling technology that creates the interface between man and machine. Touch capabilities in a large-format display would be advantageous as a large display area is informationally dense and touch provides a natural, life-size interface to that information. This paper describes a new enabling technology in the form of a camera-based man-machine input device which uses smart cameras to analyze a scene directly in front of a large-format computer display. The analysis determines where a user has touched the display, and then treats that information as a mouse click, thereby controlling the computer. Significant technological problems have been overcome to make the system robust enough for commercialization. The paper also describes camera-based system architecture and presents some interesting advantages as well as new capabilities. The technology is ideally suited to large-format computer displays, thus creating a very natural interface with familiar usage paradigms for human-computer interaction.

Keywords: Human-Computer Interaction, Machine Vision, Large-Format Touch Display, Computer Vision, Smart Cameras

1. INTRODUCTION

Research and applications in man-machine interfaces or devices often use touch as an enabler. The ability to touch a computer display and manipulate displayed objects or control applications is a natural extension of everyday human experience where touch is a critical sense. As a result, many types of touch systems have been developed to address this need. There are two types of touch interfaces available and they can be broadly classified as passive touch or active touch. Passive touch systems are the most intuitive and do not require the use of a special pointer device in order to interact with a computer. A passive pointer can be a finger, a cylinder of some material, or any suitable object that can be used to point to something of interest on a computer display. Active systems need special proprietary pointing devices that usually require some power source for normal operation of the pointer. The advantage of using a passive touch system, beside the fact that it is more natural for human computer interaction, is that the user need not be concerned about issues such as batteries, damage, theft or misplacement of the special pointer. The most common and intuitive passive pointer is the user’s finger.

Large-format displays are very natural for man-machine input devices as they allow the user to work with a computer application in sizes that are comparable with the size of the human body. Human-body sized interfaces facilitate interactivity with the user and collaboration among several users. Although many touch technology systems are suitable for small display types, not all commercially available touch screen technologies are suitable for very large format displays. Machine vision, however, is quite suitable and can scale from small display formats to very large display formats including wall-sized displays. Vision systems are attractive because they are more versatile than other input devices since they add capabilities similar to human vision. They can also be made inexpensively as more functionality is integrated into a single chip.

The novel technology presented in this paper uses smart CMOS cameras in the corners of a display to look along the surface of the display and determine the location of an object in front of, or in contact with, the display. Typically, the object of interest is the user’s finger. The images collected by the cameras are processed in such a way as to recognize various attributes of the object(s), such as location relative to the display in three dimensional space (for one example). This information is then used in feedback to the computer generating the display, enabling touch control of the application.
Note that there can be many cameras in a system, but for single large displays, the number of cameras is usually limited to four (one in each corner). Other large displays can function with as few as two cameras or as many as are required to observe the display (a wall-sized system for example, needs multiple pairs of cameras to observe the entire surface). When very large displays are touch enabled, multi-user collaboration and the ability to detect pen or finger contact are desirable functions.

This paper describes a novel touch enabling technology for large format displays that facilitates human computer interaction in a very natural way. It will outline the architecture design and will discuss the several attributes of the technology over other touch technologies. Also the ability to work with a variety of display sizes, both large format and wall-size format, and accommodate multiple simultaneous users will be presented.

2. RELATED WORK

In the past several years, there has been a lot of research conducted relating to vision-based man-machine interfaces. Some work has focused on finger tracking while others have investigated hand gesture or hand posture recognition. Work on finger tracking is usually intended for the finger to replace the mouse, and, therefore, some mechanism for mouse-click is also required. The vision architecture varies greatly in both the type of hardware used and placement of the cameras. Virtually all systems use a traditional hardware approach, connecting a full-frame camera to a frame grabber and then processing the information on desktop computers. The complicated problem of feature detection has lead to widely varied approaches and restrictions on the system design. Researchers have used skin color, special gloves, restrictions on speed of movement, non-visible light, known static backgrounds, non real-time processing, as well as other specific constraints in order to create a demonstrable system. However, a robust commercial vision system for computer application control has not been available until the recent introduction of DViT™ technology by SMART Technologies Inc. which is presented in this paper.

There is a substantial volume of work related to the field of vision based human computer interaction and some interesting efforts are summarized here. A camera-projector system, the Digital Desk, is described in [1]. This system uses image differencing and motion to track the finger. The FingerMouse system [2] uses color segmentation to detect the finger and uses the keyboard to perform mouse clicks. FingerPaint was shown in [3] where the finger was detected by correlation. Mouse clicks were simulated by pressing the spacebar on a keyboard. Freeman, et. al.[4], investigate several vision algorithms to track hand gestures for gaming or TV control applications. Although temporal averaging was used, the best results were achieved by having a fixed background with a high contrast compared to the hand. Jennings [5] shows a 3 fps system using multiple cameras and color segmentation to track a finger against a complicated background in 3D space. The Magic Board system described by Crowley, et. al. [6] projects a computer images on a standard whiteboard while a video camera captures drawings at specific intervals. The camera is also used to track the finger to provide minimal editing capabilities (copy, paste, move, etc.). Finger identification is accomplished by color segmentation and correlation. Hardenberg and Berard [7], describe a system for finger and hand posture tracking. They used a color camera 384x288 pixels to collect images at a rate of 20–25 fps and a 1 GHz Pentium processor for image processing. The system was tested under different lighting conditions with both a constant and cluttered background. Segmentation of the finger worked best with a white background and movement that was less than 2.5 m/s. The “Visual Panel” [8] uses a single camera to track a finger against a piece of blank paper in 3D space. Mouse click is non real-time and is accomplished by having the user hold the finger stationary for a predetermined times (1 second is suggested in the paper). The Visual Screen [9] is similar to the Visual Panel except that the camera tracks the finger against the complicated background computer screen. Segmentation is accomplished by color filtering the predominantly blue cast background generated by the CRT monitor. The “Magic Table” [10] is an evolution of the Magic Board [6] and uses cameras looking down on a table surface. This system allows for more than one user. Besides finger or hand tracking, the system makes use of color to track tokens placed on the table. The Perceptive Workbench project by Starner, et. al., [11] also uses the table surface concept to track hand gestures or finger movements from multiple users. Objects placed on the surface are also identified. This system uses infrared illumination and a monochrome camera with an IR filter over the lens. Images from a 320x240 pixel camera are processed by two SGI workstations and a frame rate of 14-20 fps is achieved. Stenger et. al. [12] use a single camera to detect and track hand and articulated finger movements against a cluttered background. The tree-based methods result in ½ fps on a 1GHz Pentium computer. Letessier and Berard [13] use a camera to capture the hand and then use imaging difference segmentation to identify the location of fingers in order to manipulate digital objects on large surfaces. Rekimoto and Matsushita [14] demonstrate the use of an
infrared camera looking at a large transparent to track hand movements in their *HoloWall* project. More recently, Wilson [15] reported on his *Touchlight* project that uses two video cameras behind a semi-transparent screen to track hand movements on the other side of the screen. Although not strictly related to tracking a hand or finger over a display, the work of Shen, et. al., [16] and Sumi [17] are noted for their application of machine vision in room-size human computer interactions.

All of the previous work is relatively slow for smooth handwriting or fast double-clicking and none has been commercialized. This is partly due to the research intent of many of the efforts but also due to the real technological challenge of defining a system that is robust enough for general application.

### 3. METHODOLOGY

The recent ability to embed significant processing power onto the same printed circuit board as a camera sensor, as opposed to the traditional approach of using a frame grabber and discrete computing system, has led to the terminology of “smart” cameras [18]. In fact, the eventual integration of processing power on the same silicon as the camera sensor will lead to further leading-edge embedded applications [19]. Smart cameras are ideal candidates for industrial or consumer machine vision systems as they offer greater flexibility in mechanical design, much better bandwidth handling, and are low cost. Some aspects of hardware are described in [20] but more detail is given here.

Figure 1 is a graphical depiction of the DViT™ system architecture. The camera and master components are manufactured on single circuit boards which make low cost visual based input devices possible.

The smart cameras use a 2185M digital signal processor (DSP) from Analog devices to collect and process information from a National Semiconductors CMOS camera sensor. The pixel collection is performed by having the sensor clock pixels into a FIFO buffer and then that information is collected by the DSP when a full sensor frame is available. The camera DSP then processes the pixel information and extracts relevant metrics from the analyzed scene. A new current design used the more advanced Analog Devices Blackfin™ processor which interfaces directly to the sensor and does not require the FIFO.

A smart camera is mounted in each corner of the display and is configured to use a 640x20 pixel format (although other formats can be used). The field of view of the lens is 90 degrees. Lens distortion is compensated in the image processing algorithms.

This format, along with the embedded processor design, is a critical component in bandwidth handling as only a subset of the available pixels are collected and processed. This was the primary reason for selecting a CMOS sensor over a CCD sensor since the CCD sensor would require the collection of every pixel and thereby adding to the processor load and also limiting the frame rate to 24–30 frames per second (fps). This frame rate is not high enough for a writing application where 70–100 points per second (pps) of mouse events are desired. By reading out the pixels only in the 640x20 window of interest, very high frame rates can be achieved. Indeed, it is possible to get many hundreds of fps in this configuration. It should be noted that high frame rates are required not only for smooth writing but also for small system lag times. MacKenzie and Ware [21] have shown that user performance is degraded by the perception of increasing lag in system responsiveness. They show that the speed and accuracy of mouse movement is not seriously degraded for latencies up to 75 ms. They conclude that a latency of between 25 ms and 75 ms is optimal. Ware and Balakrishnan [22] showed that an acceptable latency for selection and manipulation of tasks is between 75 ms and 100 ms. In [7] and [11] where the system tasks are similar to those reported in this paper, the authors conclude that a system latency of 50 ms could not be perceived in human computer interaction. This 50 ms latency corresponds to 20 fps. Clearly, at 100 fps and a corresponding 100 pps output events, writing will be smooth and there will be no human perception of system lag.
Once the onboard DSP has processed the pixel information and extracted the needed metrics, that information is then sent to a Master DSP. The metrics are usually information about the object of interest in the camera scene. This object is normally a finger or pointer of some type. Some important metrics are location in the scene, contact, size, type, etc. The Master DSP uses the information from each smart camera to triangulate the object’s location in three dimensional space. This information is then sent to a computer running an application. Contact events (where the user has actually touched the display) are reported as mouse-clicks. In this way, the application is controlled.

The system architecture is very bandwidth friendly as there is a staged reduction from the camera to the computer. A camera running at 100 fps (to give 100 pps mouse events), will transmit 640x20x100 pixels per second. This requires a data flow in excess of 10 Mbits per second (Mbps) for each camera. The onboard DSP processes all this information right at the camera and reduces it to a few simple metrics about the pointer. This information only requires approximately 130 Kbits per second (Kbps) for all cameras. The Master DSP then further processes the metric information and reports the XY display location and contact status to the computer over a 9600 bit per second serial connection. The bandwidth handling is very efficient as it reduces 40 Mbps down to less than 100 Kbps (a data flow reduction of more than 400).
Figure 2. The DV iT™ overlay on a plasma screen

Figure 2 shows an example of what a finished commercial system looks like on a plasma display. This system uses DV iT™ technology. The arrows point to the smart camera location enclosed in the bezel. The tray along the bottom holds several passive pen styluses for writing and an eraser.

4. IMPLEMENTATION RESULTS

In this section, results relating to image processing and usage are discussed. Although there are many stages involved in image processing, two particular areas, segmentation and object identification will be described. In the domain of usage, machine vision offers many advantages. These advantages result in the fact the every pixel of a pointer is tracked and therefore more information besides the XY location is available. Also, the size limitation of more traditional technologies is overcome thus leading to very large touch displays [23]. Finally, results relating to two interesting usage issues, namely multiple pointers (multiple users) and large displays, will be discussed.

4.1. Image Processing

As stated in the previous section, all the image processing is done exclusively at the camera. Besides the fact that bandwidth considerations are dealt with, this configuration also increases the available processing power by assigning an individual processor to each camera as opposed to the frame-grabber approach which typically sends the camera image to a single host computer.

4.1.1. Segmentation

A good segmentation algorithm can be based on numerous approaches; motion detection by optical flow (for example), change detection by differencing against a reference image, region growing from a known pixel, edge detection, pattern recognition, and the list goes on. Successful segmentation can also be accomplished by taking a systems approach. For example, the use of a known fixed background can make segmentation by differencing and thresholding easier to manage. This improves robustness and significantly reduces the required processing power, and therefore reduces the
cost of the system. In real-time, manufacturable systems, robustness and cost are important considerations. In the following discussion, two fixed-background methods of segmentation are described.

Image differencing in natural light (or ambient light) is usually accomplished with a reference image followed by statistical thresholding. The thresholding algorithm can convert the image into three types of regions: areas that became brighter, areas that remained the same, and those that became darker. Objects that are lighter or darker are candidates for consideration as possible targets. How good the implementation of the thresholding algorithm, or how much apriori information is known about the background, will determine the amount of further analysis required to isolate the legitimate target.

A major weakness of image differencing in ambient light tends to be variable lighting conditions; that is, changes in the lighting can be incorrectly detected as changes in the scene. To combat this, the algorithm can continuously update the background image, as described in [24], based on the following equation:

$$B_{n+1}(i, j) = (1-a) B_n(i, j) + a I(i, j)$$

where $B_{n+1}$ is the new background image, $B_n$ is the current background image, $I$ is the most recent image, and $i$ and $j$ are the row and column pixel coordinates of the pixel being updated. The number $a$, which lies between 0 and 1, indicates how much should be learned from image $I$. As an example of the image processing, Figure 3 presents processed (upper) and ambient light (lower) images. For illustrative purposes these images are an expanded portion of the 640x20 format discussed in section 3. The pointer (e.g., a fingertip, as shown in this case) is extracted from the other features in the camera image. For illustrative purposes, the processed image depicts horizontal and vertical lines that indicate the XY image position of the pointer tip. From this processed result, only the two metrics relating to the tip position are important for determining pointer location in a touch system. As described in section 3, 640x20 (12,800) bytes of information have been reduced to two bytes of information.

Another approach to segmentation involves the use of artificial light. If the entire bezel (the four sides) surrounding the display were illuminated, then any pointer approaching the display would block the light reaching the cameras. The resulting occlusion would allow the image processing algorithm to easily extract the pointer location through simple thresholding. Of course, visible light might be distracting to the user so infrared (IR) light would be a good choice because the user cannot see it but the cameras can. The system described here has been implemented using sparsely spaced IR diodes around the perimeter of the display. IR light is then spread out through a diffusion layer in order to create a relatively smooth light barrier to the cameras. Figure 4 depicts the image, as seen by a camera, of an IR lighted bezel with a pointer object occluding a portion of the light. The top portion shows the unobstructed lighted bezel and the bottom portion shows the occluding effect of a pointer crossing the light barrier.
4.1.2. Object Identification

An important capability of machine vision is the ability to recognize different objects. In a touch-sensitive display system, the ability to recognize the difference between pen and finger tips (or any other object) represents a significant enhancement from the user's point of view. Any recognizable object (in terms of the system's knowledge) could then be used to control or direct applications. For example, in an electronic whiteboard application, a user would like to be able to write electronic notes but would also like to switch to or between other applications. If the difference between a passive stylus (a pen for writing notes) and a finger is detected, then the stylus can be exclusively assigned to generating electronic ink while the finger can be assigned to mouse events.

Previous work [23] demonstrates that this can be done. This is quite different from non-camera based touch systems that can only report the location of contact (i.e., the XY coordinate). This extended ability will lead to further exploitation as researchers seek to understand how to improve the interface between man and machine.

4.2. Usage

Usage is an important consideration when constructing an interface between man and machine. If the interface is perceived as difficult or unnatural, the user will not use the system. This is partly why finger capable touch interfaces are so attractive as pointing and touching are natural activities of humans. The ability to point and touch a computer icon and have the machine respond in the expected way is a natural and powerful paradigm.

Multiple pointer capability is increasingly important as touch interfaces are increasing in size to human scale proportions. Indeed, as we approach a time where physical walls become the interface between man and machine, it is expected that more than one user will want to interface with the machine since the physical space allows for that option. Machine vision has the ability to support many pointers simultaneously. As opposed to traditional touch interfaces that can only report a single XY location, machine vision “sees” the entire scene and all objects in that scene. Therefore, pointer objects in the scene can be correctly identified and acted upon. To further illustrate the concept, Figure 5 shows a camera image that has been processed and then binarized. This image clearly depicts two simultaneous pointers although more than two are possible as the number of multiple pointers detected is only limited by camera resolution and processing power.

Figure 5: A processed and binarized image showing two simultaneous pointers.

Figure 6 shows a large wall display that was also manufactured for studies being conducted a Georgia Tech. The system uses 12 cameras to provide a very robust touch system and to cover the 17-foot span while still providing excellent functionality (writing, click and drag, double-click, etc.). In the figure, three projected images are seamed together to make one large homogeneous display. Conversely, in the Georgia Tech research, the displayed information is constructed from a six-projector array that overlaps so that the user shadow is significantly reduced [25]. Large format high resolution displays are becoming increasingly important in scientific and educational institutions, and projector arrays are a common way of constructing them [26]. The compelling visuals, information density and high resolution make them ideal for a variety of applications. Until now the display could only be manipulated through traditional input methods and so some applications, like those that use hand-writing for example, were awkward and ineffective. By adding touch capability, new applications and uses are possible. In fact, it is now possible to replace a traditional classroom blackboard with a fully electronic, fully interactive, and more informationally dense wall-size equivalent.
4.2.1. Human Scale Man-Machine Input Device

Machine vision can be used to create a touch interface of arbitrary size due to the fact that the resolution of the system is optically based. This means that for any projected display, there is a proportional correspondence between the display pixels and the camera pixels. Therefore, if the resolution of the cameras system allows a touch resolution of say, one display pixel, that touch resolution will not change no matter how small or how large the display pixel is. In other words, whether an SXGA display occupies a 15"LCD or is displayed on a 50' wall, the capabilities of the machine vision system remain the same as long as the camera resolution is sufficient for pointer identification. This optical scaling ability is an important feature when making large format displays touch sensitive, but it also allows for creation small display touch systems. Figures 6 and 7 show two types of large-format touch displays. In the previously discussed figure 6, two users are shown simultaneously using a 17-foot-wide display. This is presently the largest touch interactive display available. Figure 7 shows an 84” camera-based touch sensitive display that can be used for presentation, training or any type of computer application control. The upper corners of the display are not covered thereby revealing the location of the cameras.

With these systems it is possible to control the computer by touching (clicking) the icons, dragging information, writing text, or anything else that can be accomplished with a mouse. The touch enabling technology is unique as it allows for human-scale operation. Everyone is familiar with the human-scale operation of chalkboards and whiteboards, and their functionality is well known and natural to use. However, they are not machines that can run sophisticated applications or record user activity. By paralleling the paradigm of familiar non-machine input devices, a modern implementation of the interface with comfortable usage has been created.
Figure 7: An 84” diagonal camera-based man-machine input device.

5. CONCLUSIONS

This paper describes an enabling technology for a smart camera-based man-machine input device for large-format displays. The touch enabling technology copies the familiar paradigm of the traditional chalkboard or whiteboard, but allows for the user to naturally drive computer applications by using their finger or some other passive pointing device. The use of machine vision has been shown to have many attractive properties that conventional touch interfaces do not have. The attributes include multiple users, object recognition and scalability to support human-scale operation. The significant technological challenges of bandwidth control, processing power, and image processing were solved to a sufficient level to allow commercialization for general applications and environments.

ACKNOWLEDGEMENTS

The author gratefully acknowledges the support and encouragement of David Martin (CEO) and Vaughn Keenan (VP Product Development) of SMART Technologies Inc. for providing a research and development environment that allowed for a working system to be constructed. The author is also appreciative of the many individuals who assisted in mechanical design, hardware design, programming and fabrication. Specifically, the assistance of Dr. David Holmgren, Randy McCharles, Scott Su, Radek Kristof, Trevor Akitt, and David Popovich, all of SMART Technologies Inc., is noted for their significant contribution in research, algorithm development, construction and deployment. The author also wishes to thank the reviewers for their insightful comments.
REFERENCES

Overview

• Motivation
• Technology
  – 3D-approach
  – 2D-approach
• Applications
  – ButterflyNet
  – ModelCraft
  – Shared Design Space / INTOI
Motivation

Background

A digital world with 1000 interfaces
Experimental Results

- **Finger-operated touch screen**
  - Best in speed and worst in accuracy (Albert, 1982)

- **Stylus(Pen)-operated touch screen**
  - Comparable to a mouse on both speed and accuracy measures (Mack & Lang, 1989)
Finger-operated touch screens

• Pros:
  – No special hardware requirements
  – Really intuitive (especially for novices)
  – Fast & Direct Input
  – Finger is usable, any pen is usable

• Cons:
  – The user’s finger may obscure parts of the screen
  – The screen gets dirty from finger prints
  – Less precise without pen

Touch-Interaction (Fluid DTMouse)

Touch + Pen != Pen + Touch

• Finger-operated touch screens often support pen input (e.g. SmartBoard)
• Pen-operated touch screens mostly do not support finger touch

Technology

3D approach
Pen-based 3d interfaces

- Tracked Wand
  - Reflecting balls
  - ART-tracking
  - Lower accuracy
  - Fraunhofer IGD

- Wireless Pen
  - Reflecting balls
  - IMS, TU Vienna

Interacting in 3d space

- Hardware that allows the user to communicate with the system
- Input device vs. interaction technique (e.g. zoom)
- Video

(used with permission of Hannes Kaufmann)
Pen-based Tablets

- Wacom tablets
  - Very precise input
  - Absolute values in 2d coordinate space
  - Direct touch on display
- MAX 6DOF stylus from Terminal Display System
Digital Pen, Scrivo.1

- Optical navigation & mouse-hover technology, 800 dpi
- No special surface requirements (it does not work on glass surfaces)
- BT-communication

Digital Pen, ANOTO

- Captures position (x, y) in absolute coordinates, time (t), pressure (p), and status (pen up, down)

www.anoto.com
Anoto Digital Pen - 2

- What the pen sees?
- Pen camera use IR light
  - Pattern has to be printed using IR absorbing ink
  - User content should be printed with IR transparent ink

Anoto Digital Pen - 3

- C, M, Y are IR transparent
- Black content has to be printed as C+M+Y, not K
Anoto Digital Pen - 4

- Transferring data to the computer
  - Manual transfer via dock
  - Automatic transfer via Bluetooth
  - The pen provides both page ID and the pen ID

Combining Advantages

Paper Notebook: Robust, Battery, ...

Computers: Search, Storage
Different Types of pen

- All-in-one pen: Leapfrog, FlyPen
- BT-based pen: Logitech io2 with BT Maxell, DP201 Nokia, SU-27W
- Non-streaming: Logitech io2 Maxell, DP201 Nokia, SU-1B

Pens for large surfaces

- SmartBoard allows both touch and pen interaction (optical-based)
- Ultrasonic-based tracking setups (e.g. MIMIO, eBeam)
- Digital pens?
Applications

ButterflyNet

(used with permission of Scott Klemmer, Stanford University)
Automatic Association
Notes + Photos associated by Time
Back at the Lab...

Multimedia Spreadsheet
Applications

ModelCraft

ModelCraft

• SolidWorks PlugIn
• Digital Pen and Paper models
  – Logitech IO2
  – Anoto pattern on the surface of the models
ModelCraft - 2

- Using External Reference
  - Fitting a cube around a door frame
Applications

INTOI, Digital Flipchart

Shared Design Space - Motivation
Real vs. Digital Paper

Real Paper
- Real Ink + Digital Ink
- Tracking of paper
  - ARToolKit (Kato, 2001), ARTag (Fiala, 2005)

Digital Paper
- Stylus tip
- Digital Ink

Shared Design Space

- 8 pens on a single BT dongle at 50 Hz

- Large table sizes are no problem (accuracy is not depending on the size) – 3 to 4 projectors mounted on the ceiling

- Occlusion & shadow problem

- Hand interaction
Rear-projection setup

• Experiment 1: Transparent foil
  – Good tracking, problems with image

• Experiment 2: Lee filter
  – White diffusion (used for spot-lights)
  – Good tracking, bad image

• Experiment 3: Backlit foil
  – great diffusion of projected image
  – Perfect tracking

INTOI, Interchanging Ideas

www.officeoftomorrow.org
INTOI – Feedback

- HP Colorlucent Backlit UV foil
- Protecting acrylic glass (<4mm)
- Features:
  - Multi-User Interaction
  - Simultaneous interaction
  - Scalable
  - Combination of touch and pen-interaction

Real and digital data

- **Pick-and-move**
  Pick data from a printed document and move it to the interactive surface.

- **Paper device**
  The paper as an alternative control device

- **Sketch-and-send**
  Draw & store sketches and send it to the table/wall display during a presentation
Thanks

• Peter Brandl, Michael Hurnaus, Daniel Leithinger, Jakob Leitner, Verena Lugmayr, Jürgen Oberngruber, Claudia Oster, Christian Schafleitner, Thomas Seifried, Jürgen Zauner

• François Guimbretière (University of Maryland), Hannes Kaufmann (IMS – TU Vienna), Scott Klemmer (Stanford University)

Questions

Michael Haller
Upper Austria University of Applied Sciences
Hagenberg/Austria

e-mail: haller@fh-hagenberg.at
web: http://www.officeoftomorrow.org
References


References - 2


ButterflyNet: A Mobile Capture and Access System for Field Biology Research

Ron B. Yeh\textsuperscript{1}, Chunyuan Liao\textsuperscript{2}, Scott R. Klemmer\textsuperscript{1}, François Guimbretière\textsuperscript{2}, Brian Lee\textsuperscript{1}, Boyko Kakaradov\textsuperscript{1}, Jeannie Stamberger\textsuperscript{3}, Andreas Paepcke\textsuperscript{1}

\textsuperscript{1} Stanford University HCI Group
Computer Science Department
Stanford, CA 94305-9035, USA
\{ronyeh, srk, balee, boyko, paepcke\}@cs.stanford.edu

\textsuperscript{2} Human-Computer Interaction Lab
Department of Computer Science,
University of Maryland,
College Park, MD 20742, USA
\{liaomay, francois\}@cs.umd.edu

\textsuperscript{3} Stanford University
Biological Sciences
Stanford, CA 94305
jeans@stanford.edu

ABSTRACT

Through a study of field biology practices, we observed that biology fieldwork generates a wealth of heterogeneous information, requiring substantial labor to coordinate and distill. To manage this data, biologists leverage a diverse set of tools, organizing their effort in paper notebooks. These observations motivated ButterflyNet, a mobile capture and access system that integrates paper notes with digital photographs captured during field research. Through ButterflyNet, the activity of leafing through a notebook expands to browsing all associated digital photos. ButterflyNet also facilitates the transfer of captured content to spreadsheets, enabling biologists to share their work. A first-use study with 14 biologists found this system to offer rich data capture and transformation, in a manner felicitous with current practice.

Author Keywords

Mobile capture and access, augmented paper notebook.

ACM Classification Keywords

H.5.1: Multimedia Information Systems — artificial, augmented, and virtual realities. H.5.2: User Interfaces—input devices and strategies; interaction styles; prototyping.

INTRODUCTION

Every day, we witness mobile professionals at work—on the subway, at the park, in cafés. On mobile phones, they chat with business partners and write text messages. On their laptop computers, they surf the Web and post blog entries. Yet, despite the availability of these tools, many professionals rely on paper notebooks. To understand why this is, consider the advantages of each medium. Computers afford interactive computation, electronic communication, multimedia, and digital information management. Paper notebooks, on the other hand, are cheap, turn on instantly, have infinite battery life, and provide a fluid and flexible surface for jotting down ideas on the go. They are also amazingly robust. As a result, paper notebooks support many mobile practices better than computing devices do.

Field biologists struggle daily with this tradeoff. On the one hand, their practice depends on paper notebooks as the central organizing tool, considering its shortcomings necessary to gain the reliability and flexibility of paper (see Figure 1). On the other hand, field biologists depend on computers to analyze data, and must transform their work to do so. This tension suggests a wholesale replacement of paper in current practices. However, a wholesale replacement of paper can be problematic, as evidenced by Sellen and Harper’s work [33]. Instead, we argue that it is better to design technologies that complement paper tools: the bits in our computers should be aware of the atoms of our world (see e.g., [13, 14, 24]). Next generation tools should support the capture of heterogeneous data, aid the transformation process, and yet preserve the best aspects of current paper-centric practices.

Figure 1. A) Field biologists choose paper notebooks because they are portable, readable outdoors, robust to harsh field conditions, and have infinite “battery life.” B) As seen on this office desk, paper notebooks support flexible input and output. C) However, like Tracy Storer’s notes from 1925 (in CAS archives), most notes are locked in storage, their value lost to those who might benefit from them.
The rest of the paper is organized as follows. The next section summarizes our observational study of field biologists. Following that, we present the two primary contributions of the paper. The first is ButterflyNet, a system comprising interaction techniques—informed by the observational study—that leverages digitally augmented paper notebooks as the central structuring tool for capturing, organizing (through automatic and manual techniques), transforming, and sharing heterogeneous data. The second contribution is a first-use study of this system, and the lessons we learned. The study demonstrated that automatic association was highly successful, and that manual associations show promise for some users.

IN THE WILD WITH FIELD BIOLOGISTS
Part of our interest in studying field biologists stems from a desire to use an understanding of this highly mobile community to inform mobile interaction design. Designing from a deep understanding of a particular community can provide insights valuable in a broader context.

This study comprised several parts. First, the first author interviewed 23 biologists from Stanford University, the Jasper Ridge Biological Preserve (JRBP), and the California Academy of Sciences (CAS). He conducted each interview at the biologist’s work place, observing current practices in the field and in the lab. Second, he joined a field research class in the Los Tuxtlas rainforest, where he lived with 12 biologists, helping with experiments on tropical plants. Third, he became a docent at Jasper Ridge, where he has spearheaded a project to evaluate digital camera traps. In total, this study comprises 370 hours of observing, talking to, and working with field biologists, with observations captured on photographs, audio, and video. Finally, an ecologist (the seventh author) collaborated on this research. She aided our need-finding efforts and directed us toward issues most critical for biologists. From this work, we have distilled design implications that can influence future mobile tools. We summarize these implications here.

Capture and Access of Heterogeneous Data
Field biologists capture handwritten notes, digital photos, audio, video, sensor readings, GPS data, and physical specimens. By examining how these are currently managed, we make a case that new technologies must support the rich capture and access of this heterogeneous data.

Paper notebooks are a field biologist’s central organizing tool (see Figure 1). They take their notebooks everywhere, using them as the definitive record of all procedures, measurements, and results. In the field, biologists use notebooks to capture observations that may lead to new hypotheses. This practice, shaped by Joseph Grinnell’s work [27], emphasizes careful documentation with descriptions of the day’s work, the time and date, weather, participants’ names, and pictorial annotations such as maps. We examined 13 notebooks from five biologists (471 total pages), finding that notebooks primarily contain tabular data and descriptive prose, augmented with charts, pictures, sketches, pasted-in-sheets, and bulleted lists.

Field biologists supplement their notes with specimens, photos, GPS data, audio and video. Physical specimens help biologists understand ecosystems. For example, CAS owns millions of specimens. Field biologists use photos and video to record experimental data, observations, and context to supplement their notes and specimens. One use of photos is to identify species where collecting specimens is not desired. For example, some of the biologists we work with use cameras to “trap” mammals at Jasper Ridge. The biologists use the photos to identify animals, in an effort to model their movement. Photographs also aid collaboration, as they can convey the feeling of an ecosystem to other scientists. Biologists can also use photos to locate sites in locations where GPS data is not available, such as under a rainforest canopy (or as backup in case GPS data is lost). When GPS is available, many biologists use commercial receivers to capture the geographic data. One of our interviewees uses GPS to track the spread of invasive ant species. And as for audio, one ornithologist we spoke with captures bird calls while conducting his research in India. He correlates his notes with the audio of the calls, and sends ones he cannot identify to a local expert for help.

Finally, field biologists use sensors to record environmental parameters (e.g., temperature, solar radiation, wind speed, humidity, and precipitation). Portable, inexpensive, low power, and reliable sensors such as the iButton [12] have enabled environmental data collection in harsh situations, and the advent of battery-powered wireless sensor networks [9] offers even richer environmental monitoring. While sensor data can be exported to PCs, current tools cannot associate these data with a biologist’s own observations, making the understanding of natural systems fractured.

In short, field biologists gather information from a diverse set of sources, yet have little support for coordinating and distilling this information. Transforming the information into analyzable forms is labor intensive and error prone, as the information may be scattered across different locations. There is limited support for organizing, searching, and sharing. Moreover, there is no tractable method for ascertaining a particular result’s data lineage. And while scientists struggle with these tasks, valuable research remains trapped in paper notebooks and in digital storage.

Technology makes it possible to overcapture in the field; however, as we found, solutions for rapidly harnessing this rich data are limited. Improving this situation can have a significant impact. Technology that supports mobile capture and access should strive to meet several design goals. First, it should support handwritten notes and the other types of data that field biologists work with, such as specimens and digital photos. Second, it must support the robustness requirements of the domain. Finally, the design must
remain flexible, enabling biologists to include new input streams as needed.

Data Transformation and Tools Integration
While much of a biologist’s research is organized on paper, interpretation requires that data be entered into computers. We learned during our interviews that a big limitation of current practice is that transcribing data from paper notebooks to spreadsheets is painfully slow. Interviewees asked for OCR software to import handwritten tables into Excel. One interviewee described his bee experiments in Costa Rica, where he and collaborators spent six hours a night transcribing datasheets. The ornithologist who worked in India spent multiple 12-hour sessions correlating audio with his notes, transcribing the information into a database. New technologies need to support efficient transformation of data from the captured format (e.g., handwriting) to the computer world. While fully automated solutions are tempting, they will not work in all cases. Current solutions are error prone, and the process of manually transforming some data plays a cognitive role in helping the biologist assimilate her research. The design goal, then, is to provide a hybrid solution, where the biologist can oversee the computer transformation of data. One such design is where a person manually verifies handwriting recognition results.

In addition, systems in this area must also integrate with downstream tools, to enhance usability and increase adoption. For writing publications, our interviewees use Microsoft Word. For statistics, they used Excel, SAS, JMP, or SPSS. For capturing geography metadata, they use GPS receivers in the field and GIS software at the field station.

Robustness
Paper notebooks can take extraordinary amounts of abuse before failing. Data can be salvaged from a notebook that is torn in half, dropped to the ground, or subjected to a downpour. The same cannot be said about modern portable computers. Field systems should follow suit by being robust and offer graceful degradation.

THE BUTTERFLYNET SYSTEM
Informed by this study, we designed ButterflyNet, a capture and access system for notebook-centric mobile work. With ButterflyNet, field scientists can capture, organize, and share heterogeneous research data, including notes, photos, and specimens (see Figure 2). By recognizing the centrality of paper notebooks in current practice, ButterflyNet allows users to be immediately familiar with its primary interactions. This section describes these interactions and how they support field biology work practices.

Heterogeneous Capture
ButterflyNet supports the capture of handwritten notes, digital photographs, and physical specimens. To capture handwritten notes, a field biologist uses the Anoto digital pen system [2] (we use Nokia SU-1B pens with Bluetooth [30]). While ink is physically laid down on paper, the pen’s camera tracks a dot pattern printed on that paper and digitally captures which page and where on the page the writing occurs; it even annotates every stroke with the current time and date. When the user synchronizes the pen with a PC, the digitized notes are uploaded. We decided on Anoto pens because they afford graceful degradation. Unlike pure digital solutions, if the pen’s digitizer were to fail, users would still be able to record observations, as the paper and inking pen provide redundancy. Conversely, if a physical notebook is lost or otherwise unavailable, the electronic version can be used.

To capture photographs, a user employs a digital camera. For richer interactions, we prototyped a custom “smart” camera (see Figure 3A), our functionality prototype of a successor to contemporary digital cameras. With the smart camera, users can perform on-the-spot annotations of photos by marking on the LCD screen with a stylus. The smart camera also communicates wirelessly with the pen, offering real-time visual and audio feedback for in-the-field interactions. This smart camera was prototyped with an OQO handheld [31] running Windows XP with a webcam affixed to the back. This is a functionality prototype; we presume that a production implementation would provide a sleeker form factor. (Given current technology trends, we anticipate this will be a camera phone.)

To capture physical specimens, biologists use tagged coin envelopes (see Figure 4D). Using coin envelopes to collect specimens was a practice observed in our field work. The tags enable ButterflyNet to uniquely identify specimens.

Information Association
ButterflyNet provides several techniques to associate captured data. Association between heterogeneous data is important as it “glues” together pieces of data, possibly scattered among various media, into a meaningful story about the field work. Our field study found that systems...
must provide both low-threshold and high-ceiling interactions [28] — easing adoption for novices while providing control to experts.

The first technique is automatic time-based correlation, an extremely low-threshold technique that does not require biologists to alter current practices. Photos, notes, and other data that contain timestamps are automatically associated by ButterflyNet during capture. For example, if a biologist writes an observation at 3:23 PM and takes a photo shortly thereafter, the photo and those notes would be associated.

ButterflyNet provides two manual techniques to provide more precise, explicit control over media association. Explicit authoring is important, as a biologist may take many photos before batch-processing them, a use model that automatic time correlation does not support.

One technique, hotspot association, enables users to associate a photo with a specific area of a notebook page (see Figure 3). To invoke a hotspot association, the user captures a photo (or browses to a photo) and then draws two brackets in her notebook. This hotspot is later visualized as a photograph that has been resized to fit into the frame. Our smart camera provides real-time multimedia feedback for hotspots; it beeps and displays a temporary popup to confirm that the hotspot association has been created. The audio feedback is an important design feature, as in the field, users may not actually be looking at the camera while creating the hotspot. The hotspot interaction extends prior work in smart-paper systems [11, 19, 23] by enabling end users to author associations on-the-fly.

The second technique, visual specimen tagging, enables users to associate physical specimens with photos and handwritten annotations (see Figure 4). The user places the desired specimen in a coin envelope enhanced with a 2D barcode and Anoto paper. Annotations written on the paper will be associated with the barcode, and thus, the specimen. Additionally, any photo containing this barcode will also be associated with the specimen. When taking a photograph that is related to a particular specimen, the user places the envelope such that the barcode appears in the photo. ButterflyNet detects the tag in the image, extracts the ID, and establishes the association. This technique aligns well with field biologists’ existing practice of using envelopes to store specimens and other physical artifacts.

Rich Information Access
In addition to the capture and association techniques presented above, ButterflyNet supports rich information access through the ButterflyNet Browser (see Figure 5). After the biologist imports her data, she can use the browser to visualize her notes and photographs in a rich browsing interface. The content panel (Figure 5B) shows the information the user is currently focused on (digitized field notes by default). The photo context panel (Figure 5C) shows time-associated photos. For example, if a user views notes from 3:23 PM on March 23, 2005, she will see photographs taken on or near that time in the context panel.

The browser provides a direct manipulation interface for
Figure 5. The ButterflyNet Browser. A) A timeline visualization of captured notes. The browser presents digitized field notes in the main panel (B), and associated media in the context panel (C). Maps (D) and sensor data (E) were not used in the study.

navigating the data. The timeline visualization (Figure 5A) allows users to jump to content by date and time. The height of each bar represents the quantity of data at that time interval. Users can jump to specific pages with the navigation bar (Figure 5F), or show multiple pages by zooming out (via a slider on the navigation bar). The bar also lists shared notebooks, which the user can view by selecting from a dropdown menu.

ButterflyNet also enables users to access research data using their physical notebook. With this technique, a user taps the page with his digital pen, and the ButterflyNet Browser responds by presenting the digital version of that page and all associated data. With this technique, a user can also retrieve hotspot-associated photographs by tapping inside a hotspot frame (in the physical notebook) with the digital pen. The retrieved photograph appears in the browser or on the smart camera.

Enhancing Data Transformation and Integration
Finally, ButterflyNet enhances data transformation through a multimedia spreadsheet, which contributes several novel organization and visualization techniques (see Figure 6). First, the spreadsheet assists with transcription of tabular data. Users can select handwritten data in the browser and send it to a window that hovers over the spreadsheet. As the biologist types, a placeholder moves down the page to help her keep track of which row she is currently transcribing, eliminating the need to look back and forth between a physical notebook and the computer display.

The spreadsheet enables users to embed photographs and charts into individual cells. (In Excel, these objects cannot be placed in a cell; they float loose.) This feature is accessed through a context menu that is updated with new content as the browser views new pages. Like the smart camera, the spreadsheet is a prototype of the salient aspects of a future system. Currently, it serves as a ButterflyNet-integrated springboard that can export to industry standards.

To further facilitate transformation and sharing, the user can select any data in the browser, and export to the system clipboard. The physical notebooks can also be used to export data to the spreadsheet. When the spreadsheet is open, a user can draw a pair of hotspot-like brackets on a page to specify a region of interest. ButterflyNet detects the paper gesture, extracts the selection from the corresponding digital notes, and exports it to the multimedia spreadsheet.

Extensibility
ButterflyNet was architected with extensibility in mind. We are currently extending the system to associate and present a wider variety of data, including audio, video, GPS logs (Figure 5D), and sensor data (Figure 5E). If notes are georeferenced, a map will show relevant locations. If there are sensor readings that were logged at the same time as captured notes, they will also appear in the context panel.

We will also continue to explore the tangible navigation of media. With the smart camera, a biologist can now retrieve associated photos by tapping the digital pen to a relevant notebook page. This device ensemble approach for in-the-field retrieval is valuable in mobile settings, where screen real estate is intrinsically limited for individual devices.

Implementation

SYSTEM EVALUATION
We conducted a first-use study of ButterflyNet, focusing on interactions with three data types (photos, notes, and specimens) and three hypotheses:
H1 The field capture techniques (digital notebook, hotspot association, and visual specimen tagging) enable media association with minimal overhead.

H2 The ButterflyNet Browser presents a fast and rich information view by presenting photographs both in a context panel and inline with notes (through hotspots).

H3 The spreadsheet facilitates the transformation of data.

Sessions were held at the Jasper Ridge Biological Preserve, and lasted 2.5 hours per participant (we paid $45 cash). The 14 participants (six male; eight female) included JRBP docents, PhD students in biology, and professional researchers. Field experience ranged from none (for a single docent), to 1-2 years (most docents), to several years (for PhD students), up to 18 years (for one professional). Five of the 14 had more than 10 years of field research experience.

We asked participants to go to the field to collect photos, notes, and specimens, and then use that day’s data to create a spreadsheet to present to colleagues. The design of this task was informed by our field study. Specifically, we modeled the task to mimic a day of field research, as we witnessed in the Los Tuxtlas rainforest. The first hour of the study comprised fieldwork, where the participant carried a backpack (with water and equipment), a field notebook and digital pen, a digital camera (Canon SD300), the smart camera, and tagged specimen envelopes. The reason that participants carried two cameras was that at the time of the study, hotspot association required the smart camera’s features, while the consumer digital camera’s higher resolution yielded more reliable recognition of the specimen tags. We envision that future cameras will provide the smart camera functionality.

In the field, biologists used three techniques:

1. For each oak gall they found in a 2m × 40m line transect (a standard field sampling method), they recorded the distance of the gall along the transect, its size (large, medium, small), its color (dark, bright), and a photo.

2. At three points along the transect, they photographed the habitat using the smart camera, and associated it with a hotspot in their notebook (see Figure 7).

3. At three different points on the transect, they used a visual specimen tag to photograph, annotate, and collect a physical specimen of their choice.

These subtasks mirror everyday field tasks — collecting measurements, photos, and specimens. Back at the field station, the participant filled out a 15-question survey of their background and their opinions on the field task.

Next, the participant engaged in a lab task (also informed by our need finding). The participant was asked to use the browser and spreadsheet to create a spreadsheet with photos and measurements, for explaining the data to collaborators. As an incentive, we awarded the author of the “most useful spreadsheet” a $10 gift certificate (the winner was chosen after all studies were completed). The lab task ended with a second 15-question survey to gauge the lab tools and ButterflyNet in general. Finally, we conducted an informal debriefing interview with the participant. Other than this interview, and a tutorial of ButterflyNet, the participants completed the tasks on their own, while the experimenters observed (capturing video and handwritten notes).

RESULTS

This section highlights several outcomes of the user study, discussing how they will impact our future work. We organize the results around ButterflyNet’s key features, and refer to specific observations, questionnaire results, coded free-form responses, and hypotheses where appropriate.

Media Association

Participants readily understood the automatic, time-based association. However, at the time of the study, ButterflyNet associated media at a fixed (and coarse) granularity—the context panel showed all photos captured within the time span of the current page. Unfortunately, the users recorded many measurements and photos per page, and sometimes needed finer associations than ButterflyNet provided (e.g., there might be several photos of different galls all at 1:24 PM). To negotiate the spreadsheet task, then, some participants would find anchor images in the browser (e.g., the dark and small outlier gall at 1:24 PM), and then interpolate the rest (e.g., the subsequent photos at 1:24PM must be associated with the measurements immediately after the small and dark gall). Thus, we see that capture and access systems need to provide the user a way to adjust and visualize the granularity of automatic associations.

Participants were excited about the possibilities presented by hotspot association (the two-bracket gesture for associating photos to parts of a page). People mastered the gesture quickly. One participant found it efficient enough to draw in every row of measurements, to achieve a one-to-one association with photos. During the debrief interview, a different participant mentioned that ButterflyNet, with its hotspot-based and time-based association techniques, would
Figure 8. Participants found automatic association most applicable to their current work. Hotspot linking shows promise, and visual specimen tagging may suit only some biologists. Study participants were also able to quickly learn and apply the visual specimen tagging technique. However, we did notice that occasionally, the visual tag would not be recognized, due to tall grasses occluding parts of the barcode. Fortunately, in this case, the biologist would still be able make the association after the fact, as the visual tag includes a human-readable number (see Figure 4D).

Figure 8 shows the participant response to the association technique questions (we use the median to analyze the ordinal data). These results partially support H1. Participants felt that automatic association would not increase field time, and were positive toward the technique's potential usefulness. Through automatic association, ButterflyNet presents an informal UI, such that the in-the-field focus—when time is expensive—is on documentation, rather than interface manipulation. And though flexibility over the window of automatic association would improve the experience, the system was already performing better than today's jury rigged solutions.

However, the data show that participants felt that hotspot association and specimen tagging slightly increased field time, and felt that specimen tagging would have to improve before they would use it in their own work. This response to visual specimen tagging may have several explanations. First, biologists may be reluctant to use tools that increase field time by any amount. Second, not all of our subjects collect specimens in their work, and thus have no use for the tagged envelopes. Finally, it may be due to limitations in our current implementation—we currently do not provide functionality beyond linking tagged photos with annotations and do not provide solutions for the occasional barcode recognition problem (e.g., by manually presenting the barcode again in a more controlled environment).

As our study implementation addressed photo-based tasks, our data analysis partitioned the participants by how much they value photos in their work. In this case, opinions about the association techniques diverged significantly. In all cases, the likelihood that the subject would use the technique ranked higher for those who valued photos, showing that participants who use photos are particularly excited about ButterflyNet’s potential. For instance, when asked if she would use ButterflyNet’s field tools into her work, a veteran of more than 10 years responded with straight 7’s (the highest rating). She takes 10-20 photos per day, and views photos as extremely important (7 out of 7). She stated that she found the ability to find photos and associate them with spreadsheet cells perfect for her work with animal teeth. Recently, she requested a copy of ButterflyNet to use in her current work measuring jaw bones (through photos).

Rich Information Access
Participants readily understood the ButterflyNet Browser’s presentation and access interface. In our questionnaire, we attempted to determine the usefulness of this interface. Figure 9 summarizes the evaluations for 14 likelihood variables in a 1 to 7 scale (1 for very unlikely; 7 for very likely). The top advantages participants saw were that ButterflyNet would help them to capture and transcribe more data. Additionally, it would help them recall experiments better. These lend support to H2, that the browser provides rich information access.

Transformation and Integration
Participants successfully completed the lab task, and generally perceived that the transcription helper would speed up transcription. We find that the tools integrate well with current practice (12 of 14 reported regular spreadsheet use, the highest rating in a 7-point scale). In the free-form responses, eight mentioned that they liked the association of photos with notes. Six liked the tools for exporting data. Six wrote that digital backup for notes would be invaluable. Thus, ButterflyNet aided transformation (H3) and integrated well with the participants’ current practices (see Figure 9).

Graceful Degradation
The study also reflected how ButterflyNet supports graceful degradation. Very occasionally, the digital pen would miss a few letters or numbers in participants’ handwriting. Perhaps there was dirt on the page, or perhaps the pen was used too close to the edge of a page (where the pen’s camera cannot decode the dot pattern). The few users who encountered missing data in their digitized notes quickly
switched to their physical notebook, where the data was faithfully captured with actual ink. These participants seemed comfortable defaulting to the physical notebook when the digital representation was incomplete.

**Gesture Recognition**

The recognition of hotspots is good. The recall rate was 78.3% (54 of 69 attempted were correctly recognized); the precision rate was 88.5% (of the 61 recognized, 7 were false positives). However, many errors arose from a single participant’s data, whose hotspots were smaller than our threshold. Without this data, the recall rate was 93.3% (42 of 45), and the precision rate was 91.3% (4 of 46).

Our fieldwork found that participants would rather save field time, even if it resulted in more work later. Thus, we made a design tradeoff to make the hotspot gesture as lightweight as possible. The normal PapierCraft gesture engine requires a pigtail loop at the end of all gestures to enhance recognition. We removed this to achieve simpler gestures. Additionally, because ButterflyNet does not have modes to switch between gesture and ink, all strokes are potential hotspots. This design achieves simpler field interaction at the expense of recognition rates. However, given the existence of recognition errors, future iterations will enable users to edit associations between potential hotspots and their photos (e.g., by deleting false hotspots).

**Possible Limitations**

The freeform questionnaire feedback pointed to possible limitations. First, participants felt that while the aided transcription was faster, it was still tedious. To address this, we are currently integrating handwriting recognition into ButterflyNet and exploring the UI implications. Second, a few participants voiced concern about the need to use a special pen, and were worried they might lose it in the field.

The data indicate a slight negative correlation between expertise and opinions, though not all expert participants currently use photos. For example, one expert who gave low ratings studies bat calls and takes zero pictures per day. When we described in debriefing that future versions would handle audio, he said that then, ButterflyNet would prove extremely valuable to him.

The data shows that experts who use photos find the pen and notebook interaction useful. The manual techniques did not fare as well; we note that they must prove valuable beyond automatic association. Additionally, participants only had limited exposure to them in the lab.

Much of the support for our hypotheses comes from questionnaire results. While the ratings generally support ButterflyNet’s lightweight interactions (H1), fast and rich information view (H2), and efficient transformation of data (H3), one must keep in mind that each session took no more than 2.5 hours, and that a longitudinal evaluation would be much more reliable. We leave this for future work.

**Future Work**

The results from this study point toward some exciting opportunities. An important step will be to study how biologists can use ButterflyNet to interact with data outside of photos and notes. The studied system did not include any GPS or sensor data features. The freeform responses did show that while participants found the integration of photos to be useful, many stated that adding GPS integration would prove extremely helpful. We plan to integrate GPS, sensor data, audio, and video into future versions of ButterflyNet. One particular point of interest is automatic correlation based on other metadata facets, such as location.

Also, while the hotspot interaction currently works only for cameras, there is no reason why it cannot be generalized. As long as a device can record the timestamp of captured or browsed-to data, it can leverage hotspots. Thus, in the future, a field biologist may be able to associate video, GPS, or sensor data using simple hotspot gestures.

**RELATED WORK**

This research draws from prior work in three areas: interacting with paper, information capture and access, and information technology for biologists. In this section, we explain how this work contributed to our system’s design.

**Interacting with Paper**

Two systems in particular inspired much of our early ideas. Mackay’s a-book integrates a paper notebook with a PDA for laboratory biologists [25]. The “interaction lens” enables users to create a table of contents, links between pages, and links to external sources. A-book demonstrated the importance of scientists’ current artifacts and practices, and introduced techniques for augmenting notebooks. Our fieldwork results corroborate many of Mackay’s findings, that notebooks are multimedia documents, and that the notebook is the central tool for supporting the design and execution of biology experiments. The second system, Audio Notebook, introduced a paper notebook where tapping portions of a written page retrieved the audio recorded when those notes were written [34]. The elegance of imbuing a paper notebook with query capabilities was one of the main inspirations of ButterflyNet. ButterflyNet differs from these systems by providing richer capture of heterogeneous media, an efficient visualization interface, and higher-ceiling interactions for associating media.

Prior work has shown that people are comfortable using physical paper interfaces to control media. Listen Reader is an augmented paper book that allows a user to control audio streams by moving his hands near different parts of a page [5]. Books with Voices introduced paper transcripts as a physical input medium for browsing video [19]. Parikh’s work marries camera phones with the affordances of paper [32]. Users transcribe data with the phone’s keypad, and invoke computation by photographing visual codes on paper forms. Each of these systems offers a technique for
associating paper and a single digital medium. ButterflyNet builds upon these ideas, and contributes techniques to navigate heterogeneous media.

Paper PDA [4, 14], XAX [17], and PADD [13] demonstrated techniques for manipulating documents in either digital or physical form. PapierCraft [22] investigated gesture-based commands for interactive paper. NISMap [8] showed how pen-and-paper interfaces can provide robustness in field situations. Other systems (e.g., [11, 23, 29]) have explored techniques for leveraging the tangibility of paper in multimedia navigation. ButterflyNet’s association and navigation techniques are inspired by this class of work. Like some of these systems, ButterflyNet takes advantage of the Anoto digital pen system. Additionally, ButterflyNet uses PapierCraft to recognize hotspot gestures.

**Information Capture and Access**

A central research theme of ubiquitous computing has been techniques for capturing and accessing information [1]. FiloChat [37] provides such techniques for personal and shared notes. Like Audio Notebook, FiloChat showed how synchronizing notes with audio could improve later review. PARC has worked on a number of tools for capture and access of group meetings (e.g., [26]). One system, Tivoli, enabled users to revisit meetings through a time slider or by pointing to virtual pen strokes. eClass showed that capture and access can be effective in classroom lectures, where the professor’s actions are captured and made accessible to students over the web [36]. In particular, the StuPad extension enabled students to visualize lecture notes within their own personal notes. Integration of free-form ink with other streams of input (audio, video, etc.) has been a common theme in these systems (see e.g., [18, 21]).

Like AudioNotebook, ButterflyNet leverages paper as the central media. However, it extends the capture and access ideas to provide synchronization with photographs and physical artifacts. Like INCA [35], a toolkit for capture and access systems, ButterflyNet provides an infrastructure for time synchronization of data. We are extending it to also provide association by other metadata facets (e.g., location).

**Information Technology for Biologists**

While scientific data has been traditionally organized around paper notebooks, the advent of new technologies (such as tablet computing) offers important benefits for field biology. As a result, there has been recent interest in electronic systems specifically supporting biology research (e.g., [3, 7, 10]). Tablet-based electronic notebooks can work well in laboratories, where power outlets are plentiful, and an infrastructure is available to provide electronic communication and backup. PDA-based solutions are suitable for data capture in the field, but still trail behind paper in flexibility and robustness. Thus, in the field, paper notebooks still remain the medium of choice.

ButterflyNet extends the ideas for the digital lab out into the world, enabling biologists to take work between the office, the lab, and the field. What ButterflyNet contributes is a hybrid physical/digital solution for field scientists, and an information ecology approach for organizing heterogeneous data types that treats each type as a first class citizen.

**CONCLUSIONS**

In summary, we have contributed to the mobile design space in several ways. First, we described a study of one group of mobile workers—field biologists—showing how they construct a heterogeneous tool belt, featuring paper notebooks. Second, we detailed the interaction techniques in ButterflyNet, a system informed by these field observations. ButterflyNet provides several capture and structure techniques, and a device ensemble metaphor for accessing the captured information. As a result, it expands the process of leafing through a notebook into a process of browsing synchronously created media. Finally, we presented results from a first-use study.

We have released ButterflyNet as open source software (see http://hci.stanford.edu/bio), and are currently working on longitudinal evaluation with biologists and other scientists. We are exploring uses of this platform for rich interactions with visualizing research data on maps, and for richer collaboration. We plan to expand the system to include all information explicitly captured and implicitly available in field sites, including sensor data, audio, video, and GPS. We expect future iterations to lend impetus to design in the broader mobile domain.

While the domain of field biology provided the frame for this work, we expect that the research contributions will apply to mobile workers in general. For example, we are currently studying how ButterflyNet can aid mobile workers such as designers, anthropologists, archaeologists, and medical practitioners, all of whom rely on paper notes. We plan to study ButterflyNet over the long term with users in these communities.

With a phone in every pocket and a PC on every desk, the next decade promises sweeping transformations in the way we interact on the move and in the world. But for all the attention paid to these technologies, we often overlook the unassuming yet equally ubiquitous technology of the paper notebook. The ButterflyNet system, with the implications we presented in this paper, brings us closer to a future where physical and digital tools work together as one.

**ACKNOWLEDGMENTS**

We thank NSF (IIS-0430448 partially supported Yeh; IIS-0447703 partially supported Guimbretière and Liao) and Microsoft Research (for partially supporting Liao). Nokia, Intel, and HP provided technology donations. We are indebted to the biologists in our studies, and Rodolfo Dirzo, Thaala Montsi, Terry Winograd, and Hector Garcia-Molina for their insight. We thank Nona Chiariello and Philippe...
Cohen for helping to recruit subjects and provide space for the study. Finally, thanks to the reviewers for their feedback. Human subject research conducted for this paper is covered under Stanford University IRB approved protocol 82471.

REFERENCES
10 CyberTracker, CyberTracker. http://www.cybertracker.co.za
12 Embedded Data Systems, LLC., iButton. http://embeddeddatasytems.com
ModelCraft: Capturing Freehand Annotations and Edits on Physical 3D Models

Hyunyoung Song, François Guimbretière, Chang Hu
Human-Computer Interaction Lab
Department of Computer Science,
University of Maryland,
College Park, MD 20742, U.S.A
{hsong, francois, changhu}@cs.umd.edu

Hod Lipson
216 Upson Hall
Cornell University
Ithaca, NY 14852-7501, USA
hod.lipson@cornell.edu

ABSTRACT
With the availability of affordable new desktop fabrication techniques such as 3D printing and laser cutting, physical models are used increasingly often during the architectural and industrial design cycle. Models can easily be annotated to capture comments, edits and other forms of feedback. Unfortunately, these annotations remain in the physical world and cannot be easily transferred back to the digital world. Here we present a simple solution to this problem based on a tracking pattern printed on the surface of each model. Our solution is inexpensive, requires no tracking infrastructure or per object calibration, and can be used in the field without a computer nearby. It lets users not only capture annotations, but also edit the model using a simple yet versatile command system. Once captured, annotations and edits are merged into the original CAD models. There they can be easily edited or further refined. We present the design of a SolidWorks plug-in implementing this concept, and report initial feedback from potential users using our prototype. We also present how this prototype could be extended seamlessly to a fully functional system using current 3D printing technology.

ACM CLASSIFICATION: H5.2 [Information interfaces and presentation]: User Interfaces. - Graphical user interfaces.

GENERAL TERMS: Design, Human Factors

Keywords: Pen based interactions, Tangible interactions, Rapid prototyping.

INTRODUCTION
In the process of designing artifacts, today’s designers alternate between tangible, non-digital media such as paper or physical 3D models and intangible, digital media such as CAD models. An architect might start the design of a new building with sketches on paper, then, when her ideas solidify, create a rough model using cardboard, before finally creating the corresponding digital model. Once this model is finalized, it might be fabricated as a 3D object (either through rapid prototyping techniques or a modeling studio) so that her clients may have a better grasp of her vision. While a fully digital design process has long been advocated, it still seems a distant goal because tangible, non-digital media models present unique affordances often difficult to reproduce in digital media. Architectural models for example offer a unique presence that is difficult to reproduce on a screen. As a result, even projects that rely heavily on computer assisted design techniques (such as the recent Hearst building designed by Sir Foster) still employ
tangible models both for aesthetic and structural tasks \cite{9, 10} (Figure 2, Top).

Interacting with models is an intrinsic part of the design process for architects who see construction (and sometimes deconstruction) as a fundamental part of the idea forming process. For example, during the early phase of the design process called “massing” - a brainstorming practice that iterates between incremental modifications and rebuilding of models for conceptualization purposes - inexpensive, easy-to-build models are used to better understand the shape requirement of a building. As rapid prototyping technology has become more commonplace, models are now employed in other areas of design as well. Mechanical designers use models to check form and functional compatibility with the context of an object’s use. Models can also be extensively annotated (Figure 2 Bottom). Unfortunately, information captured on such models is difficult to integrate back into the digital world.

While this problem could be addressed by a conventional tracking system (either magnetic or optical) as proposed by Agrawala et al. \cite{2}, that approach is limited to a relatively small working volume. Magnetic or optical tracking systems require infrastructure and calibration on a per-model basis, and are somewhat expensive. They are also difficult to deploy in the field where models are frequently tested. This limits widespread adoption by architects or designers.

Noting that most annotations take place on the surface of the model, we present a system which uses the inexpensive, off-the-shelf Logitech io\textsuperscript{TM} digital pen \cite{20}. This pen is equipped with a built-in camera which captures position information by observing a digital pattern \cite{4} printed on each model (Figure 1). Our system can capture not only annotations but also editing commands that are subsequently applied to the original digital models. Using our command system, and auxiliary tool such as a ruler, users can alter and adjust the shape of a model, such as modifying dimensions, filleting corners, creating holes, or extruding portions of a model based on requirements in the field. The information is naturally captured in the frame of reference of the model, without the need to worry about scale or orientation. Our approach does not have a predefined working volume, and can easily scale in terms of the number of objects tracked, number of pens used, and locations of usage. Furthermore, it does not require a per-model calibration. Because our approach advocates cohabitation of tangible and digital models, it integrates seamlessly with the current usage patterns among architects and mechanical designers for whom interacting with 3D models is a fundamental part of their creative process. By capturing annotations and edits on physical 3D models, our system streamlines the design process and simplifies documentation of the design history of a given project.

In this paper, we present the first prototype of such a system developed as a plug-in for SolidWorks \cite{31}, a commercial CAD application. Starting from a model inside SolidWorks, users can print and build simple, paper-based 3D models (or create water transfers to be applied on an existing 3D model). They can then use a digital pen to annotate them or draw gestures that will be executed upon pen synchronization. After describing the system architecture as well as our editing system, we report on our experiences while designing this system. We also report initial feedback gathered from potential users, professional architects and teachers. Finally, we explore in detail possible paths for the implementation of such a system using current 3D prototyping technology.

**PREVIOUS WORK**

Several systems allow users to draw (or paint) on digital models. Hanrahan and Haeberli \cite{8} described a WYSIWYG system to paint on 3D models using a standard workstation. This approach has also been adapted to annotate CAD drawings \cite{17, 30}. While drawing on a virtual object has many advantages, such as the ability to work at any scale, we believe that physical models will always play an important role in the design process because of their appeal to designers (Figure 2). In that respect, our approach is closely related to Agrawala et al.’s \cite{2} 3D painting system. Our approach extends this work in several ways: By using a tracking system based on an optical pattern printed on the surfaces of the object, we offer a very short setup time requiring no calibration on a per-object basis. Our tracking...
approach also provides greater flexibility for users as annotations can be captured at any location. Finally, our approach is inherently scalable, both in terms of number of models and in terms of annotating devices—a property difficult to achieve by either optical or magnetic tracking techniques. Using a different approach, Grasset et al. [6] proposed to use augmented reality techniques to annotate objects directly. On the one hand, by relying on passive props, our system is less powerful than such systems as it does not offer direct feedback. On the other hand, the simplicity of our system makes its cost of use very low (no need to wear or set up any equipment)—a key aspect for acceptance by designers and architects.

We believe that future systems should allow users to interact directly with the representation of a given object that is most convenient for the task at hand—be it a digital model on a screen or a 3D printout of that model, or a combination of the field sketching proposed here with augmented reality feedback. In that respect, our work is similar in spirit to the work by Guimbretiere [7] on digital annotations of document printouts.

Our work is also related to the large body of work on 3D sketching in systems like Sketch [37], Teddy [14], SketchUp [1] and the 3D Journal project [23]. Our system complements these systems by addressing the need to capture modifications sketched directly on the models at later stages of the design process. In particular, our system makes it easy for users to capture real world geometric information. Our command system is also quite different. While the systems mentioned above focus on a gesture-based interface, we adopt a syntax-based approach inspired by recent work on Tablet-PC-based interfaces such as Scriboli [13], Fluid Inking [36] and paper-based interfaces such as PapierCraft [18]. We believe that this approach allows for a more flexible and extensive command set while retaining a sketch-like style.

Our work is also closely related to tangible interfaces [12, 16, 32-34] which let users interact with digital information through the use of tangible artifacts. All these systems leverage users’ familiarity with spatial interactions to allow them to perform complex interactions with ease. Our system extends and complements these systems by offering a tighter correspondence between the tangible proxy and its digital representation. In doing so, we offer users the opportunity to modify the digital representation in the real world. In that respect, our system is also closely related to the Illuminating Clay system [26] and Liu’s work on editing digital models using physical material [19], as they allow users to see modifications made in the real world applied to the equivalent 3D model. Sheng’s thesis on modeling shapes using fingers and physical props [29] is also related to our approach, but Sheng’s work focuses on free form shapes such as shaping clay. All these system require the use of somewhat complex tracking equipment only available in a lab setting, while our approach is very light-weight.

**MODELCRAFT IN ACTION**

While our vision is to have a traceable pattern generated automatically while 3D-printing a model, our current prototype uses simple paper models instead. While building a paper model seems arduous, interviews with architects confirmed that they often build model out of paper (Figure 3), sometimes starting from a printout of an unfolded CAD model. Hence our approach augments current practice.

Each SolidWorks model is printed as an unfolded paper cutout on a page of paper that has been pre-printed with a unique Anoto pattern [4]. This pattern provides a very large space of uniquely identifiable pages (in excess of $2^{48}$ letter sized pages). Importantly, this makes it possible to interact with different objects or different printouts of the same object at once. Practical paper models are usually quite simple since early designs often rely on a vocabulary of basic shapes (cube, cylinder, pyramid, cone, sphere) as proposed by D. K. Ching [5]. We show a typical example in Figure 3. Moreover, complex shapes are currently supported by printing the object with a 3D printer and printing a slide transfer to be applied to each face.

All interactions are carried out with the Logitech io™ pen [20], a commercial implementation of the Anoto system. As each Anoto digital pen has a unique ID, it is also possible to distinguish several different pens interacting on one object. Our system also lets us designate special objects as tools. For example, we instrumented one of our rulers by taping a strip of Anoto pattern onto it (Figure 6). Making marks on this ruler is interpreted by the system as making measurements during command operation.

**Annotations**

Annotating a model is straightforward: Simply pick up the model and annotate directly on any surface (Figure 4a). Upon pen synchronization, the marks will be merged onto the corresponding surface of SolidWorks model. Users can use several pens for different colors. Marks created by annotation pens are not interpreted by the system.
Command Syntax. All commands are performed with a command pen which lays ink in a different color (red in our system). We choose a “command” pen approach as it fits well with the current practice of using color coded annotations. Other solutions such as having a command button on the pen are also possible.

All commands follow a uniform syntax (Figure 5) inspired by Scriboli [13] and PapierCraft [18]. First, users draw on objects or on tools (like our ruler) a set of strokes that represent the parameters of the action to be performed (Figure 5a). Then they draw a pigtail gesture which is used as a separator between the parameter strokes and the command name, (Figure 5b). Next, they write the name of the command they wish to execute (a simple letter in our current implementation) on top of the pigtail (Figure 5c). During pen synchronization, the command is then executed using the area on which the pigtail started as the primary command parameter. For example, to create a hole through an object (Figure 5), the user would draw the shape of the hole onto the object surface, then draw a pigtail starting inside the shape, and then write a C (for cut) on top of the pigtail. Note that starting the pigtail outside of the shape would have created a pillar instead.

The use of the pigtail proved to be very reliable for pen-based interaction [13] and is well-adapted to our case as it does not require any feedback besides the ink laid on the surface [18]. For our system, the pigtail has two advantages. First, it serves as a natural callout mark when one needs to execute a command on a small area (like cutting a hole for a screw). Under such conditions, it would be difficult to write the name of the command directly on the area of interest because the area is too small or too close to the surface border. Second, the pigtail provides a natural orientation for the surface. While up and down are well understood in a Tablet-PC context, this is not the case on 3D objects which people may place in arbitrary orientations to facilitate the
annotation process. Accordingly, when interpreting a command, we consider the pigtail as the baseline for the command name (Figure 5d).

As shown in Figure 4, some operations may require several sets of strokes. For example, to create a groove on an object (Figure 4d) one first draws the profile of the groove on one surface, then the extent of the groove on an adjacent surface, then one uses a pigtail to indicate the inner region, and finally one writes a $G$ (for “groove”) on top of the pigtail. Another example is the creation of a cut of a given depth. To do so, the user first creates the shape of the cut, then marks the depth of the cut on another face, and then uses a pigtail to issue the cut command. As shown in Figure 6, top, this syntax makes it very easy to use real world objects as references without the need for further measurements.

So far, we have considered cut operations, as they are easily specifiable by drawing on the available sides of the models. Our system also provides a way to create extrusions through the use of our “digital” ruler. For example, to extrude a shape from a surface, one simply draws the shape on the surface, then draws a mark on the ruler to indicate the extrusion length, and finally, using the pigtail, issues the extrude command (E) on the area to extrude. Figure 4e provides an example of such an operation. As in the case of cutouts, this command facilitates the process of using real-world objects as references (Figure 6 bottom).

Together the Cut, Groove and Extrude operations represent the fundamental modification tools (removing material and adding material) defined by D. K. Ching [5], these are also the interactions most frequently described by architects. But of course our system is easily extendable to more complex commands as we will describe later.

Dealing with Errors in Batch Processing. In our system, the annotations and commands are captured in batch mode. There are several reasons for this choice. First, as explained previously, it is important for the intended pattern of use of our system that interactions can take place away from a computer. Second, by delaying execution, a batch approach might help keeping users in the “flow” of their task by avoiding unnecessary interruptions.

Of course, errors will need to be corrected eventually. Our interface offers two main mechanisms to deal with errors. For marking errors in annotations and commands we use a simple scratch out gesture to indicate that the underlying gestures should be removed, or that the underlying command should not be performed. For execution errors, it is important to remember that while our system might misrecognize gestures and command names, it accurately captures the parameters of the commands on the correct faces. Since this information is directly transferred to SolidWorks, it becomes a trivial matter to make corrections because all the relevant command parameters are already in place which saves transcription time.

In our system it is also possible to issue several “alternative” commands by simply drawing the new command over the last command, a common pattern in practice. Each command will be recognized as a different operator (or “feature” in SolidWorks terminology) and appear in the feature tree managed by SolidWorks. Once the strokes have been transferred to SolidWorks, the user can compare the results of different commands, pick the best of them, and delete...
alternative executions. Also, to document the design process, each of the alternative commands could be applied in a different configuration using the configuration management provided by SolidWorks.

**IMPLEMENTATION**

As shown in Figure 7, the life cycle of a model in our prototype can be broken down into 3 phases: 1) print the 3D model as a paper prototype with a unique pattern on each side; 2) capture the strokes made on the paper prototype and map the strokes onto the correct virtual 3D model’s face; 3) execute the commands themselves. We are now considering each phase in turn.

**Printing the Model**

To print the paper model or water slide paper patch, we start from a SolidWorks file and unfold it as described below. The unfolded model is then printed on paper pre-printed with the Anoto pattern. During printing, we use the PADD infrastructure [7] to maintain the relationship between a given model and the unique page ID on which it has been printed, and to record the calibration data and the geometric transformation used during printing. This information is used during the synchronization process, to identify on which digital model a stroke has been made. Note that we are using the Anoto system as it is the only commercially available tracking system at this time, but other systems such Data Glyphs [11] could have been used as well.

**Unfolding the Model.** While many unfolding approaches are possible, such as [24] (see [27] for a review), we focused on limiting the number of discontinuities because they interfere with the tracking system. We use a heuristic approach which starts at the triangle with the longest perimeter, and transforms it onto a plane. Then it recursively visits all neighboring triangles and attempt to repeat the transformation, while maintaining adjacency of neighboring triangles and disallowing triangle overlap in the plane. Since the level of discontinuity is proportional to the length of seams, the recursion order follows a greedy heuristic of taking the next untransformed triangle sharing the longest boundary with transformed triangles that can be transformed without creating an overlap in the plane. When no such triangle is found, a new pattern is initiated and the process repeats. Once a set of patterns is generated, the patterns are collected and nested in a page by recursively packing pattern bounding boxes.

**Importing Back Captured Strokes**

During pen synchronization, our application receives all the strokes captured by a pen. Strokes are recorded with a time stamp and the page ID on which they were made. We use this ID to recover the model that the strokes were drawn on as well as the calibration and geometric transformation stored during the printing process. Once the user is ready to process the strokes, they are imported from our application onto the unfolded model; each stroke point is mapped from page coordinates back into 3D coordinates by applying the inverse of the transformation that was originally used to move the local triangle from 3D onto the plane.

**Executing Commands**

As pointed out above, all command strokes have been made by a special “command” pen, so it is easy for our system to distinguish them. Our first step in processing these strokes is to segment the stream of command strokes into individual commands. To do so, we first detect strokes that might look like a valid pigtail using a set of simple heuristics such as looking for gestures with a relatively small loop and large outside tails. Once these are detected, we observe if there is a stroke recognizable as a character that has been drawn above the candidate pigtail within a pre-set time out. If this is the case, the stroke is recognized as a valid pigtail, and the strokes drawn since the last command are used as parame-
ters for the command execution. We also check for natural command separators (such as creating an annotation) and check that the parameter set matches the command. For example, the face IDs associated with shape and pigtail delimiter and command character should all be the same. In practice, this approach worked well for our prototype. Once command syntax has been validated, the command is executed. Commands that use input from tools as parameters are processed in a similar way, but the strokes that were performed on the tool are processed in a tool-dependent way. If a command is not recognized, it is skipped, but its strokes are still presented on the surface of the model.

DISCUSSION
As we were developing our prototype we conducted several formative studies about the potential use of our approach. So far, we have conducted six semi-structured interviews, including a demonstration and a hands-on test. Our participant population covered a wide range of architectural backgrounds and included a student in an architecture school working as a drafter, several young architects, a senior architect, a senior partner and a faculty member at a school of architecture. Despite the current shortcomings of our prototype (such as the requirement that each contour parameter uses only one stroke, and the use of handwriting recognition without training) seasoned architects’ reaction to the system was very positive. Several architects pointed out that our system would be perfect for massing a building. During massing, new models are built based on marks or shapes that were suggested in the previous iterative cycle. This type of practice is well suited for the ModelCraft interactions.

The professor remarked that our system would allow students to explore prototyping and develop 3D thinking skills. For example, visualizing the 3D results of subtractive operations drawn on a face of a cube is a common task in architecture training. ModelCraft might also create a natural bridge between the traditional approach to architecture (based mostly on paper-based sketching) and the use of modern applications such as SketchUp [1]. Architects further pointed out that annotations on paper models could be useful for capturing feedback from some of their clients who might be intimidated by digital models. The response to the system was more muted for younger participants (one a student drafter and one a CAD modeler), since their work did not require extensive use of tangible 3D models. Yet, the architecture student pointed out that the system would be very useful for teaching and would support current practice taught at school. The CAD modeler, while skilled in building models, was not using them at work. This participant also pointed out that she often “deconstructed” her models in order to reconfigure them, so annotations did not seem as useful for her. We are considering ways to support this type of approach with our system.

Several users were concerned about the limitations of the digital pen (mainly that one has to remember to aim the pen correctly). One user suggested that this problem could be alleviated by slightly modifying the design of the pen. Overall, our interviews confirmed our hypothesis that a system bridging the gap between the digital and physical worlds would be useful for practitioners and teachers alike.

Editing the Models
The design of our command language followed a different path than that of Teddy and Sketch. While those systems adopted a gesture-based approach well-suited for sketching, we used a structured approach based on a simple extendable command structure and a pigtail as a separator between parameter strokes and command selection [13, 18]. One of the strengths of our approach is that, while keeping an informal feel, it can be easily extended to more complex commands and a wider set of commands by using longer command names. We implemented a Window (W) command to create windows of a certain depth in buildings (Figure 1). Two additional commands supported by SolidWorks were implemented as part of our system: first, (S)hell, a command that creates a shell given a volume, and second, a (F)illet command used to round out a selected edge. Using techniques described in the PapierCraft system [18], we could also transfer a shape captured on transfer paper onto a given surface and extrude it. It would also be easy to extend the system to accept post command parameters like numerical arguments.

Another important difference to other systems is that in our system, there is not always a plane on which to draw. This limitation is not merely the result of our tracking technology. Even if more complex tracking systems were used, it would still be difficult for people to draw in free space. This makes several techniques (such as free form extrusion) that were used by Teddy more difficult to implement in the present system. However, as discussed above, we were able to address this problem through the use of simple tools such as rulers. We are also exploring how users could use sketches drawn on a drawing board to create new geometry.

Tracking Performance and Limitations
One of our goals during this project was to better understand the limitations of a tracking method based on a pattern printed on the model surface. We now discuss our observations derived from working with our prototype.

Printing Models. Our current prototype used only printed, paper-based models. For simple models, this approach worked extremely well. Cutting and scoring (to simplify folding) the models by hand proved to be easy and accurate. Using a laser cutter would greatly simplify and increase the accuracy of this process. As shown Figure 4 this approach works well with basic shapes such as a cube, cone, cylinder, pyramid, and tetrahedron. Using a more advanced unfolding algorithm like the one proposed by Mitani and Suzuki [24] would allow for more complex shapes. Yet, it is clear that the paper-based approach seriously limits the complexity of objects. One simple alternative is to add the pattern to existing models once they have been built. For example, using our system, one can print the unfolded surface on a water slide transfer paper, and apply the transfer onto the model. Our tests showed that this technique is a viable option for models printed with a ZCorp printer. This approach allows
for more complex shapes to be built rapidly and only adds a small time to the production process. Yet, for the system to stay accurate, one needs to be careful while applying the transfer.

Of course the preferred solution would be to have the 3D printer print the pattern at the same time as the 3D object itself. Some 3D printers (ZCorp Z510) can already print at a resolution up to 600 dpi in the plane of the printing bed and 540 dpi vertically [35]. This is in the same range as for laser printers able to reproduce the Anoto pattern. Unfortunately, our tests showed that a pattern printed with the ZCorp Z510 printer was not recognized by the digital pen. To understand why, we show in Figure 9 segments of patterns printed on a laser printer and on a Z510. As can be seen on Figure 9, left, the dots produced by our laser printer are of somewhat irregular shape but use black ink to provide a highly contrasted image. The dots printed on the Z510 (Figure 9, right) are diffuse and do not use true black ink but a combination of C, M, and Y inks to simulate black. As a result, they are likely invisible to the infrared pen sensor. We believe that this problem can be readily addressed by introducing a truly CMYK printing process and using finer grained printing material. Another solution would be to use another tracking system like the Data Glyph designed for 300 dpi printing on par with the minimum layer thickness of .089 mm (286 layers per inch) of the ZCorp process, or a more robust encoding scheme.

Accuracy. The Anoto tracking system reports points with 678 dpi accuracy, but, taking into account the errors introduced by pen orientation and the printing process, the system’s maximum error is around 1 mm. Of course, the overall accuracy of the system also depends on the accuracy at which the paper is cut and folded (around 1 mm in our current manual process). Using a laser cutter would further improve accuracy.

Optical Tracking of Passive Patterns. Another problem inherent to optical tracking is that the system might lose tracking because the pen camera overhangs on a face or because users are trying to draw inside a groove or on an indented face. At overhangs, the pen loses track when the tip is about 3 mm from the border, at which point it vibrates. As a result, the smallest square surface on which a command can be issued is 12 mm wide. For indented faces the problem is exacerbated by the fact that the Anoto firmware is expecting a continuous pattern in the field of view. In our tests, the pen was able to track a pattern at the bottom of a 4.8 mm x 4.8 mm groove or mark a 6.4 mm diameter circle using a 1.6 mm thick template. Finally, because the pen was developed for tracking on flat surfaces, the system cannot track strokes on cylinders (or cones) whose radius of curvature is smaller than 12 mm. It is not clear how significant these limitations will be in practice and future work will be necessary to evaluate their impact. Other 3D encoding schemes may remediate this.

Another limitation of our tracking system is that it cannot track in free space. As demonstrated above, instrumentation of traditional tools used by wood workers (such as rulers, squares, tracing paper) may help to address this problem. For example, we used our instrumented ruler to indicate the height of an extrusion.

Finally, the current version of our digital pen does not provide orientation information for the object itself. So far, this limitation proved to be mainly relevant for handwriting recognition and our use of the pigtail as a reference mark addressed the problem successfully.

Character Recognition. Character recognition and pigtail recognition determine the total number of successfully recognized editing commands. Several problems might affect the recognition rate: first the pen provides samples at a relatively low temporal resolution which might influence the recognizer. To address this problem we add/subtract points so that the points are sampled not according to the time stamp but equidistantly. The orientation of the command might also have some effects. Our informal tests showed that using the pigtail as a baseline of character recognition was quite successful as the orientation of the characters seems to have little influence over the recognition rate. Finally, we observed that when users write on curved surfaces, letters are slightly deformed. This is due to users writing from a planar surface perspective. This problem can easily be addressed by projecting each letter on the plane normal to the surface at the centroid of the letter. Overall our tests show that our pigtail recognition rate is about 99% and, given our small dictionary of commands, we could reach a command recognition rate of about 92%. Further empirical evaluation will be needed to confirm these numbers.

FUTURE WORK
The system presented in this paper is built as an exploration tool allowing us to investigate the feasibility of our approach and provide us with a hands-on demonstration for potential users. In the near future we are planning to expand the system so that it can accommodate more complicated models and can be used during long term studies.

Dealing with Non-Developable Surfaces
Non-developable surface are problematic for our system because unfolding of such surfaces leads to multiple discontinuities in the pattern space (Figure 10) and creates gaps in tracking. Our tests suggest that the pen’s field of view is
often the case that architects create new designs by stacking objects creating many discontinuities in the pattern space. A closer look at the design of the Anoto pattern [21] reveals that this is merely a limitation of the current implementation. In principle, one could uniquely resolve a position if any 2.4 mm x 2.4 mm patch is visible. We believe that if the firmware were modified to detect the edge of each continuous pattern region (maybe by recognizing printed edges) and each face of the model was wider than 2.4 mm, the pen would be able to uniquely identify its position even around discontinuities in the pattern. Another solution to this problem would be to adopt a different approach to tracking altogether. Instead of mapping a 2D pattern onto our models, we could tile them with small (2-3 mm) optical tags which can be tracked by the pen. For example, one could use the system proposed by Sekendur [28], or the Data Glyph system [25], or, of course, the Anoto position pattern itself. All of these provide the large number of unique identifiers that is necessary. In all cases, the requirement of the minimum patch size can be accomplished using subdivision-based techniques such as the one used in the Skin system [22] and extended by Igarashi and Hughes [15].

We would also like to examine in more detail how our system could be adapted to 3D printing systems. In particular, we would like explore the feasibility of a 3D version of the Anoto pattern. This would not only simplify the printing process and alleviate the pattern discontinuity problem but also allow for annotations on newly exposed, cut, or fractured surfaces of objects.

Extended Feature Set
Our interviews with architects pointed to several directions in which the current system could be extended. One of them is to provide a better support for “free space” sketching by using information sketched on free paper to be incorporated as parameters to commands. Another one is to provide an operation to “glue” objects together. This multiple object operation will be very useful in early design phases as it is often the case that architects create new designs by stacking or joining available building blocks. This will provide a functionality similar with Anderson et al.’s system [3].

Finally, while our current system focuses on batch processing, new Anoto pens can transmit the strokes they capture in near real-time. One of the appeals of our system is that it can be used without a nearby computer. Nevertheless, several applications might benefit from streaming capabilities (for example by combining our system with the Urp system [34]). It will be a simple matter to adapt our system to streaming-based interactions.

With these new functionalities in place we intend to conduct longer term usability studies to better understand how our system will be accepted and how it might change current design practices.

CONCLUSION
We presented a new system which lets users capture annotations and editing commands on physical 3D models and transfer them onto the corresponding digital models. Our system is inexpensive and easily scalable in term of objects, pens, and interaction volume. Our command system reflects current practices of model builders and integrates seamlessly with current practice. Our system allows users to bridge the gap between the digital and the physical worlds by allowing them to deploy resources of both media for the task at hand. We believe that our approach will provide an efficient tool for the early phases of design in both architecture and product design.

ACKNOWLEDGEMENTS
This work was supported in by NSF Grant IIS-0447703 and Microsoft Research (as part of the Microsoft Center for Interaction Design and Visualization at the University of Maryland) and a graduate fellowship from the department of Computer Science at the University of Maryland. We would like to thank the architectural and interiors firm of BeeryRío for their support during the interview process (with special thanks to Rosana Keleher), Irena Savakova of DMJM H&N and all our participants. Corinna Löckenhoff and Adam Bender provided many useful comments to help improve this document. We would also like to thank Ben Bederson, Bobby Bhattacharjee and Bill Pugh for their support. Foster & Partners kindly provided us with the picture shown in Figure 2 top. ZCorp kindly provided us with the picture shown in Figure 2 bottom.

REFERENCES
1. @Last Software, SketchUp. 2005.
6. Grasset, R., L. Boissieux, J.D. Gascuel, and D. Schmalstieg. Interactive mediated reality. Proceedings of Pro-


29. Sheng, J., A Gestural 3D modeling Interface using Fingers and a Physical Prop Traceked in 3D, PhD thesis, University of Toronto. 2005


35. ZCorp, ZCorp 3D printing system. 2005.


Carpeno: Interfacing Remote Collaborative Virtual Environments with Table-Top Interaction

HOLGER REGENBRECHT

University of Otago, New Zealand
Information Science, P.O. Box 56, Dunedin
holger@infoscience.otago.ac.nz
Phone: ++64 3 479 8322
Fax: ++64 3 479 8311

MICHAEL HALLER

Upper Austria University of Applied Sciences, Austria

JOERG HAUBER

University of Canterbury, New Zealand

MARK BILLINGHURST

University of Canterbury, New Zealand
Abstract

Creativity is enhanced by communication and collaboration. Thus, the increasing number of distributed creative tasks requires better support from computer-mediated communication and collaborative tools. In this paper we introduce “Carpeno”, a new system for facilitating intuitive face-to-face and remote collaboration on creative tasks.

Normally the most popular and efficient way for people to collaborate is face-to-face, sitting around a table. Computer augmented surface environments, in particular interactive table-top environments, are increasingly used to support face-to-face meetings. They help co-located teams to develop new ideas by facilitating the presentation, manipulation, and exchange of shared digital documents displayed on the table-top surface. Users can see each other at the same time as the information they are talking about. In this way the task space and communication space can be brought together in a more natural and intuitive way. The discussion of digital content is redirected from a computer screen, back to a table that people can gather around.

In contrast, Collaborative Virtual Environments (CVE) are used to support remote collaboration. They frequently create familiar discussion scenarios for remote interlocutors by utilizing room metaphors. Here, virtual avatars and table metaphors are used, where the participants can get together and communicate with each other in a way that allows behaviour that is as close to face-to-face collaboration as possible.

The Carpeno system described here combines table-top interaction with a CVE to support intuitive face-to-face and remote collaboration. This allows for simultaneous co-located and remote collaboration around a common, interactive table.

Keywords

Collaborative work, CSCW, Virtual Environments, Tabletop Interfaces, Teleconferencing

Introduction

In recent years computing and communication has become tightly connected so it is easier than ever before for remote teams to work together. Despite this, current remote collaborative tools do not support the easy interchange of ideas that occur in a face to face brainstorming session. In this case people are able to use speech, gesture, gaze, interaction with real objects and other non-verbal cues to rapidly explore different ideas. In addition, there is a need to provide technology that can capture and enhance face to face meetings, such as digital whiteboards and interactive tables.
The central question that we are interested in exploring is: how can we create a computer supported environment which enhances face-to-face collaboration while at the same time allowing remote team members to work as closely together as if they were all sitting around a single real table.

A tool dedicated to group processes has to support the inherent requirements of a creative environment [1]:

□ The group members have to be able to communicate their ideas verbally and non-verbally, so they can build on top of each other’s ideas.

□ Group members need to be able to visualize ideas through use of sketching, image presentation and document sharing.

□ Group members need to be able to work with real world objects, including creating new or modify objects and showing examples to others.

The tool to be developed has to deal with three elements: creative people working in a creative space focusing on the creative task. Creative people are the target users, such as designers, and architects, who work in domains requiring original idea generation. The creative space is an environment which should be as close as possible to a face-to-face situation, which generally prove to be the most creative settings. Creative tasks are those where the goal is divergent rather than convergent thinking and where group result is supposed to be better than any individual outcome.

These requirements are challenging, however in this paper we present a prototype system that has many of the elements of an ideal interface for supporting face to face and remote collaboration. In the next section we review related work from earlier research in enhancing face to face collaboration and enabling remote collaboration. Then we describe two of our earlier prototype systems, cAR/PE! and Coeno, and our current integrated system, Carpeno, which uses elements from both of these prototypes. Finally we present an exploratory usability study which evaluates the Carpeno prototype and gives some directions for future research.

**Related Work**

*Enhancing Face-To-Face Collaboration*

Early attempts at computer enhanced face-to-face collaboration involved conference rooms in which each participant had their own networked desktop
computer that allowed them to send text or data to each other. However, these computer conference rooms were largely unsuccessful partly because of the lack of a common workspace [2].

An early improvement was using a video projector to provide a public display space. For example the Colab room at Xerox PARC [3] had an electronic whiteboard that any participant could use to display information to others. The importance of a central display for supporting face-to-face meetings has been recognized by the developers of large interactive commercial displays (such as the SMARTBoard DViT\(^1\)).

In normal face-to-face conversation, people are able to equally contribute and interact with each other and with objects in the real world. However with large shared displays it is difficult to have equal collaboration when only one of the users has the input device, or the software doesn’t support parallel input. In recent years Stewart et al. coined the term Single Display Groupware (SDG) to describe groupware systems which support multiple input channels coupled to a single display [4]. They have found that SDG systems eliminate conflict among users for input devices, enabling more work to be done in parallel by reducing turn-taking, and strengthening communication and collaboration.

In general, traditional desktop interface metaphors are less usable on large displays. For example, pull down menus may no longer be accessible, keyboard input may be difficult, and the mouse requires movement over large distances [5]. A greater problem is that traditional desktop input devices do not allow people to use free-hand gesture or object-based interaction as they normally would in face-to-face collaboration. Researchers such as Ishii and Ullmer [6] have explored the use of tangible object interfaces for tabletop collaboration while Streitz et al. [7] use natural gesture and object based interaction in their i-Land smart space. In both cases people find the interfaces easy to use and a natural extension of how they normally interact with the real world.

In many interfaces there is a shared projected display visible by all participants; however, collaborative spaces can also support private data viewing. In Rekimoto’s Augmented Surface interface [8], users are able to bring their own

\(^1\) http://www.smarttech.com/
laptop computers to a face-to-face meeting and drag data from their private
desktops onto a table or wall display area. They use an interaction technique
called hyper-dragging which allows the projected display to become an extension
of their own personal desktop. Hyper-dragging allows users to see the information
their partner is manipulating in the shared space, so it becomes an extension of the
normal non-verbal gestures used in face-to-face collaboration. In this way the task
space becomes a part of the personal space.

**Enabling Remote Collaboration**

Although being in one place and talking to another person face to face can be
considered the gold standard for collaboration, it is not always possible,
economical, or otherwise desirable for people to come together in the same
location. In that case they alternatively rely on teleconferencing systems that
support effective collaboration at a distance.

Many researchers from the fields of CSCW (Computer Supported Cooperative
Work), HCI (Human Computer Interaction) [9, 10] and Social Psychology [11]
have explored the complex issues around distant communication and remote
collaboration. They have tried to understand how systems for remote
collaboration should be designed to mediate human activities in a way that allows
people at a distance to accomplish tasks with the same efficiency and satisfaction
as if being co-located - ideally even going beyond that [12].

In that context, videoconferencing (VC) technology has always played and still
plays an increasingly important role as it provides a rich communication
environment that allows the real-time exchange of visual information including
facial expression and hand gestures. A growing number of organisations
nowadays use advanced video based collaboration-networks like for example the
AccessGrid\(^2\), or Halo\(^3\) system developed by HP for group-to-group meetings on a
daily basis. Although the installation and operation costs for these systems seem
high, they still prove effective at supporting tasks over a distance, thus making
travel redundant. However, although systems like these are capable of producing
videos with high grade audio and image quality, a remote encounter for people in
front of the cameras often feels rather formal and artificial. The spontaneity and

---

\(^2\) [http://www.accessgrid.org/](http://www.accessgrid.org/)

natural interaction that we take for granted in face to face meetings is inhibited by the absence of spatial cues (such as eye-contact), by the lack of a shared social and physical context, and by a limited possibility for informal communication. In fact, as various studies have proven, people’s communication behaviour while being connected through a standard audio-video link more closely resembles that of people talking over a phone than of people talking from face to face. [2] [1]. While this might not greatly affect tasks that involve the exchange and the presentation of existing information and documents, it does have a negative impact on tasks of a more creative nature.

In an attempt to simulate traditional face-to-face meetings more closely and eventually overcome the formal and mediated character of standard videoconferencing interfaces, various three-dimensional metaphors have been developed in videoconferencing applications. Early work introduced spatially positioned video and audio streams into the conferencing space (FreeWalk [13], Gaze [14], VIRTUE [15]), but without the addition of virtual content to be discussed in such a meeting. In contrast, SmartMeeting⁴ provides a highly realistic conference environment with virtual rooms with chairs, whiteboards, multi-media projectors, and even an interactive chessboard, but without spatially placed video representations of the participants. AliceStreet⁵ makes use of a similar concept, although with a more minimalist virtual room design, but the participants are represented here as rotating video planes sitting around a virtual table at fixed positions and watching each other or a shared presentation screen capable of displaying presentation slides.

The common goal of all of these approaches is to improve the usability of remote collaboration systems by decreasing the artificial character of a remote encounter.

**Mixed Presence Groupware**

Systems that support multiple simultaneous users interacting on a single shared display are categorized as Single Display Groupware (SDG) [4]. If a shared visual workspace also supports distributed participants in real-time, one can label such a system as Multiple Presence Groupware (MPG) ([16], see also [17]). If placed

into a place/time groupware matrix (see figure 1) it spans over the two places segments while still being synchronous.

<table>
<thead>
<tr>
<th>same place</th>
<th>different place</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>same time</strong></td>
<td><strong>different</strong></td>
</tr>
<tr>
<td>face-to-face collaboration</td>
<td>real-time remote collaboration</td>
</tr>
<tr>
<td>e.g. Coeno</td>
<td>e.g. cAR/PE!</td>
</tr>
<tr>
<td>co-located on-going work</td>
<td>asynchronous distrib. work</td>
</tr>
<tr>
<td>e.g. message boards</td>
<td>e.g. email work</td>
</tr>
</tbody>
</table>

Figure 1: Mixed Presence Groupware in place/time matrix

Tang et. al. identified only few MPG systems to date, a CAVE-like environment by SICS (Touch Desktop), Microsoft’s Halo, a split screen environment for the Xbox, and two video-overlaying systems without spatial arrangements of the participants. They found two main problems in using MPG systems: (1) Display disparity: considering the appropriate arrangement of persons and artefacts when using a mix of horizontal and vertical displays and (2) Presence disparity: the perception of the presence of others depending on whether s/he is co-located or remote. In our research presented in this article we will address both problems and try to find (partial) solutions.

**System Concepts Used**

Our approach is novel in that it combines and integrates several vital features found in other earlier work:

- We make use of a horizontal, interactive workspace to support creative group processes in a natural way and allow remote group members to be part of that process avoiding presence disparities.
- We combine interfaces of the remote and co-located worlds in a natural and easy-to-use way.
• We provide a system seamlessly combining a vertical and horizontal display system in a way that minimizes display disparities.
• We integrate the task space (data) within the work space (table environment) providing both with a task to focus on and a creative atmosphere.
• We offer private and public workspaces at different levels for all group members regardless of their location.

In the following we present in brief our earlier existing systems and how we combined them to create a novel collaborative environment.

3D Teleconferencing System: cAR/PE!

cAR/PE! is a teleconferencing system used with commonly available equipment: a PC with a web camera and a headset. It is designed for small group collaboration between Internet networked computers and it integrates data distribution and presentation with communication capabilities. cAR/PE! simulates a face-to-face meeting in a room and therefore uses the metaphor of a three-dimensional conference room [18].

All participants meet in this room and are represented by video avatars. The virtual room is “furnished” with a meeting table and several presentation screens to be used in a way as close as possible to a real world meeting (see figure 2). The participants can freely move around within this room, can place slides, movies, or pictures on the virtual screens or on the table, can share remote computer screens in an interactive way, and can put three dimensional virtual models onto the table to be discussed with others. The person’s movement within the room is visible to all other participants easing gaze and workspace awareness. This awareness is
further supported by the provision of three-dimensional sound (in particular to hear others from the right direction even they are not in the current field of view).

![Figure 3: cAR/PE! connection scheme](image)

From a technological point of view, cAR/PE! stations are connected via standard Internet as shown in figure 3. Up to six stations can be connected forming one virtual meeting space. The maximal number of stations depends on the bandwidth available and with standard ADSL connections three stations can be used with a good overall quality. All audio and video streams as well as the data distribution are implemented point-to-point, mainly for security reasons. All interactions occurring in a session (e.g. the movement of the participants within the room or changing slides on the virtual projection screen) are sent to a common request broker, which delivers the results to all stations. Supplemental remote computers can be connected to this cAR/PE! network. The content of the displays of these computers is displayed within the virtual cAR/PE! environment and can be operated interactively from within the meeting room.

Given these capabilities, the cAR/PE! system allows for synchronous collaboration over a distance while trying to maintain the metaphor of a traditional face-to-face meeting. Remotely located participants are able to focus
on their task and data (shared place) and to communicate in a natural way (shared space), because of the integration of both domains: data and communication. The system has been used in pilot installations in industry and academia and usability and social presence successfully evaluated with hundreds of subjects [18, 19, 20].

Some desired interface functionality cannot be supported yet, because of the technology used, or the inherent limitations of this dedicated distant communication and collaboration tool. For instance, by its very nature tangibility input is not supported by any means. Users operate the system using a traditional mouse and therefore all interactions are virtual. To visualize ideas in a real world scenario one would probably use paper and pen or a whiteboard, in a mouse operated virtual room this is inconvenient and less natural. In addition, co-located collaboration and the transmission of most non-verbal cues is poorly supported, even when used in combination with a projection system.

**Co-located Table-top System: Coeno**

Collaborative table-top setups are becoming increasingly popular for creative tasks. Coeno, is a collaborative table-top environment that is designed for brainstorming and discussion meetings. In Coeno, we particularly focus on a novel ubiquitous environment for creative sketching, drawing, and brainstorming (cf. Figure 4).

![Figure 4: People can discuss and brainstorm by directly interacting with the table and presenting their results on a rear-projection screen (a). Moreover, we support natural input devices (e.g. digital pens) (b).](image)

The application incorporates multiple devices and novel interaction metaphors supporting content creation in an easy-to-use environment. Our installation offers
a cooperative and social experience by allowing multiple face-to-face participants to interact easily around the shared workspace, while also having access to their own private information space and a public presentation space.

Figure 5: Coeno system configuration.

The installation itself consists of two main modules (cf. Figure 5):

1/ An Interactive Table, combining the benefits of a traditional table with all the functionalities of an interactive surface and display. The table allows people to easily access digital data and re-arrange both scribbles and virtual sketches in an intuitive way using different interaction tools.

2/ An Interactive Wall, consisting of an optically tracked rear-projection screen that displays digital content and captures gesture input. Combined with the Interactive Table, data can be seamlessly transformed from all presentation sources to the presentation wall.

The interface consists of two ceiling and one wall mounted projectors showing data on a table surface (Interactive Table) and on a rear-projection screen (Interactive Wall). All users can sit at the table and connect their own laptop and/or tablet PC computer to the display server. There is no limit as to how many clients can connect simultaneously to the system and the amount of co-located participants depends on the space around the table. In our case, typically 4-5 participants are involved in a meeting, where one of the participants usually leads the session.
Participants can interact with the table in several ways. They can either use their personal devices (e.g. tablet PC) wirelessly connected to the server, or a digital pen. Designers can create imagery on their own personal computers and “move” them to the interactive table for further discussion using hyper-dragging as proposed by Rekimoto et al [8].

Unlike Rekimoto’s work, users can also use real paper in the interface. To digitally capture handwritten notes, participants use the Anoto\(^6\) digital pen system. These are ballpoint-pens with an embedded IR camera that tracks the pen movement on a specially printed paper covered with a pattern of tiny dots. We use the Maxell Pen-It device with Bluetooth wireless connectivity. In our tabletop interface, we also augment the real paper with projected virtual graphics. The paper itself is tracked by using ARTag\(^7\) markers, placed on top of each piece of paper. Thus, participants can make annotations on real content that is combined with digital content projected on top of the paper surface.

Participants are able to use the Interactive Table as a traditional whiteboard for brainstorming tasks. We integrated a MIMIO device\(^8\), with ultrasonic tracking, which enables participants to draw on the interactive table and create annotations in real-time. Finally, the Interactive Wall is a rear-projection system which allows an intuitive gesture based interaction on a wall screen. We use a transparent rear-projection screen and track the user’s gestures with an infra-red (IR) camera setup. All of these devices can be used simultaneously and they combine input and output on one surface using several novel interaction metaphors. A closer description of the implemented interaction metaphors including a first pilot study is presented in Haller et al. [21, 22].

In summary, the Coeno interface combines three different display spaces:

- **Private Space**: The users’ own hardware device (e.g. laptop/tablet PC screen) and/or the area on the table around each participant. Other users cannot see the private information of the others.

- **Design Space**: The shared table surface (the interactive table), only visible to those sitting around the table. This space is mainly used during the brainstorming process.

---

\(^6\) [www.anoto.com](http://www.anoto.com)

\(^7\) [http://www.cv.iit.nrc.ca/research/ar/artag/](http://www.cv.iit.nrc.ca/research/ar/artag/)

\(^8\) [www.mimio.com](http://www.mimio.com)
Presentation Space: The digital whiteboard which is visible to all people in the room and therefore part of the presentation space. However, Coeno does not offer a remote, collaborative functionality. Therefore, we combined the advantages of cAR/PE! and Coeno into a first prototype, Carpeno, which is described in the next section.

A Combined Approach: Carpeno

Carpeno tries to overcome the barrier between co-located and remote collaboration while maintaining the interface advantages of table-top environments for creative group processes. Therefore a combination of the cAR/PE! and Coeno systems seems to be a promising approach. We will briefly introduce our conceptual idea and show a proof of concept with an initial, exploratory user study based on a first implementation of the concept. Our general concept is based around the obvious idea of combining the two approaches: (1) the table-top part of the Coeno environment and (2) the teleconferencing elements of cAR/PE! in a wall projection mode. The goal is to link these systems as closely together as possible to allow for a borderless communication and interaction space. Figure 6 shows the setup in a simplified manner.
Coeno’s private space is preserved and the data and interface components are still used in the same or even enhanced way as the design space introduced earlier. The presentation space is replaced by a screen projection showing the remote cAR/PE! virtual meeting room environment. This should create the impression for the local participants of two tables placed next to each other: the physical local table and the remote virtual table, both interactive and suitable for information display. The remote cAR/PE! participants can still freely move around in the virtual space. With this they are able to form an own shared space out of reach and sight of the local participants (similar to their local shared space). Both sides of the setup are coupled via (1) the display of the video and audio streams, including their (changing) locations and (2) data transfer and interactions coupled between the systems. Figure 7 illustrates the new communication and interaction spaces with Carpeno.
The central shared element between all participants (local and remote) is the virtual table within the (former) cAR/PE! environment, called the Common Shared Space. Local spaces are provided for each group: the local shared space on top of the physical table and the remote shared space everywhere within the cAR/PE! environment outside the reach of the local group. For example, the remote participants can choose a corner (and virtual table or presentation screen if needed) within the virtual environment and come back to the common shared space (virtual table) for discussions concerning the entire group.

The private spaces are on each side personal information systems (in most cases laptop computers or tablet PC’s) connected to the Carpeno system, but only visible to the individuals. Digital content can be shared via hyper-dragging or screen sharing, visible to a sub-group (e.g. local only) or the whole group (e.g. on the virtual table). Furthermore the virtual presentation screen within the cAR/PE! environment can be made visible to all for group discussions.
With this concept a new technological infrastructure and features have to be
developed. Figure 8 illustrates how Coeno and cAR/PE! are linked together to
form the seamless Carpeno system. As shown, the networked part of the cAR/PE!
system remains almost entirely unchanged, while the data and interaction
components are extended by the Coeno interface. We adopt a loosely coupled
approach, where network messaging techniques are used as the main software
technical method. With this we are able to control almost all of the aspects of the
cAR/PE! part of the system with the Coeno part and vice versa.

A virtually infinite number of even mixed local and remote stations can be linked
together without any system-inherent limitations. The main reasons not to do so
are: (1) limited bandwidth and other networking issues, (2) the (virtual) placement
of a certain number of persons and parties around one virtual table, and (3)
interface issues that have to be solved beforehand (e.g. orientation of documents,
pointers indicating interacting persons, etc). Currently two to six co-operating
parties can be brought together in one Carpeno system without serious problems.

**Prototype Implementation**

The first implementation of our conceptual approach serves as a test bed for
evaluating the feasibility of the Carpeno concept. Our focus therefore is set on
building a functioning and tangible system to be used for testing rather than on
providing the most comprehensive and complex solution first. We decided not to implement and integrate all features available in cAR/PE! and Coeno but rather to develop a system which can be initially tested in exploratory studies.

**System**

The initial version includes the following elements (see figure 9):

A vertical Plasma projection screen (WXGA resolution) displaying the remote shared space. The size of this screen was chosen to provide a wide field of view for the local party. The screen is accompanied with speakers to display the (spatially arranged) voices of the remote participants to the local group in a convenient way.

The local shared space is defined by a touch sensitive surface\(^9\) on which a projector (XGA resolution) shows the augmented surface content. With this setup one person at a time from the local group can directly interact with the digital content displayed simply by using his or her finger.

---

\(^9\) www.nextwindow.com
The augmented surface content is provided by the cAR/PE! system: An additional computer is rendering the same environment as shown on the vertical screen, but from a correct perspective from above the physical and virtual table. With this pre-configured setup we can ensure that both sides, local and remote, see the same content on the table.

To capture the live video stream of the local participant(s), we placed an Apple iSight camera on top of the Plasma display. While the image quality of the camera is superior for teleconferencing purposes, no real eye-to-eye contact can be achieved. In a standard situation, where the remote and local participants are sitting, this is still the best camera position, because it is close to the remote participant’s eyes.

Within the shared cAR/PE! environment the virtual content on the table is provided via a VNC application sharing component. The Coeno system connected to the network is providing this screen stream and resides on an additional computer.

In summary, three components from the cAR/PE! system are involved in the Carpeno setup: (1) the remote participant working at a standard PC screen, (2) the vertical screen (Plasma) of the local setup, and (3) the horizontal screen (touch screen) of the local setup. We have configured and calibrated these three components in a way that they form one, consistent spatial environment.

The local private space is provided by a tablet PC standing beside the touch sensitive surface. It is used to prepare content to be discussed in the group and to drag and drop it to and from the local shared space using the hyper-dragging metaphor. While for the users this interaction is a transparent one, the actual technical process is implemented via VNC application sharing feeding the cAR/PE! applications. All three cAR/PE! components receive the same VNC stream and display it on top of the virtual table.

All computers involved in this initial Carpeno setup are linked via a dedicated network switch, ensuring the highest possible networking performance. While we could have chosen virtually any video and audio codecs in this network setup, eventually we opted for high quality videoconferencing standards (G.711 uLaw and H.261 CIF) to emulate an Internet connection.

In this version we have reduced the conceptual number of possible spaces to three to ease our exploratory studies. The virtual table (common shared space) and the
projection onto the physical table (local shared space) are exactly overlaid to give the impression of one single table surface. Therefore, what the remote participants see on the virtual table is exactly the same what the local participants see. In addition, we abandoned the use of additional PC’s on the remote side (remote private spaces) to avoid confusion about the interface in the first instance.

Figure 10 illustrates our implementation. The Coeno system delivers all content via the application sharing functionality of cAR/PE! (sharing parts of the computer screen), while the interaction with the content of the common shared space is controlled by the touch sensitive surface. This system allows for actual communication and interaction within the Carpeno concept and serves as the basis for our exploratory user study described in the next section.

**Exploratory Study**

We conducted an informal exploratory study with our first prototype system. In total forty visitors at the ICAT2005 and Graphite2005 conferences participated in a hands-on evaluation during the exhibition of our system (see figure 11).
Task
Two persons at a time took a seat at different parts of our booth. One part was configured as a Carpeno station as described in the Implementation section and the other part was set up as a cAR/PE! station using a standard PC and Monitor equipped with a headset and a web cam. If only one volunteer was available, one of the exhibitors took on the role of the second person at the cAR/PE! side. Photographs of interesting looking devices that were invented during the last 200 years (taken from [23]) were then dragged onto the shared table by a moderator. The task for the participants was to collaboratively discuss what exactly the purpose of the displayed objects might be. If a device’s function could be guessed correctly, that picture got removed from the table by the moderator. All pairs had to discuss five to six different photographs in order to clear the table while playfully exploring the features of the Carpeno setup at the same time. To complete one round typically took between 5 and 8 minutes.

Questionnaire
After a team completed the task, both participants were asked to fill out a short questionnaire. Besides usability issues we were especially interested in finding
potential research variables that would arise from the asymmetrical nature of our setup. Most results that are presented in the following section are therefore presented separately for cAR/PE! and Carpeno users.

Results

After each session users were asked to subjectively rate the experience by answering nine seven-point Likert-scale questions. The questions and their normalised scores are summarized in Figure 12.

The scores in the satisfaction questions Q1 and Q2 show that both user groups liked the system. With the exception of question Q6, the answers on general usability issues (Q3 to Q7) further show an overall positive response. The lower score of Q6 uncovers that users of both sides could not easily infer where the other person was looking at. This deserves further investigation but could be influenced by the fact that there was a very high task focus. No major differences in the usability scores emerged between the Carpeno and cAR/PE! side. However, cAR/PE! users were more aware of the other person’s presence, as can be seen in the scores of question Q8, probably due to their undisturbed concentration on one screen surface (the monitor). The biggest difference between both user groups
emerged in question Q9. Carpeno users felt much more that the meeting with the other person occurred “locally”, i.e. around the physical table in front of them. On the other hand, cAR/PE! users thought the meeting took place more “remotely”, situated somewhere in the middle between their and the other person’s location.

Although we haven’t carried out formal statistical tests in this exploratory study, we can derive some initial lessons:

1) The low gaze awareness that appeared in question Q6 suggests that this issue demands some more attention in our setup. Applying head tracking technology that allows users to control their video avatar simply by moving their heads could deliver some improvements and would get rid of the need for mouse-based navigation. In addition, other gaze awareness support could be integrated such as the “miner’s helmet” metaphor [14] that displays a lightspot at a person’s centre.

2) The lower awareness of the partner’s presence in the Carpeno setup might be a result of the carpe user “disappearing” from the Carpeno-user’s screen when navigating to the other side of the table in the cAR/PE!room. This often led to confusion on the Carpeno side. Seeing the other person at all times therefore seems to be crucial for the awareness of the other’s presence, even if the audio connection is maintained. In future experimental setups, we therefore have to limit the navigation space for the cAR/PE!-user to an area where s/he is always visible to the Carpeno user.

3) The clear result about the experienced location of the meeting (Q9) suggests that users are very much able to associate a remote encounter with a spatial reference frame somewhere between “here” and “there” as it is defined by the interface. To understand the effects on the user and how exactly we can move both interface types along this dimension will be part of our future research.
Discussion & Future Work

Our conceptual approach in bringing together co-located and remote collaboration into a single system as well as our first implementation suggests that the Carpeno interface has indeed great potential for enhancing remote face-to-face collaborative creative experiences. Our initial, exploratory user study with Carpeno and the numerous experiences with the single systems cAR/PE! and Coeno lead us to develop requirements a future Carpeno system should have and opens up new research areas to work on.

Our initial assumption was supported, that the combination of our two systems can compensate for the flaws in interfaces detected in the separated systems. In particular the incorporation of remote participants into the co-located collaboration is possible and the provision of a table-top environment for the remote participants is of great value, especially in creative tasks like brainstorming or general discussions involving some sort of media. Eventually we can provide a common shared space as well as local shared and private spaces at the same time.

Direct manipulation on the interactive table is intuitive and can be supported by different interfaces, depending on the particular task to be addressed. For our picture sharing application finger pointing was very appropriate. Participants have different preferences and different tasks require different input devices (e.g. digital pen, tablet PC, Mimio tracking device, etc.). Therefore, one of our goals is to test the different benefits of these devices.

The incorporation of a table as the central element of our interface (real and virtual) and the consequent integration into a meeting environment (also both real and virtual) leads to the reasonable approach of (“re-“) introducing spatial objects into the process and interface. On the physical side (real world) real objects can be used as part of the creative group processes or as part of the interface (tangible user interface, see [24, 25, 26]). On the virtual side (and within the virtual space) the use of 3D virtual objects representing the real world can be used also either as the object of discussion or as interface elements. Further research is needed here and should be based on existing findings and systems (in particular tangible and perceptual user interfaces, ubiquitous computing, 3D user interfaces).

For the sake of simplicity and to rapidly allow for an early exploratory study we’ve excluded some interfaces, which would be very relevant in non-
experimental situations. We are going to amend the system with a shared digital whiteboard, better support for gesture communication, and pen-based interaction. Also, the (simultaneous) placement of documents in the shared spaces will be approached based on the experiences made with the single systems. For example, mechanisms already built-in into the Coeno system can be used for a “real estate” saving arrangement of documents onto the limited virtual and real table space. While general gaze awareness could be provided with our Carpeno system, eye-to-eye contact is still not possible because of the different locations of the real camera and the virtual participant representation as a video stream. We are working on optical and/or IT solutions to allow for this essential aspect in certain task scenarios (like negotiations). The form of representation of the avatars itself (video stream on a moving virtual plane) was acceptable. This was already tested in earlier studies with the cAR/PE! system. However, to provide even better communication cues and channels, we are going to test, whether other forms of representations (e.g. with background eliminating methods) can even enhance the overall quality.

In addition, our first implementation was mainly limited to one remote and one local person. We are exploring how the system has to be modified to add more local or remote participants. Issues that must be addressed include concerns such as: Do all of the participants meet in the local (physical) or the remote (virtual) place? or How does informal communication between co-located participants affects the entire, creative process with the remote participants? These questions have to be answered in the future, involving more creative tasks besides brainstorming and/or picture sharing.

With our current, integrated approach the development of new interface metaphors and techniques considers the combined support for local and remote collaborative tasks at an early stage. It can be assumed that this consideration leads to more comprehensive and efficient interfaces suitable for both worlds, the local and the distant one. This could be a satisfactory contribution to tool and process development of a converging world of communication and information. Last but not least, communication quality can be improved in using the Carpeno approach. Especially support of non-verbal communication cues in relation to a high level of social presence seems to be essential and can be implemented on our current basis. For instance the introduction and evaluation of head-tracking, gaze
and workspace awareness supporting techniques for natural gesture recognition, and eye-to-eye contact in remote settings are part of our future research.

Acknowledgements

We would like to thank Claudia Ott, Michael Wagner, Graham Copson and the Technical Support Group at Otago University, and all the participants in our experiments for their great support. In addition, we would like to thank DaimlerChrysler Research and Technology for supporting our work and the anonymous reviewers with their comments, which lead to some very relevant improvements. The Office of Tomorrow project is sponsored by the Austrian Science Fund FFG (FHplus, contract no. 811407) and VoestAlpine Informationstechnologie. Moreover, the authors would like to thank Daniel Leithinger, Jakob Leitner, and Thomas Seifried for the great work in the Coeno project.

References


Abstract

The Shared Design Space is a novel interface for enhancing face-to-face collaboration using multiple displays and input surfaces. The system supports natural gestures and paper-pen input and overcomes the limitations of using traditional technology in co-located meetings and brainstorming activities.

CR Categories: H.5.1 [Information Interfaces and Presentation]: Multimedia Information Systems—[H.5.3]: Collaborative computing—Computer-supported cooperative work

Keywords: Augmented Reality, Tabletop Environment, Sketching, Design Room, Design Environment

1 Overview

The project is part of the research project “Office of Tomorrow”, a collaborative tabletop environment, designed for presentations and discussion meetings. In this installation we particularly focus on a novel ubiquitous environment for sketching, drawing, and brainstorming. The application incorporates multiple devices and novel interaction metaphors to create an easy-to-use environment. The installation offers a cooperative and social experience by allowing multiple face-to-face participants to interact easily around a shared workspace, while also having access to their own private information space and a public presentation space. The installation itself consists of two modules:

1. An Interactive Table, which combines the benefits of a traditional table with all the functionalities of a touch sensitive digital table and display.
2. A Digital Whiteboard, consisting of an optically tracked rear-projection-screen that displays digital content and captures user gesture input. Combined with the Interactive Table, data can be seamlessly transformed from all presentation sources to the presentation wall.

Both devices can be used simultaneously and combines input and output on one surface. Based on these two devices, we implemented several novel interaction metaphors. Moreover, we combined the usage of traditional input devices (e.g. laptops) with digital pen input using real and virtual paper. Users can create imagery (e.g. scenario sequences, scribbles, 3d content) on their own personal computers, move them to the Interactive Table for discussion, and present them to the audience by using the Digital Whiteboard. The combination of digital information and real paper is realized by using the Anoto pen and its paper technology. Anoto-based pens are ballpoint-pens with an embedded camera that tracks the movements simultaneously. The pen has to be used on a specially printed paper with a pattern of tiny dots. In our setup, we use this technology to combine real paper with digital augmented content. Participants can make annotations on digital content that is projected on the top of the paper surface. The paper itself is tracked by using computer vision techniques and special ARTag markers [Fiala 2004]. Special control elements on the paper allow a copy of the digital data to be made to the private spaces of the collaborators. Thus, once one participant finishes making an annotation, he or she can send it to the other’s workspace on the Interactive Table and they can modify it accordingly. While one participant is writing on real paper, the collaborators get a ”digital copy“ projected on their own paper and all modifications by the first participant are projected onto the paper sheets of his/her collaborators.

Figure 1: Videos, images, and 3d objects can be projected on the paper and be combined with annotations made with the digital pen.

The Shared Design Space allows people to communicate as they normally would around a table and points to a future where computers will be able to naturally support face-to-face meetings and creative design sessions. More details (papers, pictures, movies), illustrating our system in action, can be found online at: http://www.coeno.org.

References

Shared Design Space: Sketching ideas using digital pens and a large augmented tabletop setup

Michael Haller¹, Peter Brandl¹, Daniel Leithinger¹, Jakob Leitner¹, Thomas Seifried¹, and Mark Billinghurst²

¹ Digital Media, Upper Austria University of Applied Sciences, Austria
² HITLabNZ, University of Canterbury, New Zealand
coeno@fh-hagenberg.at

Abstract. Collaborative Augmented Reality (AR) setups are becoming increasingly popular. We have developed a collaborative tabletop environment that is designed for brainstorming and discussion meetings. Using a digital pen, participants can annotate not only virtual paper, but also real printouts. By integrating both forms of physical and digital paper, we combine virtual and real 2d drawings, and digital data which are overlaid into a single information space. In this paper, we describe why we have integrated these devices together in a unique way and how they can be used efficiently during a design process.

1 Introduction

In recent years, Augmented Reality applications have been developed for many different platforms, such as mobile phones and handheld devices and also tabletop environments. Kiyokawa et al. describe the communication behaviors in a tabletop collaborative AR interface setup and they present several ways to improve the face-to-face collaboration by using AR [1]. In this paper, we describe a novel tabletop AR environment suitable for enhancing face to face collaboration, especially in the design process.

Designers and people who are discussing and brainstorming usually work in a studio surrounded with sketches, which are either pinned on a wall or placed on large surfaces. Currently, new sketches are mainly created directly on paper on the drafting table before developing a digital mock-up model on the computer. With new technology it may be possible to enhance this process. Blinn [2] postulates that the creative process is a two-phase process: firstly, moving from chaos to order and secondly, from ideation to implementation. Most computer-based design tools are primarily focussed on the second phase, and there is limited support for digital tools where people can play with ideas in a free form manner. Digital tabletop setups would be an ideal interface for sketching out a crude version of an idea. In the creative process, people still prefer using paper and large tables to capture their ideas. Therefore, the table still remains the main interaction device during the creative process. Augmented with virtual elements, a tabletop setup becomes an ideal input and output device around which people can share a wide range of verbal and non-verbal cues to collaborative effectively (cf. figure 1).

In this paper, we describe the combination of different hardware devices that can be combined to develop an AR-based tabletop environment for creating efficient applications [3]. Now, that it is technically possible to develop large augmented surfaces, it is
important to conduct research on the different types of collaborative AR applications that are ideally suited for these types of tabletop setups and to present "user interface guidelines" for developing these applications. In the next section, we review related work on tabletop collaboration environments. Next, we present our system, focusing on hardware and design decisions. Section 3 focuses on interaction techniques. From the early user feedback, we conclude with design guidelines for collaborative tabletop AR systems and directions for future research.

2 Related Work

Early attempts at computer enhanced face-to-face collaboration involved conference rooms in which each participant had their own networked desktop computer that allowed them to send text or data to each other. However, these computer conference rooms were largely unsuccessful partly because of the lack of a common workspace [4]. An early improvement was using a video projector to provide a public display space. For example the Colab room at Xerox PARC [5] had an electronic whiteboard that any participant could use to display information to others. The importance of a central display for supporting face-to-face meetings has been recognized by the developers of large interactive commercial displays (such as the SMARTBoard DViT ).

In traditional face-to-face conversation, people are able to equally contribute and interact with each other and with objects in the real world. However with large shared displays it is difficult to have equal collaboration when only one of the users has the input device, or the software doesn’t support parallel input. In recent years, Stewart et al. coined the term Single Display Groupware (SDG) to describe groupware systems which support multiple input channels coupled to a single display [6]. They have found that SDG systems eliminate conflict among users for input devices, enable more work to be done in parallel by reducing turn-taking, and strengthen communication and collaboration. In general traditional desktop interface metaphors are less usable on large
displays. For example, pull down menus may no longer be accessible, keyboard input may be difficult, and mouse input may require movement over large distances. A greater problem is that traditional desktop input devices do not allow people to use free-hand gesture or object-based interaction as they normally would in face-to-face collaboration. Researchers such as Ishii [7] have explored the use of tangible object interfaces for tabletop collaboration, while Streitz et al. [8] use natural gesture and object based interaction in their i-Land smart space. Regenbrecht et al. extend the idea of tangible user interfaces and demonstrates the benefits in novel video conference system [9]. In many interfaces there is a shared projected display visible by all participants; however, collaborative spaces can also support private data viewing. In Rekimoto’s Augmented Surface’s interface [10] users are able to bring their own laptop computers to a face-to-face meeting and drag data from their private desktops onto a table or wall display area. They use an interaction technique called hyper-dragging which allows the projected display to become an extension of the user’s personal desktop.

In the next section, we describe our AR tabletop system which combines these features. Unlike previous work, our system benefits from the following features:

- Seamless combination of both real and virtual data combined with augmented content,
- Intuitive data exchange using multiple (heterogeneous) devices based on natural and easy-to-use metaphors,
- Adapted and modified interaction methods (e.g. improved hyper-dragging for objects, Pick-and-Drop etc.), and
- Detailed discussion of what customers expect while using such a system. From the first meetings with our customers, we achieved a couple of interesting ideas which somehow diverge from the requirements seen by the developers.

3 System Overview

Our interface consists of four ceiling and a wall mounted projector showing data on a table surface (Interactive Table) and on an adjacent wall (Interactive Wall). All ceiling projectors are connected to a single display computer (cf. figure 2). Users can sit at the table and either connect their own laptop and/or tablet PC computer to the display server or interact directly with the table using digital pens. There is no limit as to how many clients can connect simultaneously to the system, and the amount of co-located participants depends on the space around the table. In our case, typically more than 5 participants are involved in a meeting, where one of the participants usually leads the session.

The tracking is realized by using large Anoto³ patterns and digital pens from Maxell. Anoto-based pens are ballpoint-pens with an embedded camera that tracks the movements simultaneously (cf. figure 2). The pen has to be used on a specially printed paper with a pattern of tiny dots. Each paper sheet is unique and has its ID. The pen with its inbuilt infrared (IR) camera tracks the dots of the paper to retrieve both the ID of the paper and the position relative to the upper left corner of the paper. All required

³ www.anoto.com
The system consists of an interactive table and an interactive wall. The tracking on the table is realized using digital pens which can track the tiny dots of the Anoto pattern. To protect the Anoto pattern, we put a Plexiglas cover, which however, does not interfere the accurate tracking results.

Information can then be sent in real-time to the PC using Bluetooth. Usually, the digital pens are used in combination with real printouts and real notebooks. However, we use this technology for tracking the users’ movements and interactions on the whole table surface. To do this we put two A0-sized Anoto-patterns under the Plexiglas cover placed on the table (cf. figure 2). This allows us to track users’ gesture input while holding the digital pen. Once the user touches the table with the pen, the camera tracks the underlying Anoto paper. Images projected onto to the table do not interfere with the tracking. Since every digital pen has its own personal ID, we can easily identify who is interacting with the interface without any additional hardware requirements (e.g. capacity measurement on the chairs [11]). Notice that the pens are also pressure sensitive which allows additional functionalities (i.e., for a better control in a sketching/drawing application).

We also allow real paper to be used in the interface. Real paper is socially well-accepted in meetings and does not interfere with the face-to-face collaboration. In design meetings, participants often have to handle paper documents and drawings. Paper
also has the advantage that it gives a fast and quick overview, provides a high resolution surface and is easy to carry. However, it can be difficult to transfer sketches on paper into digital applications. In our AR tabletop interface, we augment real paper with projected virtual graphics [12]. Participants can make annotations on the real content that is combined with digital content projected on top of the paper surface. The paper itself is tracked by using ARTag [13] markers, printed on each piece of paper.

Whenever a sketch is finished, it can be presented to the audience by moving it to the Interactive Wall. For the Interactive Wall setup we use a 60” diffuse Plexiglas display surface with special qualities for rear-projection. The anti-reflex surface homogeneously distributes the light to achieve a balanced image. A single camera tracking solution provides the interaction data for the sketch presentation. An IR pass filter that is applied to the camera avoids interferences with the projected image. As the tracking is done via shadows in the IR spectrum, IR lights are needed to provide a homogenous illumination in front of the screen. Touching the screen with one hand, the user can scroll through the sketches. A timeline on top provides a map-like overview that can also be used to navigate in a fast-forward style. Using both hands activates the zoom mode on screen, changing the distance between the hands affects the zoom-level accordingly. Notice that Anoto pens cannot be used on the interactive wall due to technical limitations: Although the embedded IR camera can track the graphite ink even on transparent surfaces (e.g. a pattern printed on a transparent foil), the rear-projected light interferes the tracking results.

4 Interaction Techniques

![Shared Design Space Interaction](image)

**Fig. 3.** Shared Design Space Interaction. (a) Designers can choose different color from a tangible interface (b) and sketch directly on the private workspace. (c) Optionally, users can also use pens with different caps.
In this section, we mainly focus on novel interaction techniques we used in our project. We also present the different interaction possibilities and highlight why we have chosen the individual interface metaphors that we use. During a face-to-face brainstorming process, we can identify the following main tasks, which require adequate interaction techniques:

– Data Creation and Selection,
– Data Transfer, and
– Data Manipulation.

### 4.1 Data Creation and Selection

Usually, in the design process participants already have some sketches, notes, or drawings they want to share. On the tabletop system, users have two methods for data sharing, depending on the device they are using. If they have already some sketches on their private devices (e.g. tablet PC, laptop), the user can use traditional desktop tools for data creation. They can also create new sketches on the table using real/virtual paper and the digital pen. By observing users, we noticed that people preferred using the digital pen more; especially for sketching quick ideas. In fact, our system allows a seamless combination of digital and real content, where the augmented content has mainly be picked up from the table’s surface. Thus, the table becomes a storage pool of digital content, where participants can share data. It is still a challenging task in which way the content initially comes onto the table. As mentioned before, we use additional devices, where the content can be moved smoothly to the table by hyper-dragging [10]. In the early phase, we did also experiments with a 1-degree of freedom (dof) menu, which was directly projected on the table surface (cf. figure 4). After drawing an ink-rectangle with

![Fig. 4. 1-degree of freedom menu for selecting different 3d geometries.](image-url)
are printed on the real paper sheet. We used a default directory, which was browsed and shown in the menu. Once the corresponding geometry has been selected (again the user has to select the corresponding checkbox on the paper), the printout can be moved including the augmented content (e.g. videos). Notice that people can also select 3d geometries simultaneously (cf. figure 4).

4.2 Data Transfer

In interfaces that contain multiple displays, an important research question is how to transfer data from one source to the other. In our setup, users sitting around the Interactive Table can move notes to the Interactive Wall by touching special control elements, projected onto the table surface and using the digital pen (cf. figure 5).

![Figure 5](image)

**Fig. 5.** Special control elements projected on the table surface allow to share the sketches, but also to transfer them to the Interactive Wall.

As noticed by Rekimoto in [14], people often have to transfer data from one device to another - especially in an augmented tabletop environment. In most cases, the traditional interaction metaphors (e.g. drag-and-drop) fail and more natural and quick data transfer metaphors have to be found.

In our setup, participants have three different possibilities of moving data from one screen to another depending on the device they are using (i.e. wireless presenter, personal device (e.g. laptop, tablet PC), and the Maxell pen).

4.3 Moving Data with the Laptop PC (Hyper-Dragging)

In our setup, all participants around the table can quickly re-arrange the content on the table using their personal devices and the user’s mouse attached to their own devices.
Therefore, users can click on a virtual image on their laptop computer desktop and drag it. Once the mouse reaches the edge of the physical display, the content appears on the table connected by a virtual line to the center of the desktop (cf. figure 6). Dragging with the mouse continues to move the note across the table.

![Figure 6](image)

**Fig. 6.** Using the hyper-dragging metaphor allows an easy-to-use transfer of data between two devices (e.g. tablet PC and tabletop desk).

Participants can create new content on their private device and then move it to the public space. However, the direct interaction with content on the table surface is in many ways more natural that using mouse cursor input. In pilot studies people expected to be able to interact with their images once they were on the table surface and were surprised when this was initially not possible. We found that some people preferred using a device rather than interacting with their fingers. For this reason we added support for input from the Anoto digital pen.

### 4.4 Moving Data with the Digital Pen (Pick-and-Drop)

The combination of both real and virtual data requires a novel metaphor for data transfer. Instead of using the Drag-and-Drop metaphor, which is mainly used in 2d-based GUI applications, Rekimoto proposed the Pick-and-Drop metaphor [14]. It is more natural if users can manipulate and insert data in a computer environment as if it would be a real, physical object. In our setup, users can pick-up digital data from the table (e.g. video, images, 3d geometry) and drop-it on the real paper by using the digital pen. Once the users pick an object, the digital pen tracks the underlying Anoto paper.

Once the user taps a digital object on the interactive table, our content manager automatically binds it (virtually) to the pen (cf. figure 7). Whenever the user moves the same pen onto the real paper, the manager transfers the data to the server that displays it on the interactive table. During our first tests, we recognized two reasons that the Pick-and-Drop metaphor is more convenient than the Drag-and-Drop interaction. Firstly, our pens did not have additional buttons, which would have been necessary for realizing
Fig. 7. Pick-and-Drop. Users can pick (a) an object from the interactive table to drop it (b) onto the real paper.

the Drag-and-Drop metaphor. Secondly, large movements on the table can be really uncomfortable and time-consuming, and are not required in the Pick-and-Drop metaphor. Robertson et al. presents improved Drag-and-Drop metaphors for wall-size displays for multiple data sets that could be integrated in the future [15].

4.5 Moving Data on the Interactive Wall

Once the digital data has been sent to the interactive wall, participants can move the data with users’ gestures. We noticed during our first tests with end-users that the interactive wall is mainly used for presenting intermediate results. We also noticed that in most cases one person (mostly the coordinator of a session) stands up and presents the results to other participants.

4.6 Data Manipulation

On the interactive table, people can just move the sketches and perform simple transformations, such as translating, scaling, and rotating. Once participants want to manipulate the data sheet, they have to move it to private space (either to the personal tablet PC by hyper-dragging or to the personal workspace, as depicted in figure 3a).

An easy-to-use sketching tool allows the creation of simple strokes which are layered either onto the virtual data or onto the real printout (cf. [16][17]). The colors and brush strokes can be changed by using different tangible tools (e.g. the color chooser as depicted in figure 3). Again, the tiles of the color palette are printed with the special Anoto pattern. As mentioned before, users can also integrate virtual content (i.e., 2d images, movies, or 3d geometries) into the printout and/or the virtual paper. Not surprisingly, people also expect to have the possibility to transform the geometry or to start and/or stop a projected video. The ultimate goal would be that people draw their own control elements onto the paper and control the virtual content accordingly.
4.7 Group Interaction

Using a collaborative interactive table means sharing information and working together. Often, in traditional meetings, people have neither enough hard-copies nor enough space for the large sketches (e.g. CAD sketches). Group interaction often becomes difficult and in many cases it is hard for user to see where their colleagues are pointing to. Moreover, the manual annotations (mainly done with markers or pens) cannot be stored digitally. Figure 1b depicts a scenario where both users are working on the same paper sheet. All modifications are sent from the digital pen through Bluetooth to the server and the individuals’ view of the projected comments is updated accordingly. The "shared desktop", which is embedded as a window in front of the collaborators, becomes a WYSIWIS (What You See Is What I See) window and all participants can see the same thing. The content control is mainly left to social norms. Discussing with participants, they have not felt the need to explicitly lock modification control over data objects, because they could always see who was attempting to modify their objects. However, it is important that users can identify who is manipulating each object (e.g. by using different colors).

5 Implementation

A system overview is presented in figure ??, which depicts the most important components. The system is written in C++ based on OpenGL for rendering the virtual content and on ClanLib for rendering the graphical user interface. The advantage of the ClanLib library is that the system is based on top of OpenGL and can therefore be combined easily with the basic rendering system of the 3d graphics library. In addition, we used the component model of the ClanLib library for the inter-component communication and the network extension for the communication between the clients and the server. The tracking of the markers attached on the printout was performed by the ARTag tracking library. The real-time tracking of the Maxell pen was implemented in C# by using Maxell’s Streaming API. Moreover, we use an in-house library for recognizing the axis-aligned rectangles drawn on the paper. The tracking on the rear-projection screen is mainly based on both the JazzUp library and OpenCV.

6 Conclusions and Future Work

In this paper, we presented the results of combining different easy-to-use interaction metaphors in a tabletop setup using digital pens and paper. Previous research demonstrated that AR technology can be used to enhance face-to-face collaboration in a natural way. The first results were strongly influenced by the first informal tests with our customers. The setup incorporates multiple devices and novel interaction metaphors to create an easy-to-use environment. The installation offers a cooperative and social experience by allowing multiple face-to-face participants to interact easily around a shared workspace, while also having access to their own private information space and a public presentation space combining both virtual and real sketches. The project is different
from typical screen-based collaboration, because it uses advanced interface technologies to merge the person and task space. It is also uses multiple display surfaces to support private, group and public data viewing and interaction. The Anoto-based pen allows an ideal interface for interacting with both the real and virtual paper. Special control elements, projected on the desk, allow to share the own work for further discussion with the participants. Moreover, we included intuitive and natural objects (e.g. color palette etc.) and combined the real paper with digital content. We also investigated adequate and intuitive interaction metaphors for data movement (e.g. hyper-dragging, pick-and-drop etc.) and demonstrated their usage in the tabletop setup. We have also shown that the simultaneous use of the digital pen for the printout and for the interaction on the interactive table makes a lot of sense. The Anoto pattern used for tracking the users’ movements seems to be an ideal configuration for similar tabletop applications.

Currently, we are starting with a formal evaluation. The main goal is to find out how the participants are using the real paper and in which sense they benefit from the augmented annotations. Finally, we also want to find out, in which sense a centralized data placement differs from a replicated note (cf. [18]). In a centralized design, participants only see one copy on the table. This often causes orientation and viewing problems by the participants who are sitting on the opposite of the table. Moreover, people have often problems, if one person is pointing to an object. To address these limitations, we also implemented the replicated view, where everybody can get the same view projected onto the personal paper sheet. Moreover, everybody can edit the paper simultaneously. In which sense traditional meetings will be influenced by a simultaneous interaction, will be investigated in a next user-study.

7 Acknowledgements

This research is funded in part by the Austrian FFG consortium (FHplus framework, no. 811407) and voestalpine Informationstechnologie. We also thank both Maxell and Anoto for their support. Finally, we thank our ”Office of Tomorrow” team for providing inspirational ideas and constructive support for our work.

References

What’s next? Outlook to the future

*Bruce H. Thomas*

**Overview**

- **Wearable Computing Interfaces**
  - Information on the go
  - Garment integrated technologies
- **Augmented/Mixed Reality Interactions**
  - What is augmented/mixed reality?
  - Indoor application examples
  - Outdoor application examples
  - Through walls collaboration
Wearable Computing Interfaces

Information on the go

Information Not Just Data

• Access
  – Fast
  – Reliable

• Collaboration
  – Co-located
  – Remotely located
  – Real-time
  – In Situ Visualisation and Manipulation of Data
Usability Challenges

• What do people need?
• What do people want?
• More information promotes information overload
• Smaller devices maybe harder to control
• People do not want to look like freaks!
• “Walk up and ready to use”
• People will want to be able to rely on it.

Usage Challenges

• Mobile contextual information
  – Access to appropriate information all the time
  – Understand where the user is
  – Understand what the user is trying to do
• Collaboration tools
  – Intense time critical work
  – Reach-back for experts
  – Making it easy for groups of groups of people to work
Lugable

Wearable Computing Interfaces

Garment integrated technologies
Wearables (Uniform of the Day)

Components of the e-SUIT

- **Inputs**
  - Multi button capacitive keyboard

- **Displays**
  - 2 Line X 64 Character Watch Display
  - 3 colored LEDs in jacket cuff
  - Single pager motor OR
  - Six pager motor shoulder pad
  - iPAQ PDA
The E-textile Construction Kit


Wearable Displays
String-Based Interaction


Stock Browser Example
Augmented / Mixed Reality Interactions

What is augmented/mixed reality?

Video Augmented Reality

- Real world is captured by video camera
- Mixed with virtual world in computer
- Displayed on VR head mounted display
- High quality output for capture

Diagram:
- Camera
- Real World
- Virtual World
- Computer
- VGA Output
Mixed Reality: Two Definitions

- Milgram, Takemura, Utsumi, and Kishino

“… a generic Mixed Reality (MR) environment as one in which real world and virtual world objects are presented together within a single display, that is, anywhere between the extremes of the RV continuum.” (SPIE Vol. 2351, Telemanipulator and Telepresence Technologies, 1994)

Augmented / Mixed Reality Interactions

Indoor application examples
VITA: Visual Interaction Tool for Archæology

- Hybrid UI
- Combines multiple
  - Users
  - Displays
  - Interaction devices

(Used with permission of Steven Feiner, Columbia University)

Cross-Dimensional Gestures
Bridging the Gap between 2D and 3D

Cross-Dimensional Gestural Interaction Techniques for Hybrid Immersive Environments

Hrvoje Benko, Edward Ishak, Steven Feiner
Columbia University

IEEE VR 2005
Display-based Tracking & Control System

- Can be used for both input and output

(Used with permission of M. Inami, University of Electro-Communications)

Augmented / Mixed Reality Interactions

Outdoor application examples
TINMITH

- Wearable computer
- Full 3D graphics
- 50cm position
- InertiaCube3
- Pinch gloves
- Tracked thumbs


Outdoor AR Interactions

- Manipulation of objects
  - Two handed
  - Direct Manipulation
  - AR working planes
- Town size walkthroughs
  - Life size
  - ~ 300 metres by 300 meters
Augmented / Mixed Reality Interactions

Through walls collaboration

Through Walls Collaboration

- The Tinmith software is fully distributable across networks
- Connected to a ubiquitous workspace
- Experts indoors <-> Experts in the field
- Initial project supported with ARC Discovery grant

Hand of God


Questions?