A HIMI Model for Collaborative Multi-Touch Multimedia Education

Irene Cheng
University of Alberta
Edmonton, AB
Canada
lin@cs.ualberta.ca

Damien Michel
Institute for Comp Sci
Forth, Crete
Greece
michel@ics.forth.gr

Antonis Argyros
Institute for Comp Sci
Forth, Crete
Greece
argyros@ics.forth.gr

Anup Basu
University of Alberta
Edmonton, AB
Canada
anup@cs.ualberta.ca

ABSTRACT

Educational testing and learning have evolved from using standard True/False, fill-in-the-blank and multiple choice on paper to more visually enriched formats using interactive multimedia content on digital displays. However, traditional educational application interfaces are primarily mouse-driven, which prevents multiple users working simultaneously. Although touch-based displays have emerged and inspired new developments, they are mainly used in simple tasks. In this paper we show how the multi-touch technology can be extended to collaborative learning and testing at a larger scale, using an existing education implementation for illustration. We propose a Human-Intention-Machine-Interpretation (HIMI) model, which applies a graph-based approach to recognize hand gestures and interpret user intentions. Our focus is not to build a new multi-touch system but to make use of the existing multi-touch technology to enhance learning performance. The HIMI model not only facilitates natural interactions using hand movements on simple tasks, but also supports complex collaborative operations. Our contribution lies in embedding the multi-touch technology in multimedia education, providing a multi-user learning and testing environment which would not have been possible using traditional input devices. We formalize a conceptual model to uniquely interpret user intentions via *touch states, state transitions and transition associations*. We also propose a set of hand gestures for working with multimedia educational items. User evaluations are conducted to show the feasibility of the proposed hand gestures.

Categories and Subject Descriptors

K.3.1 [**Computer Milieux**]: Computers and Education – *Computer Uses in Education; Collaborative learning*.

General Terms: Experimentation, Human Factors.

Keywords: Multimedia education, multi-touch display.

1. INTRODUCTION

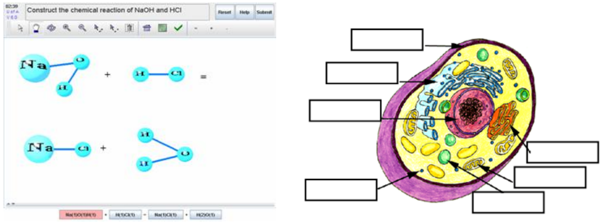
Interactive multimedia content has become increasingly popular in educational applications because it is more effective in conveying abstract concepts than the traditional text and image only formats, such as multiple-choice. Cairncross and Mannion [9] suggested that the key features of multimedia are: user control over the delivery of information and interactivity, which can assist the learners to acquire a deeper understanding of new materials presented. Active involvement in the learning process can also promote internal reflection. There has been significant research on using interactive multimedia for learning and training. The authors in [14] discuss the delivery of Tele-health related material for teaching medical curriculum. An example of how to deal with traumatic head injuries is demonstrated as an example of the system. The Access Grid is used to communicate among students at various locations in New Mexico and Hawaii, USA. Collaborative learning promotes active exchange of ideas between peers in a group. The exchange process increases interest among participants, as well as enhances critical thinking [13]. Johnson and Johnson [15] showed that cooperative teams achieve higher levels of thought and retain information longer than individual learners working quietly on their own. However, traditional user interfaces associated with a single mouse cursor is insufficient to promote collaborative learning and training.

In recent years, touch screens have emerged as an alternative to the mouse input devices. It is believed that allowing users to express themselves using natural hand movements provokes better communication and understanding. Multi-touch technology is incorporated in iPhone, which can respond to one and two-fingered gestures. Perceptual Pixel has created a wall-size multi-touch monitor. Microsoft Surface produces up to 52 points of touch on a tabletop screen [27]. UnMouse Pad supports multiple fingers to write and draw on the pad, creating the corresponding content on a computer screen. Discussion of these commercial successes can be found in [23]. Despite these advances, applying multi-touch technology in educational specific applications, particularly in a collaborative learning and testing environment, has not yet been adequately explored. Furthermore, how to best use multiple fingers and hand movements for multi-touch input is still a research challenge.

In this work we investigate the use of hand gestures on a multi-touch interface in a collaborative environment for multimedia education applications. Educational interfaces [5, 6, 7] often require students to use a mouse device to interact with the multimedia educational items. Only a single user can control the mouse at any given time. Touch-based interfaces not only enable multiple users to work together, but also allow users to express themselves naturally using hand gestures. While a mouse has a limited number of buttons, different finger compositions and movements can present more expressions. We introduce a Human-Intention-Machine-Interpretation (HIMI) multi-touch model to support multimedia education in a collaborative environment. An example of a collaborative chemistry item is given in Fig. 1 (a). To answer this chemistry question, each

AMC'09, October 23, 2009, Beijing, China.

student in the group has to construct a molecule defined in the chemical reaction formula. In a mouse-driven interface, students have to take turns using a single mouse but the HIMI model supports concurrent user operations. In particular, the model can easily execute multiple drag and drop (DND) operations simultaneously. In the example shown in Figure 1 (b), multiple students can touch the correct labels and then slide their hands to the appropriate answer boxes on the screen separately. In a mouse-driven interface, the mouse cursor has to point at the correct label. Then, with a mouse button pressed and dragged to the answer box before the button can be released, and the mouse is passed to the next student. Moreover, the standard usage of the left and right mouse buttons can vary between applications. A user may be inclined to press the left button to drag although the right button is defined for such purpose in the application. The HIMI model recognizes fingers, regardless of which finger or which hand. We define user intentions in an interpretation graph composed of touch states, state transitions and transition associations. This graph is application dependent. Multimedia items can have their own sub-graphs.



(a) (b)

Figure 1. (a) A collaborative chemistry question requires the construction of multiple molecules. (b) A drag and drop question requires the correct label to be placed in the corresponding answer box.

The rest of this paper is organized as follows: Section 2 introduces our Human-Intention-Machine-Interpretation (HIMI) model. Section 3 shows how the HIMI model can be applied to educational multimedia item types. Examples items are quoted from the CROME implementation. Section 4 presents the vision-based multi-touch setup used in our tests. Section 5 proposes a simple set of hand gestures and conducts user evaluations to show their effectiveness. Conclusions and future work is given in Section 6.

2. THE HUMAN INTENTION MACHINE INTERPRETATION (HIMI) MODEL

On a multi-touch display, user intention is communicated through finger and hand movements. The projected images on the display are then analyzed and computer operations are executed accordingly. Hand tracking has been used in augmented reality [18] to inspect virtual objects that can be placed and manipulated on top of a real hand. Template based hand pose recognition was considered in [24]. This method is optimized for several pre-defined hand poses, but may not be suitable for tracking hand movements continuously. The use of multiple fingers and hands for touch screen interaction was studied in [20]. The benefits of single handed vs. two handed interactions in 2D shape creation were analyzed. The single handed interface was found to be faster and more efficient for a limited number of shapes that were tested on a number of subjects. However, the work does not study the advantages and disadvantages of single handed vs. two handed interactions in a broader application domain, such as interactive multimedia in education. Precisely selecting small targets on an interactive panel is a challenging task. This limits the possible contact to fingers instead of the entire hand. The use of "dual finger techniques" is discussed in [4]. The disadvantage with this method is that the speed of human interaction with an application can be slowed down in the process. The difficulty in designing freehand gestures and the tracking of hand movements have been addressed by several researchers [1, 2, 28]. Their work describes using a probabilistic learning framework and skin-colored region modeling for robust tracking.

We consider the needs for a collaborative multi-touch system in educational learning and testing. The goal is to use a set of natural hand gestures that is easy to use and can replace all existing mouse input, as well as easy to memorize. This set of gestures has to be reliably detected on or close to a tactile surface. We choose the 2D image detection approach to avoid the more complex calibration and synchronization process required by the 3D approach. Simple setup that can be handled with minimal technical knowledge is essential because our target users of the collaborative multimedia education applications are students and teachers working at home and in high schools, who do not have sufficient technical knowledge to handle camera networks and multi-view calibrations.

We propose the HIMI model to allow users expressing their intentions naturally and collaboratively using a combination of hand and finger movements. The HIMI model captures human intentions (through a multi-touch screen), validates them in an interpretation graph, and responds by executing computer actions.

2.1 Human Intentions

2.1.1 Touch State, Transition and Association

To execute the example shown in Figure 1 (b) (drag a text label to an answer box) on a multi-touch screen involves touching the label with a finger (Touch), sliding the finger to the answer box (Transition) and removing the finger from the screen (Idle). We use a composition of the following notation to define user intentions.

Notation

ϑ - Idle State (No user intention detected)

Ψ_{id} - Touch State

Ω_{id} - State Transition

Θ_{id} - Transition Association

A *touch state* is uniquely identified by an *id* and five other parameters:

$$\Psi_{id} = (T_{start}, T_{end}, x, y, \Xi)\ , \ (1)$$

where T_{start} and T_{end} record the duration of the touch, and (x, y) is the centroid of the detected contact region; the coordinates in the 2D application interface. Ξ records the contour of the contact region. Each contact region creates an independent touch state, which expires when the contact is removed (Idle state). In other words two unconnected finger images generate two different touch ids. An example is given in Figure 2.

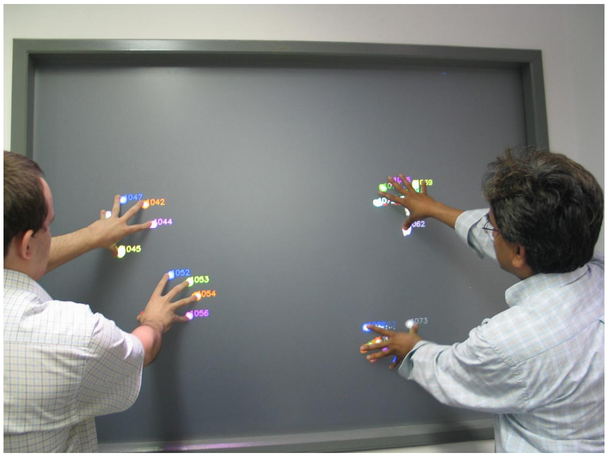


Figure 2. Unconnected finger images are labeled and colored differently indicating different touch states.

A *state transition* defines a temporal or spatial trajectory between touches. A *temporal transition* has to satisfy two conditions: First, the two touches need to be created within a predefined time:

$$|T_i - T_j| < \varepsilon_1, \qquad (2)$$

where T_i and T_j are the end time of first state and the start time of the second state respectively, *i.e.* $T_i < T_j$. Second, the two centroids need to be coincident:

$$\sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} < \varepsilon_2. \qquad (3)$$

ε_1 and ε_2 are threshold values defined by the application. A *spatial trajectory* exists when a finger slides across the screen. The path follows a list of (x, y) coordinates. A state transition can thus be defined by a vector:

$$\Omega_{id} = (\Psi_{id}, x_i, y_i, x_{i+1}, y_{i+1}, \ldots, x_j, y_j), \qquad (4)$$

where i and j denote the start and end points of the path respectively. Each state transition is initiated from a touch state and has to satisfy the constraint:

$$\sqrt{(x_{k+1} - x_k)^2 + (y_{k+1} - y_k)^2} \geq \varepsilon_2, \textit{ i.e. } i \leq k < j. \qquad (5)$$

A *transition association*:

$$\Theta_{id} = \{\Omega_i, \Omega_{i+1} \ldots, \Omega_n\}, \qquad (6)$$

Θ aims to relate unconnected transitions into the same operation.

Given the above definitions of touch state, state transition and transition association, the HIMI model describes the intention of an individual user *i* as:

$$I_i = \bigcup_1^p \Psi_i + \bigcup_1^q \Omega_i + \bigcup_1^r \Theta_i \qquad (8)$$

At a specified time instance *t*, the *s* participating users' collective intention on the multi-touch screen is:

$$\bigcup_1^s I_i \mid t \;\; , \textit{i.e.} \; 1 \leq i \leq s \qquad (8)$$

2.2 Human Intention (HI) Graphs and String

In this section, we discuss how the notation symbols can form a graph to express user intentions. A basic navigation operation, *i.e.* zoom, move and rotate, is often composed of multiple symbols, except "selecting an object" or "selecting a tool", which can be expressed by a single touch symbol. In the HIMI model, each operation on the multi-touch display is recorded as a Human Intention (HI) graph. For example, a user intends to pick up an apple (virtually) and drops it into a basket displayed on the screen. However, he puts a pear into the basket instead and thus needs to make a correction. Furthermore, he withdraws his hand after touching the pear and then touches it again. The HI graph is illustrated in Figure 3 and the sequence of hand movements is described by the HI string:

$$\vartheta\Psi_1\vartheta\Psi_2\Omega_3\Psi_4\vartheta\Psi_5\Omega_6\Psi_7\vartheta\Psi_8\Omega_9\Psi_{10}\vartheta \qquad (9)$$

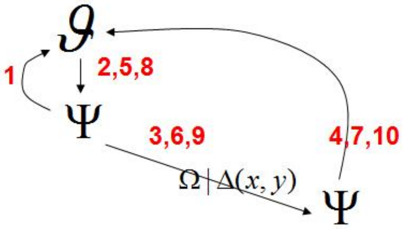


Figure 3. User intention graph example: Six touch states and five state transitions are created.

$\Psi\Omega\Psi$ in the HI string, in between idle states, indicates fetching and dragging to a new location, while Ψ indicates touching (selecting an object or a tool) in an application. The sequence of touches and transitions are numbered along the path in the HI graph. Note that in order to transit from the first touch to a second touch, either spatial ($\Omega \mid \Delta(x, y)$) or temporal ($\Omega \mid \Delta t$) conditions have to be satisfied. These conditions are given in Equations (2), (3), (4) and (5). How the HIMI model recognizes user intentions and rejects undefined gestures by matching the HI graph with the Machine Interpretation (MI) graph will be discussed in Section 2.3.

The design of hand gestures is dictated to a certain extent by the dimension of the display. When operating on a personal iPhone, the tracking surface is limited and probably a maximum of two fingers is optimal. In this paper, we focus on a larger wall mounted type of display which can support multiple user collaboration. The HIMI model allows an application to define its Machine Interpretation graph appropriate for the display dimension.

2.3 Machine Interpretations Graph

A Machine Interpretation (MI) graph is composed of touch states, state transitions and transition associations. The graph is application defined depending on specific requirements. The MI graph used in this paper is shown in Figure 4. Note that a HI string can be generated by traversing the MI graph. A continuous path between idle states is recognized as a valid operation. Paths not defined in the MI graph are invalid user instructions and cannot be executed by the application. Dotted lines are used to indicate that a different finger is needed in conjunction with the first one (solid lines). Constraints on time duration and location proximity are denoted by Δt and $\Delta(x, y)$ respectively. For example, if two consecutive touches satisfy both constraints $\Delta(x, y)$ and Δt, an association Θ is established. Otherwise, the most recent touch state Ψ prevails.

When hand movements are detected, a HI graph is recorded. A valid HI graph should be a subset of the application defined MI graph. The HIMI model validates user intention in two stages. In the first stage, the HI string is parsed to examine whether a correct combination of symbols appear between idle states. In the second stage, the MI graph is traversed to find a matching sub-graph for the HI graph. If a match is found, the user instruction is executed. The full set of proposed gestures is described in Appendix A, and will be discussed further in Section 5.

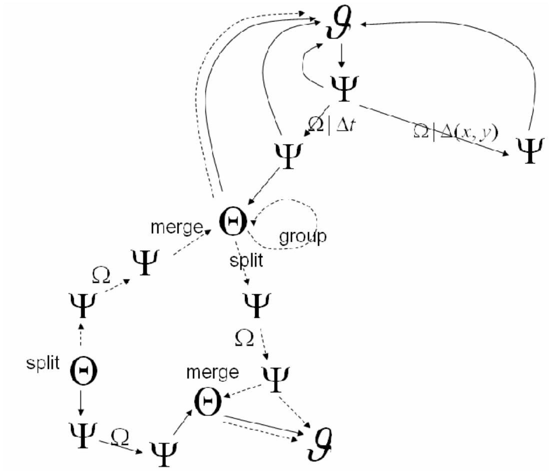


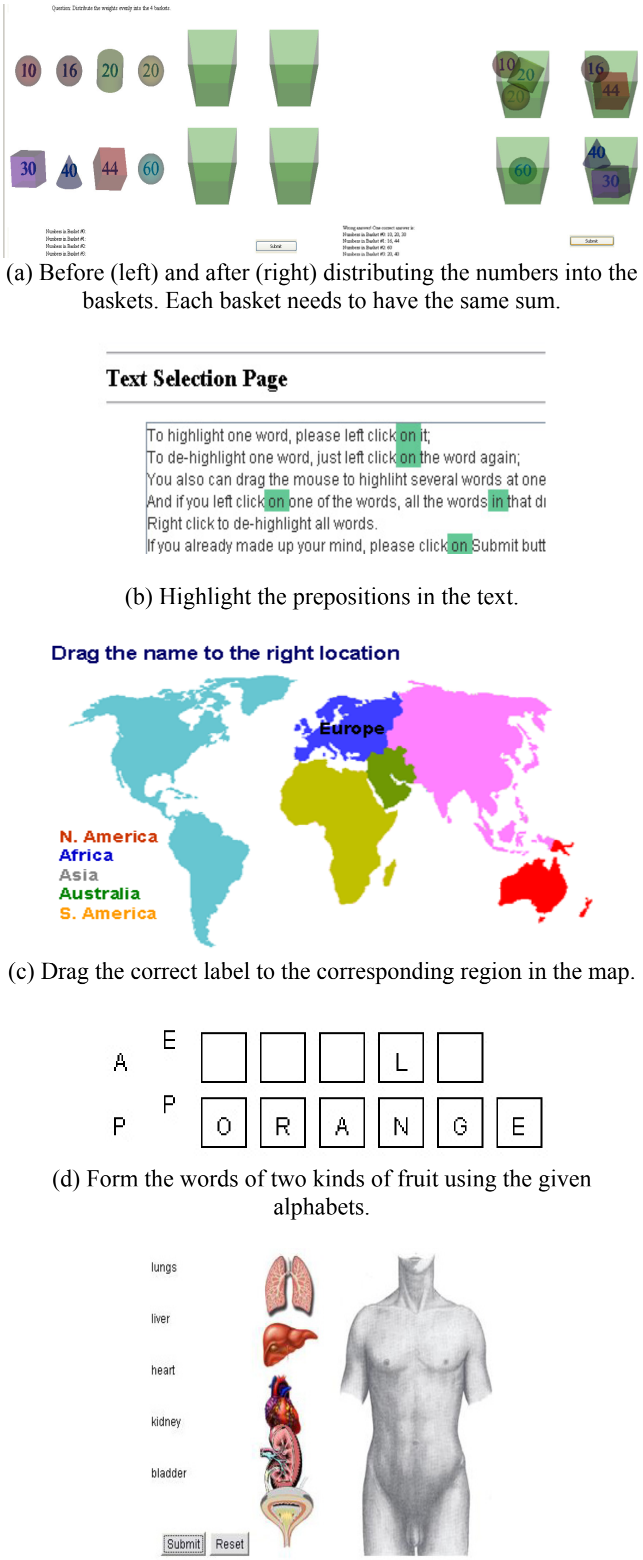
Figure 4: **An application Machine Interpretation (MI) graph.**

In the next section, we will give item examples to illustrate how the HIMI multi-touch model and the MI graph are used in a multimedia education application.

3. COLLABORATIVE MULTI-TOUCH IN MULTIMEDIA EDUCATIONAL APPLICATION

Computer-based education provides flexibility in the design of question items. Instead of the conventional items, such as multiple-choice, true/false and fill-in-the-blank, multimedia items using audio, video, graphics and animation, are used for more effective learning and testing [3, 22]. Multimedia items can convey subject knowledge and abstract concepts more effectively [5, 6, 7]. However, these items are designed mainly for a single user using a computer mouse as the input device. In order to advance multimedia education applications to support interpersonal communication, peer assistance, project and time management, we incorporate the multi-touch technology. We use an existing education implementation to show how the HIMI model works in an interactive and collaboration context, in particular for multimedia education applications.

Figure 5 shows six single-user multimedia items taken from an education application initially designed for single users and mouse input. Each item is categorized by an item type. Items generated from the same item type have similar screen layouts. Item types are template-based data structures used for creating question items. A simple item type is a multiple choice template which is populated with a question text, four possible choices and a correct answer. In contrast, innovative multimedia item types have more visually enhanced designs. Figure 5 (a) to (f) are examples of items generated from six item types defining very different layouts.


Question: Distribute the weights evenly into the 4 baskets.

Numbers in Basket #0:
Numbers in Basket #1:
Numbers in Basket #2:
Numbers in Basket #3:

Wrong answer! One correct answer is:
Numbers in Basket #0: 10, 20, 30
Numbers in Basket #1: 16, 44
Numbers in Basket #2: 60
Numbers in Basket #3: 20, 40

(a) Before (left) and after (right) distributing the numbers into the baskets. Each basket needs to have the same sum.

Text Selection Page

To highlight one word, please left click on it;
To de-highlight one word, just left click on the word again;
You also can drag the mouse to highliht several words at one
And if you left click on one of the words, all the words in that di
Right click to de-highlight all words.
If you already made up your mind, please click on Submit butt

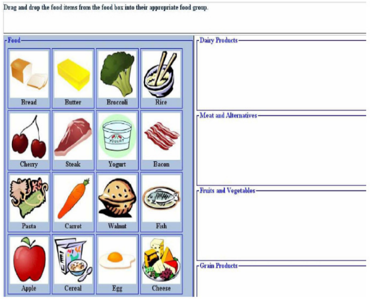
(b) Highlight the prepositions in the text.

Drag the name to the right location

(c) Drag the correct label to the corresponding region in the map.

(d) Form the words of two kinds of fruit using the given alphabets.

(e) Place the organs at the correct locations of the body.


Drag and drop the food items from the food box into their appropriate food group.

(f) Classify the food into the correct categories.

Figure 5: **Examples of multimedia educational items.**

Note that except Figure 5 (b), the other items involve drag-and-drop (DND). The DND operation can be applied in a variety of subject areas, including Math (Figure 5(a) – distribute the numbers into the baskets so that the sum in each basket is equal), English (Figure 5(d) – spell the correct words), Geography (Figure 5(c) – label the regions), Biology (Figure 5(e) – place the organs in correct locations), and Health Science (Figure 5(f) – categorize the given items). Figure 5(b) also involves DND but the path of the dragging is highlighted. When using a mouse-driven interface, pressing a mouse button at an object selects it. After dragging the mouse to the correct location and releasing the button drops the object. When executing DND operations on a multi-touch screen, multiple users can perform multiple DND operations concurrently using their fingers. Since each trajectory is continuous, the system is able to identify individual finger paths and execute the instructions without confusion. DND is a sub-path in the MI graph as shown in Figure 6. Therefore, the HI string ($\Psi\Omega\Psi$) and HI graph can be verified as a valid user intention by the HIMI model.

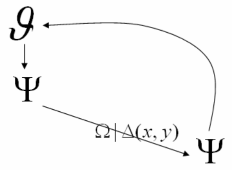
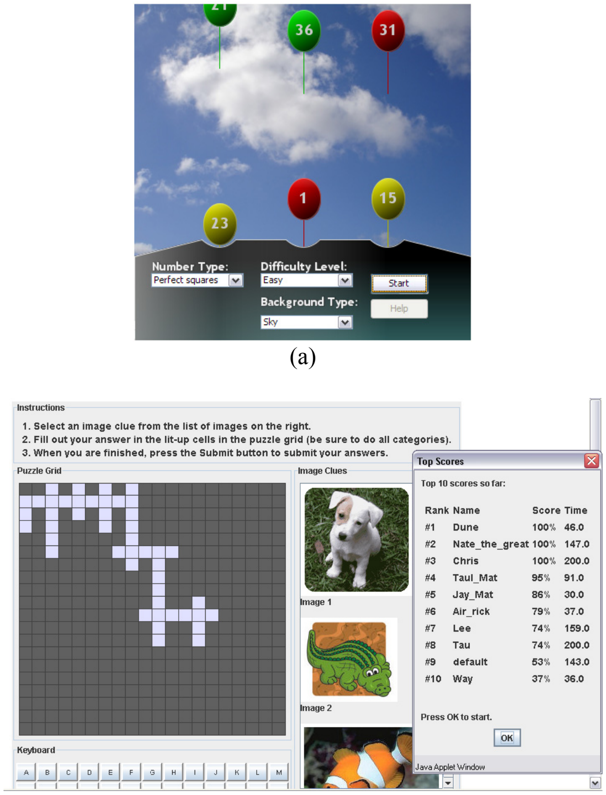


Figure 6: **The MI graph shows that DND is a valid path.**

In education research, it is believed that educational games possess an engaging and rewarding factor, which can inspire students to stay in the game and thus learn. There are an increasing number of multimedia question items presented in a game context. The balloon shooting game (Figure 7(a)) is designed to test a player's understanding about prime numbers, multiples, etc. The ascending balloons contain randomly generated numbers, some of which are the targets. By increasing the speed of the ascending balloons, the game becomes more challenging and is used to test how fast the player can respond to the visual stimuli in addition to Math knowledge. We invited 20 high school students to play the prime number game with balloons generated at a rate of one balloon in every two seconds, and to use prime numbers less than the numeric 100. Among the 20 participants only one student scored 100%. The main challenge is not because they could not identify the prime. It is because they could not respond fast enough to the three streams of simultaneously ascending balloons. When presenting this game on a collaborative multi-touch screen, we can test how well team members collaborate. A team can choose a suitable strategy based on the members' skills, *e.g.* assign different balloon streams or number ranges to different members. The team which devises the best strategy will get the highest score. The HI string of this item type is a sequence of single touches ($\vartheta\Psi\,\vartheta\Psi\ldots\Psi\,\vartheta$). It can be seen that a single touch is a valid path in the MI graph.


Number Type: Perfect squares

Difficulty Level: Easy

Background Type: Sky

(a)

Instructions

1. Select an image clue from the list of images on the right.
2. Fill out your answer in the lit-up cells in the puzzle grid (be sure to do all categories).
3. When you are finished, press the Submit button to submit your answers.

Top Scores

Top 10 scores so far:

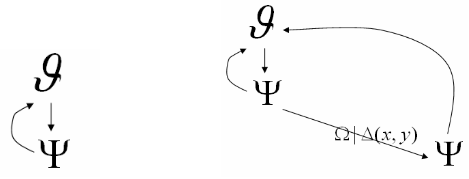
Rank	Name	Score	Time
#1	Dune	100%	46.0
#2	Nate_the_great	100%	147.0
#3	Chris	100%	200.0
#4	Taul_Mat	95%	91.0
#5	Jay_Mat	86%	30.0
#6	Air_rick	79%	37.0
#7	Lee	74%	159.0
#8	Tau	74%	200.0
#9	default	53%	143.0
#10	Way	37%	36.0

Press OK to start.

(b)

Figure 7: **Examples of educational game items.**

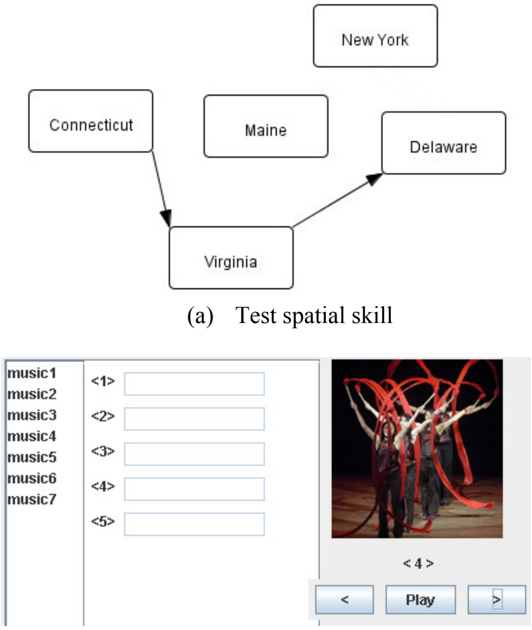
Instead of balloon shooting, vocabulary can also be learned and tested in a collaborative environment. Multiple students can participate to solve a word puzzle together (Figure 7(b)). By touching the picture, the corresponding blanks will be highlighted. Students can then drag letters with their fingers from the soft keyboard. Note that each item type can define its own valid paths. Figure 8 (a) and (b) show the valid paths permitted in the item types illustrated in Figure 7 (a) and (b) respectively. Different from DND, Figure 8 (a) only permits touch, *i.e.* shooting a target balloon. Figure 8 (b) permits both touch and DND.



(a)

(b)

Figure 8: **Valid paths for (a) the balloon shooting game, and (b) the word puzzle.**



(a) Test spatial skill

(b) Test musical skill

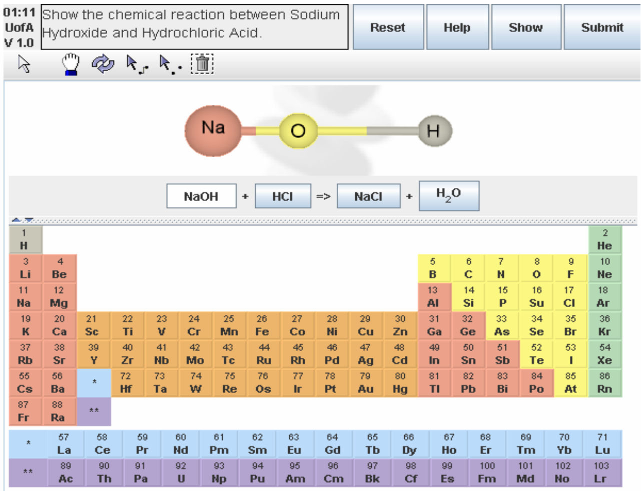
Figure 9: **Testing cognitive intelligence using multimedia items.**

Gardner defines the seven human intelligences [11, 12], namely: Linguistic, Logical-Mathematical, Spatial, Bodily-Kinesthetic, Musical, Interpersonal and Intrapersonal. It is impossible to test all these cognitive skills using simple multiple choice and true/false format. Multimedia educational items, on the other hand, can involve using special cognitive skills [8]. For example, in Figure 9 (a), the students need to order the cities from north to south by drawing arrows. Since the city boxes are moving randomly at high speed and occluding each other from time to time on the screen, good *visual-spatial* and hand control coordination is necessary. Multiple students can work together and draw arrows between the cities they know about.

Figure 9 (b) shows how the ability to perceive, discriminate, express, and transform musical forms (*musical intelligence*) can be tested. In this example, there is a sequence of silent dancing videos (<1> to <5>) displayed on the right. Music clips extracted from these videos together with some unrelated sound tracks can be heard by touching the audio icons displayed on the left. Students can listen to the sound tracks and associate the corresponding music rhythm with the body gestures in a dance. While one student scrolls through the video clips, another student can switch between music clips.

3.1 Extend beyond Simple Touch and Dragging

It is interesting to note that the execution of the item shown in Figure 9 (a) requires some new gestures other than touch and DND. Since DND is used to relocate an object, the same hand movement cannot be used to link two objects by drawing an arrow. Displaying command icons in the tool bar is common in many interfaces. However, in a multi-touch environment, a touch state vanishes as soon as the contact is removed. In other words dual fingers are necessary. While one finger touches an object and another finger touches a tool icon, the question is how to associate these independent gestures so that the system can recognize that they constitute a single operation.



Show the chemical reaction between Sodium Hydroxide and Hydrochloric Acid.

NaOH + HCl => NaCl + H_2O

1 H																	2 He
3 Li	4 Be											5 B	6 C	7 N	8 O	9 F	10 Ne
11 Na	12 Mg											13 Al	14 Si	15 P	16 Su	17 Cl	18 Ar
19 K	20 Ca	21 Sc	22 Ti	23 V	24 Cr	25 Mn	26 Fe	27 Co	28 Ni	29 Cu	30 Zn	31 Ga	32 Ge	33 As	34 Se	35 Br	36 Kr
37 Rb	38 Sr	39 Y	40 Zr	41 Nb	42 Mo	43 Tc	44 Ru	45 Rh	46 Pd	47 Ag	48 Cd	49 In	50 Sn	51 Sb	52 Te	53 I	54 Xe
55 Cs	56 Ba	*	72 Hf	73 Ta	74 W	75 Re	76 Os	77 Ir	78 Pt	79 Au	80 Hg	81 Tl	82 Pb	83 Bi	84 Po	85 At	86 Rn
87 Fr	88 Ra	**															
*	57 La	58 Ce	59 Pr	60 Nd	61 Pm	62 Sm	63 Eu	64 Gd	65 Tb	66 Dy	67 Ho	68 Er	69 Tm	70 Yb	71 Lu		
**	89 Ac	90 Th	91 Pa	92 U	93 Np	94 Pu	95 Am	96 Cm	97 Bk	98 Cf	99 Es	100 Fm	101 Md	102 No	103 Lr		

Figure 10: **3D visualization in chemical molecules construction questions.**

In the HIMI model, this issue is addressed by the notion of association using *anchor* to identify the main task, which is executed with some sub-tasks like *split, merge* and *group* to complete one single operation. Figure 10 shows an example which supports a number of commands displayed in the tool bar. This chemistry item requires students to build four molecules defined in the reaction equation:

$$NaOH + HCL = NaCl + H_2O \qquad (10)$$

An atom will appear on the canvas when the corresponding atom symbol in the Periodic table is touched. The next step is for each student to drag his or her atoms to the appropriate position in the equation. When there is only one user, putting a bond between two atoms is simply done by touching the “link” icon, followed by touching the two atoms. However, when multiple users touch separate atoms simultaneously, the system is not able to pair touches correctly. In the HIMI model, we use double touches to indicate anchoring an object. While this finger continues to touch one atom, another finger touches the “link” icon and drags it to the second atom, followed by sliding the finger back to the anchor (Figure 11). The merging of the two fingers advises the system that these movements are associated. When a bond already exists between the two atoms, using the same gesture but touching the “unlink” icon instead of the “link” icon will disconnect the bond. A similar association operation using “link” is used to draw an arrow in the item shown in Figure 9 (a).

The capability of visualizing and interacting with 3D objects is one of the reasons why multimedia educational items are so appealing. 3D graphics and animation are particularly useful in conveying the concepts of atomic bonds and molecular structures. Translation, rotation and scaling are valid paths in the MI graph. A more detailed description of the corresponding gestures is given in Appendix A. In the next section, we describe the multi-touch system setup used in our experiments.

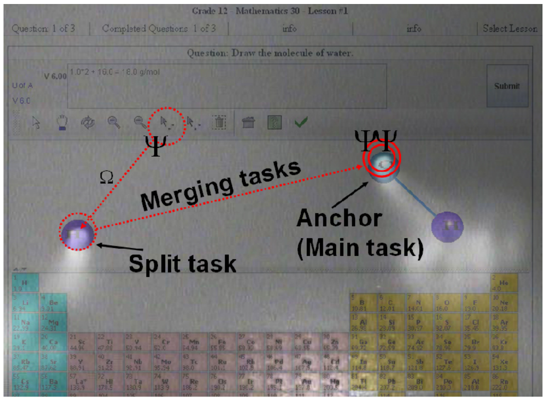


Figure 11. An illustration of using two fingers to complete an association operation. The bright blob on the right (first finger) executes the main task (anchor) and the left blob (second finger) executes the split task, which is then merged back with the anchor.

4. A VISION-BASED MULTI-TOUCH SMARTBOARD SETUP

4.1 Multi-Touch System

Multi-touch detection can be vision-based or sensor-based [16, 17, 19, 25]. Sensor-based equipment is more costly and requires a longer manufacturing time. Our goal is to employ low-cost technology in education. Instead of setting up a sensor-based system with a large number of wires and expensive circuitry, we focus on vision-based techniques, using off-the-shelf equipment such as video camera and projector, in order to provide an affordable, scalable and easy to setup system that is portable and accessible to the general public. Multiple cameras can be used to capture multiple viewpoints and can detect more variations of hand gestures [21, 26] without the users touching the screen. However, such setup requires very precise calibration which cannot be achieved without sufficient technical skills. We leave 3D hand tracking for a separate paper and focus on the recognition of projected 2D hand gestures using a single camera in this work.

In our experiment, we use the setup described in [10] and the configuration is shown in Figure 12. The hardware includes a grey level camera equipped with a low pass infrared filter, a video projector and two infrared illuminators, an acrylic screen backed by a layer of glass. The image tracking software is operated on a PC. The infrared illuminators emit light which passes through the wall-mounted screen and is reflected by the object on the screen surface, *i.e.* hand and fingers. The projected hand image is acquired by the camera, while the infrared filter blocks the visible spectrum – the display of the application interface. Figure 13 (top) shows the original question and Figure 13 (bottom) shows the paths of multiple DND trajectories relocating the alphabets forming the words "apple" and "orange." Figure 14 shows that a contact region can be created by dual connected fingers.

From Figure 11, it can be seen that depending on the lighting conditions in the surrounding, the illumination on the display may not be evenly distributed. To address this issue, we use an online photometric calibration algorithm which is based on the one reported in [10].

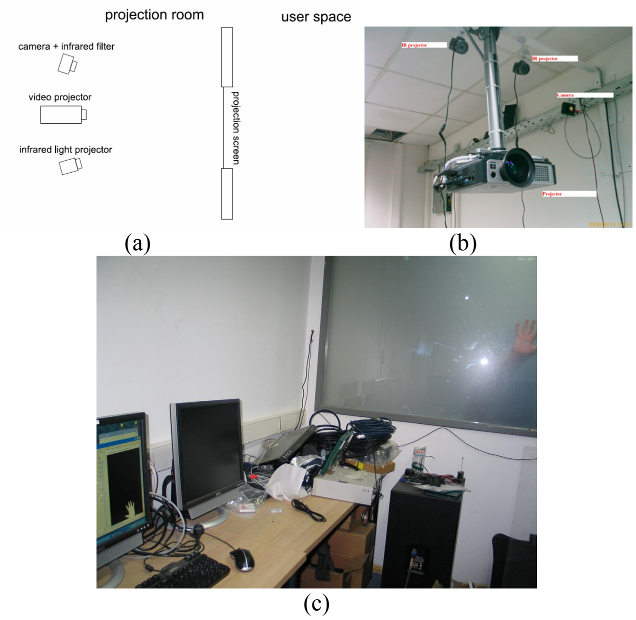


Figure 12. Multi-touch system used in our experiments. (a & b) System setup; (c) An illustration of a user hand touching the screen (top right) and the corresponding extracted image on the computer screen (bottom left).

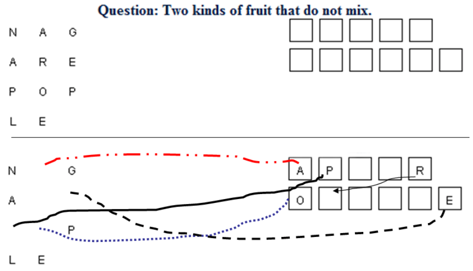


Figure 13: **An example of concurrent DND operations executed by multiple users.**

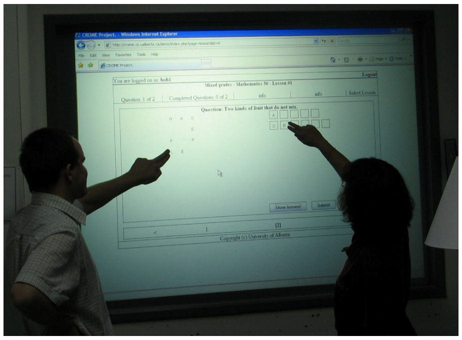


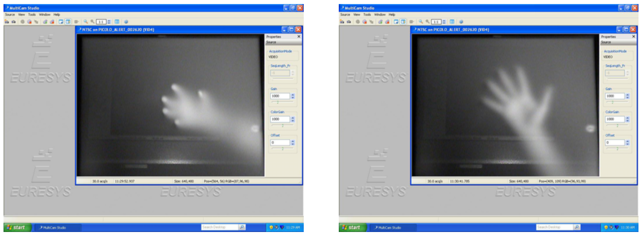
Figure 14: **Two users work on a DND item. Note that two connected fingers are recognized as a single touch state.**

More specifically, the entire projection screen is covered with a 5x5 grid of locations. At each location (i, j), a rectangular area small enough to be covered by a hand is displayed. Two camera responses are measured: $\Re_{ij}^{0}$ when no hand is placed at the location, and $\Re_{ij}^{1}$ when the user's hand is there. The next step of the calibration consists in performing a 2D interpolation of these responses for each individual pixel. During runtime, the value of a pixel P_{xy} is adjusted according to:

$$P^{c}_{xy} = (P_{xy} - \Re^{0}_{xy}) / \Re^{1}_{xy} \quad (11)$$

Then, high pass filtering is applied to the resulting image. This allows to get rid of the shadows of the objects that are close to the screen but do not touch it. Figure 15 shows the difference in the response in the case where fingers touch the screen (Fig.15(a)) and in the case where a hand is close to the screen without touching it (Fig.15(b)). Finally a threshold $\Gamma = 1 - s$ is applied to get a binary response of the presence of objects, where s is a sensitivity factor in the range [0..1]. The differentiation between touching hands and fingers can be achieved by looking at the size and, possibly, the contours of the corresponding blobs.

This calibration process is fast, effective and can easily be carried out by non-technicians.



(a) (b)

Figure 15: The difference in camera response in the cases of touching fingers and a non-touching hand.

5. HAND GESTURES AND EVALUATION

We propose a set of gestures using both fingers and hand movements to increase the number of expressions. When the hand is too close to or resting on the screen, the projected image can contain not only the finger tip but also a part of the hand (Figure 11). Part of this problem is resolved through the high pass filtering operation described in section 4.1. Additionally, such situations could be detected by analyzing the contour parameter Ξ recorded in each touch state Ψ. If it is a finger, only the finger tip is considered when computing the centroid. In the HIMI model, hand gestures are used in a global context. For example, when translating, rotating or scaling the entire canvas, hand is used instead of fingers. When manipulating objects on the canvas, fingers are used (Appendix A).

We have analyzed the HIMI model using fifty multimedia educational item types and find that the proposed set of gestures is sufficient to express the necessary user intentions. In order to test the effectiveness of the proposed gestures, we conducted a user evaluation by inviting a group of volunteers. The goals of the experiments were (i) to find out whether the gestures are easy to learn for first time users without intense memorization, and (ii) whether the users feel the gestures are natural in expressing their intentions when working with multimedia educational items. We chose six subjects who were good at using mouse-driven interfaces but have never worked on touch screens before. They were all first time users of hand gestures. The test was given to each subject separately. Before the test, four minutes tutorial was given to explain how the gestures work (Appendix A). They were asked to solve seven educational items, which require them to apply different gestures. Multiple gestures may be needed in a single item. The subject had to respond immediately. (S)he had to say "forget" if (s)he could not remember the correct gesture and moved onto the next item. The subject was not allowed to refer to any written note when answering the questions. The time required to answer each item was in the range of [2…5] seconds, with an average of 3 seconds. We summarize the findings on each gesture in Table 1.

Table 1: **User evaluations of using hand gestures in solving multimedia educational items.**

Responses from subjects	Correct
Touch an object and relocate it	6/6
Rotate	5/6
Apply a tool bar command to two objects, e.g. Link,	8/12
Zoom	6/6
Apply a tool bar command to an object, e.g. delete	6/6
Group objects	5/6

It is interesting to note that four mistakes were made by the same subject, who forgot to double touch when anchoring an object (Rotate, Link/Unlink and Group). The remaining two mistakes were made by another subject, who forgot the "unlink" gesture in one item and did not touch the linked object in another item. When the system is properly implemented, a user will notice that the link does not appear if (s)he misses the consecutive touches (anchor). In a mouse-driven interface, it is also possible that double clicks are not detected by the system and the user has to repeat the clicks. Based on this preliminary experiment, we can say that the proposed gestures are easy to use. The subjects also feel that these are natural gestures to express user intentions.

6. CONCLUSIONS AND FUTURE WORK

The goal of this paper was not to propose a new multi-touch system but to adopt multi-touch technology in educational learning and testing. In particular, we wanted to advance multimedia items to support multi-user collaboration. By inspiring more studies in this topic, the ultimate goal is to establish a set of universally adopted hand gestures that is easy to use and is effective in expressing user intentions. This is especially important for the collaborative multimedia educational items discussed in this paper. Collaboration can be online through the Internet where participants can be from diverse geographical locations. If different set of gestures are employed, the application has to implement multiple gesture interpretation algorithms.

We introduced a HIMI model together with a set of proposed gestures. The model applied a graph-based approach to detect valid hand gestures and express user intentions by using only three basic notations: touch state, state transition and transition association. The Machine Interpretation graph was application specific and had the flexibility of allowing individual item types to define their sub-graphs. The model was tested using an existing education system and proved to be adequate to handle all the innovative multimedia items defined in the system. User evaluations showed the effectiveness of the proposed gestures. In future work, we will test the model in a remote multi-user collaborative environment.

7. ACKNOWLEDGMENTS

This project is partially funded by Informatics of Research Excellence (iCORE) Alberta and Natural Sciences and Engineering Research Council (NSERC) Canada. Additionally, it has been supported by the FORTH-ICS internal RTD programme 'Ambient Intelligence and Smart Environments'.

8. REFERENCES

[1] A.A. Argyros and M.I.A. Lourakis, "Real time Tracking of Multiple Skin-Colored Objects with a Possibly Moving Camera," in Proceedings of the 2004 European Conference on Computer Vision (ECCV'04), Springer-Verlag, vol. 3, pp. 368-379, May 11-14, 2004, Prague, Czech Republic.

[2] A.A. Argyros and M.I.A. Lourakis, "Vision-Based Interpretation of Hand Gestures for Remote Control of a Computer Mouse," in Proceedings of the 2006 European Conference on Computer Vision, LNCS Springer-Verlag, pp. 40-51, May, 2006, Graz, Austria.

[3] R. Allen, "The Web: Interactive and Multimedia Education," Computer Networks and ISDN Systems, Vol.30, pp1717-1727, 1998.

[4] H. Benko, A.D. Wilson and P. Baudisch, "Precise selection techniques for multi-touch screens," ACM CHI, Montreal, Canada, April 2006.

[5] I. Cheng and A. Basu, "An effective multimedia item shell design for individualized education: The CROME project," Advances in Multimedia, Vol. 2008 Article ID 825671, 10 pages.

[6] I. Cheng and A. Basu, "Graphics based computer adaptive testing and beyond," EUROGRAPHICS Education Track 2008, 8 pages.

[7] I. Cheng, A. Basu and R. Goebel, "Interactive multimedia for adaptive online education," IEEE Multimedia Magazine 2008 (in press).

[8] I. Cheng, C. Kerr and W. Bischof, "Assessing Rhythm Recognition Skills in a Multimedia Environment," IEEE Int. Conf. on Multimedia, 4 pages, 2008.

[9] S. Cairncross and M. Mannion, "Interactive Multimedia and Learning: realizing the Benefits," Innovations in Education and Teaching International (IETI) 2001, 38(2), ISSN 1470-3300, Taylor & Francis Ltd.

[10] D. Michel, A. Argyros, D. Grammenos, X. Zabulis and T. Sarmis, "building a Multi-Touch Display Based on Computer Vision Techniques," IAPR Conf. on Machine vision Applications (MVA) 2009, Yokohama, Japan.

[11] H. Gardner, "Frames of Mind: The Theory of Multiple Intelligences," New York: Basic Books, 1983.

[12] H. Gardner, "Artistic Intelligences", Art Education, Vol. 36, 1983, pp. 47-49.

[13] A.A. Gokhale, "Collaborative Learning Enhances Critical Thinking," Journal of Technology Education, Vol. 7 No. 1, 1995, pp22-30.

[14] J. Jacobs et al., "Integration of Advanced Technologies to Enhance Problem-based Learning over Distance: Project Touch," The Anatomical Record (Part B: New Anat.) 270 B:16–22, 2003.

[15] R.T. Johnson and D.W. Johnson, "Action Research: Cooperative Learning in the Science Classroom," Science and Children, 24, 31-32.

[16] W.D. Hillis, "A High resolution imaging touch sensor," Int. J. of Robotics Research, 1982, 1, 2, 33-44.

[17] S. Lee, W. Buxton and K.C. Smith, "A multi-touch three dimensional touch-sensitive tablet," Proc. Of the SIGCHI Conf. on Human Factors in Computing Systems, 1985, pp21-25, San Francisco, USA.

[18] T. Lee and T. Hllerer, "Handy AR: Markerless inspection of augmented reality objects using fingertip tracking," IEEE Int. Sym. On Wearable Computers (ISWC) 2007.

[19] K. Kamiyama, K. Vlack, T. Mizota, H. Kajimoto, N. Kawakami and S. Tachi, "Vision-based sensor for real-time measuring of surface traction fields," IEEE Computer Graphics and Applications 2005, Vol. 25, No. 1, pp68-75.

[20] T. Moscovich and J.F. Hughes, "Indirect mappings of multi-touch input using one and two hands," ACM CHI, Florence, Italy, April 2008.

[21] S. Malik and J. Laszlo, "Visual touchpad: A two-handed gestural input device," Int. Conf. on Multimodal Interfaces (ICMI) 2004, pp289-296, New York, USA.

[22] C.G. Parshall, T. Davey and P.J. Pashley, "Innovative Item Types for Computerized Testing," Computerized Adaptive Testing: Theory and Practice. W. van der Linden & C. Glas (Editors), pp129-148, 2000.

[23] T. Selker, "Touching the Future," Communication of the ACM Magazine, pp14-16, Vol. 51, No. 12, December 2008.

[24] B. Stenger, "Template based hand pose recognition using multiple cues," Proceedings Asian Conference on Computer Vision, LNCS 3852, 551-560, 2006.

[25] Tactex multi-touch interfaces http://www.tactex.com (last visited Oct 2008).

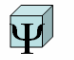
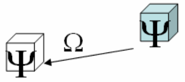
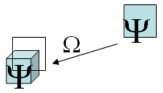
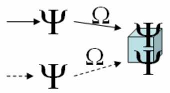
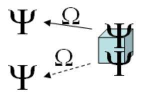
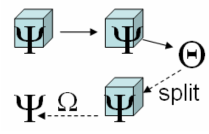
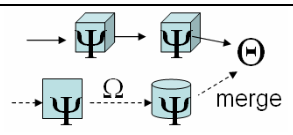
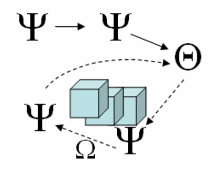
[26] A.D. Wilson, "TouchLight: An imaging touch screen and display for gesture-based interaction," Int. Conf. on Multimodal Interfaces (ICMI) 2004, pp69-76, New York, USA.

[27] A.D. Wilson, "PlayAnywhere: A Compact Interactive Tabletop Projection-Vision System," In Proc. of the UIST'05, Seattle, USA, Oct. 23-26, 2005.

[28] M. Wu et al., "Gesture registration, relaxation and reuse for multi-point direct-touch surfaces," TR 2005-19, MERL, Oct. 2005.

Appendix A

Table A1: A proposed set of gestures

User Intentions / Gestures	**MI Graph Description (Idle state is omitted)** Cube and Cylinder – objects Square – icon Solid and Dotted lines indicate different fingers
Select an object or a tool bar command icon, e.g. undo. Touch with a finger.	
Relocate an object (Drag-and-Drop) Touch the object with a finger, drag it to the desired location and remove the finger.	
Apply a tool bar command icon to an object, e.g. delete Touch the icon with a finger, drag it to the object and remove the finger.	
Shrink an object Put two fingers outside the object (blank canvas area) and slide them towards each other until they meet inside the object.	
Enlarge an object Put two fingers inside the object and slide them away from each other.	
Anchor an object (This is always associated with a sub-task) Use one finger to double touch the object to be anchored. While keeping the finger at the anchor, use another finger to perform the sub-task (Refer rotate, zoom, group, link and unlink).	
Rotate an object (in 2D or 3D as defined by the item type) Use one finger to anchor the object. Use another finger to touch the anchored finger and then split the second finger in the rotation direction. The second finger can repeat the process as long as the first finger is anchored.	
Apply the tool bar command icons, e.g. link and unlink, to two objects Use one finger to anchor one object. Touch the icon with another finger. Slide the second finger to the second object and then merge the second finger with the anchored finger. **When more than one object is touched before merging with the anchor while dragging the icon, the most recent object is the one to be recognized.**	
Apply a directional link between two objects Same as link but the anchored object will be the one being pointed to by the arrow. The user has to choose the correct object to anchor.	
Group unconnected or connected objects Anchor at a blank space using one finger. Use another finger to split from the anchor, draw a path to enclose the required objects and merge the second finger with the anchor. **By keeping the anchored finger in place, the second finger can perform a sub-task. In this case, the sub-task will apply to the group.**	
Rotate, Zoom and Translate the entire canvas Same as applying rotate, zoom and translate to an object but use hands instead of fingers.	