

RemoAct: Portable Projected Interface with Hand Gesture Interaction

Hussam Saad Adeen, Ayman Atia, Ahmad Amin, Andrew Victor, Abdelrahman Essam, Ehab Gharib and Mohamed Hussien

Human Computer Interaction Lab., Computer Science,
Faculty of Computers and Information, Helwan University, Cairo, Egypt

Article history

Received: 05-11-2014

Revised: 29-06-2015

Accepted: 24-07-2015

Corresponding Author:

Ayman Atia

Human Computer Interaction

Lab., Computer Science,

Faculty of Computers and

Information, Helwan

University, Cairo, Egypt

Email: ayman@fci.helwan.edu.eg

Abstract: RemoAct is a wearable depth sensing and projection system that enables interaction on many surfaces. It makes interaction with the environment more intuitive through sharing and sending data with surrounding devices by applying certain gestures. This system offers a mobile and intuitive solution for interacting using a projected surface on habitual flat surfaces. Every user has their public and private areas, where the user can create tiles on the fly and share it with others and these public tiles are shown to other users through augmented reality. Interaction is made through hand gestures, finger tracking and hand tracking. This gives the user more freedom in movement. Different experiments were conducted to calculate the accuracy and RemoAct ran against different conventional methods to compare its accuracy, time and user experience. RemoAct takes less time for two users to draw one chart. As the system enables the users to work simultaneously, it reduces the needed time, short compared to successive drawing. For gesture recognition, accuracy reached 90-95%. Object recognition and face identification accuracy varied with the variation of light.

Keywords: Hand Gesture, Finger Tracking, Face Identification, Object Recognition, Gesture Recognition, Portable Interface

Introduction

Recent advancement in mobile touch displays has made people familiar with new ways of dealing with computational devices. However, dealing with a touch display is now considered to be conventional. Further, it is inconvenient to hold a device of a size of a hand palm or to deal with a small device with tiny menus and icons. Sharing views of visual data and images with others is restricted by the mobile device small display.

Interacting, sharing and sending data with other people has been restricted to using the device itself rather than a sort of free-device environment. Recent research in Human Computer Interaction (HCI) has opened a new vision for dealing with computational devices so that the user does not have to deal with the device itself. Rather, users can transform everyday surfaces into interactive interfaces to interact with their mobile devices. Sixth Sense uses markers that are worn on the finger in order to interact with the system (Mistry *et al.*, 2009). New projection technologies have introduced portable projectors used to project displays on everyday surfaces commonly used. It can transfer the

handheld device display to almost any surface the user deal with in an everyday environment. Portable projectors offers great flexibility to project displays on walls, tabletops, hand-held objects, or the user's body. Thus offers a great flexibility to control the size of display to fit different environments and offers better sharing views of data and images between multiple users. New ways of interaction have been implemented opening the way to more intuitive and easier to learn interaction techniques. The user can interact with our everyday environment like walls, tabletops and different held objects to be used as interfaces. However, the adaptation of this new interaction technique implies using particular hardware and sensing devices.

According to Vision-Based Hand-Gesture Applications (Wachs *et al.*, 2011), Hand gestures are useful for computer interaction as they are the most primary and expressive form of human interaction. People communicate normally using hand gestures, finger pointing and body movement. Hand gestures interaction could yet become more important in different applications as projected interface applications due to their ease of access and naturalness of control.

Further, Projected interface applications users usually do not hold a handheld device where the hand-gesture use would be the most expressive and comfortable way to interact.

Augmented Reality (AR) is a variation of Virtual Environments, or Virtual Reality as it is more commonly called as referred in A Survey in Augmented Reality (Azuma, 1997). It is used to improve the user's real environment by adding more information. This augmented information gives the user more utility to reduce risk, if it is used to alert or simulate some situations. The added information can also reduce costs, such as the cost reduced in Virtual Reality and Augmented Reality in Digestive Surgery (Soler *et al.*, 2004). AR is used, as well, to add more intuitive user experience. We use it to assign every user to their public area. Users facing each other will be able to see each other's public area augmented beside them.

In this study, we explore a shoulder-worn system that provides a new interaction technique with different devices using a projected display easily projected on different surfaces, such as hand, tabletop or walls and thus fitting into different environments. The system uses a depth camera to capture hand and finger gestures and movements to deliver the same touch experience as dealing with conventional touch display. Through hand gestures, the user can interact with projected display easily. The system enables merging data of two users where each user can work simultaneously on the merged data and offering better data visualization and sharing.

The system contributes a novel implementation of new projected interface techniques where a user's projected display is split into a number of areas, each can be set to either private or public. These areas contain tiles where a tile is a space generated by the user to place their data in. A tile assigned by the user in the public area can be seen by anyone, whereas a tile assigned in private area can only be seen by that user. The system also introduces Augmented Reality applications that can be implemented and used by the system.

Our system, named RemoAct, is a wearable system that uses a projection and capturing devices and that delivers better interacting experience with different devices. Through this system, users can smoothly deal with a projected display on everyday surfaces as well as to share and send data to different objects and people more easily. Furthermore, it uses AR techniques for delivering better experience in dealing with real-world objects.

Related Work

Research related to the RemoAct can be divided primarily into four main categories: Interactive Projected Interface, Augmented Reality, interaction by gestures and Communication through Distance.

Interactive Projected Interface

There are several systems that project digital information onto the surface of a physical environment and interact with it (Harrison *et al.*, 2011; Shilkrot *et al.*, 2011; Kim *et al.*, 2010; Boring *et al.*, 2010). Providing touch based interactivity with arbitrary projected objects was a challenging point for all systems. Omni Touch (Harrison *et al.*, 2011) introduced a wearable depth-sensing and projection system that enables interactive multitouch applications on very day surfaces. Omni touch needs some enhancement to their system like adding postures and gestures of the arms and hands could be used for input. In RemoAct the user can make sending gesture to send their public data to (other users, screens, or printers). The Bonfire (Kane *et al.*, 2009) system offered a portable form-factor to extend the desktop computing experience onto the tabletop using several hand held projectors. The use of arbitrary surfaces for content projection was explored with The Everywhere Displays Projector (Pinhanez, 2001) to augment indoor spaces with projected content at any location using a steerable projector.

Motion Beam (Willis *et al.*, 2011b) presents metaphor for character interaction with handheld projectors. Sixth Sense (WUW-wear Ur world) (Mistry *et al.*, 2009) featured a worn camera/ projector combination. Finger tracking was achieved by wearing fingertip markers (e.g., colour or IR reflective). Side By Side (Willis *et al.*, 2011a) Introduced novel hardware and software platform for ad-hoc multi-user interaction with handheld projectors. Through the use of a device-mounted camera and a hybrid visible/IR light handheld projector they have shown the viability of tracking projected content with invisible fiducial markers.

RemoAct explore a shoulder-worn system that provides a new interaction technique with different devices using a display easily projected on different surfaces, like hand, tabletop or walls and thus fitting different environments.

Augmented Reality

Augmented reality used to improve the user's real environment by adding virtual information on it. This augmented information gives the user more utility to reduce risk and cost and also gives some entertainment applications.

Second Surface (Kasahara *et al.*, 2010) introduced a novel multi-user augmented reality system that fosters a real-time interaction for user generated contents on top of the physical environment. This interaction takes place in the physical surroundings of everyday objects such as trees or houses. The Augmented Surfaces (Rekimoto and Saitoh, 1999) project established a direct spatial relationship between laptop screen content and content

projected onto nearby surfaces. RemoAct use AR to assign every user to their public area. Users facing each other will be able to see each other's public area augmented beside them.

Interaction with Gestures

Detecting fingers, hands (Mao *et al.*, 2009; Comport *et al.*, 2006; Wang *et al.*, 2011; Frati and Prattichizzo, 2011; Gustafson *et al.*, 2010) and gestures (Zhi-gang and Lei Zhu, 2009; Leithinger *et al.*, 2011; Lin *et al.*, 2011) were the most challenging points in our system; we use the depth-camera (Francese *et al.*, 2012; Frati and Prattichizzo, 2011; Zafrulla *et al.*, 2011) tracking without augmenting the user or environment. Light Space (Wilson and Benko, 2010) uses an array of depth cameras to track users and arm-level manipulations in an augmented room.

Kinect Fusion (Izadi *et al.*, 2011) enables a user holding and moving a standard Kinect camera to rapidly create detailed 3D reconstructions of an indoor scene. Yin and Xie (2001), segment hand images from complex backgrounds with the color segmentation approach based on the RCE neural network and recognize 2D hand postures by analyzing the topological features of the segmented hand.

RemoAct uses single finger gestures to give commands to the system. We have used the 3d coordinates extracted from the Kinect depth camera for gesture recognition with preprocessing based mainly on the \$I used in (Wobbrock *et al.*, 2007), Unistroke and Procrator preprocessing algorithms (Wobbrock *et al.*, 2007; Li, 2010).

These gestures were classified using Support Vector Machine (SVM) learning algorithm.

Communication through Distance

Current screen-less devices only support buttons and gestures. Pointing is not supported because users have nothing to point at. Lee *et al.* (2007) introduce Ubiquitous Fashionable Computer (UFC) and developed a wireless gesture recognition device, called iThrow, which is worn on one's finger like a ring. The UFC, with the help of iThrow, can control ubiquitous environment using an intuitive hand motion. The UFC user can point to another UFC user (who make a receiving gesture) and send to him data or can point to a device (Like Printer) to print data. Imaginary Interfaces: Spatial Interaction with Empty Hands and without Visual Feedback presents Imaginary Interfaces (Gustafson *et al.*, 2010), screen-less devices that allow users to perform spatial interaction with empty hands and without visual feedback.

Chunming Jin *et al.* (2006) have proposed a system to support public space communication by using a large screen in public space and a PDA.

Detecting Objects

Mariano *et al.* (2002) propose a novel method to evaluate the performance of object detection algorithms in video sequences. He *et al.* (2011), proposes an ovel object detection algorithm with the combination of appearance, structural and shape features. Portico (Avrahami *et al.*, 2011) uses tangible interactions to enable the user to work in a wide area on and around the table.

RemoAct helps the user to point to a device (Like Printer) to print some data, point to a screen to display some data, or even help the user to point to another user to send the public data to the another user's public data. Table 1 shows a summarized table for RemoAct features and other related work.

Table 1. Related work comparison to RemoAct

	Interactive projected Interface	Augmented reality	Interaction by gestures	Communication through distance
OmniTouch (Harrison <i>et al.</i> , 2011)	√		√	
Bonfire (Kane <i>et al.</i> , 2009)	√			
The everywhere displays projector (Pinhanez, 2001)	√			
Motion beam (Willis <i>et al.</i> , 2011b)	√			
Sixth sense	√		√	
Side by side (Willis <i>et al.</i> , 2011a)	√			
Second surface (Kasahara <i>et al.</i> , 2010)		√		√
The augmented surface (Rekimoto and Saitoh, 1999)	√	√		
Light space (Wilson and Benko, 2010)			√	
KinectFusion (Izadi <i>et al.</i> , 2011)			√	
Lee <i>et al.</i> (2007)			√	√
Jin <i>et al.</i> (2006)	√			√
RemoAct	√	√	√	√

System Overview

RemoAct has two interaction techniques: Mid-air interaction and touch interaction. Using these two techniques, it consists mainly of five modules all of these modules: Require a Kinect and a pocket projector both connected to the user's device and attached to the user. The Kinect is to capture the user's hands and fingers; the pocket projector is to display the device interface on a flat surface within a range from 0.5 to 0.8 m from the Kinect.

As shown in Fig. 1, each user has two areas, one public area and one private area. In each of the two areas, the user can create many tiles on the fly. Each tile is a separate workspace that the user works on. Tiles in the private area will not be shared and will only be seen by that user. However, all tiles in the public area will be shared with all the identified users. Users can drag and drop any tile any time from the private area to the public area to be shared and vice versa to stop sharing an item. The user can also send a specific item to another user, or send it to other devices like printers to be printed. Two users can draw one chart together; each of them draws part and then they merge the parts together.

Mid-Air Interaction: All system modules can be manipulated using mid-air interaction. Mid air interaction is mainly used when the user is at a distance from the display. The user can only wear the kinect and can use a fixed projector a monitor as a display. Interaction is made through hand gestures, finger tracking and hand tracking. This gives the user more freedom in movement. Figure 2 shows a user manipulating the system using Mid-Air interaction.

Touch Interaction: The same modules can also be manipulated using touch interaction. The user, in this case, must be using a projector and must be near enough from the projected display. We use different algorithms to calibrate the projected 2D display with the kinect 3D vision. The hand is first recognized then different algorithms are applied to decide whether the user touched the surface or not. Figure 3 shows a user manipulating the system using touch interaction.

Interactive Projected Interface

The user will be able to project their device interface through the pocket projector attached to him/her, this projected interface will enable the user to interact through it as if (s) he was interacting directly with their device. This will be done through capturing the user's fingers using the Kinect to control the projected surface. By providing each user with public and private areas, *RemoAct* facilitates sharing data among users. By dragging a tile from the private area to the public area, the tile content will be available to all users identified. Figure 3 shows Interactive Projected Interface overview.

Merging Surfaces

We aim at facilitating group work and files sharing. We create applications that would enable the system users to merge their projected surfaces. The users can also draw one chart together simultaneously, each on their device. This enables the users to comment on other users' charts or to complete the min real time.

In Fig. 4 the first user is drawing a painting in black color using their finger. Figure 5 shows the second user drawing another painting using red color, also using their finger. When the two users make the merging gesture, Fig. 6 appears after the two paintings are merged.

Public and Private Areas

As *RemoAct* aims at facilitating collaborative work, each user has private and public areas. The user can easily create tiles and work on them while on the private area. The maximum number of tiles that could be created by a user depends on the projected surface size. However, in the current implementation we have no limit on the number of private tiles created by the user as we allow overlapping tiles. If desired to share it with others, by dragging the tile from the private to the public area the tile content will be broad casted to all users identified on the system and connected to the network. The public area is shared by all users and any tiles added or removed from it will be seen by all users. In this way, if a user moves a tile back to the private area, no other user can display it until it is moved to the public area again.

Figure 3 shows two users, each having their own areas. User 1 has dragged one picture to the public area and makes the confirmation gesture. Then the other user's public area contains what the first has shared. Then user 2 is dragging the picture to their private area.

Augmented Reality

RemoAct uses augmented reality with no marker usage. AR techniques are used to deliver a new experience of sending data in a real-world environment. Using face and object detection techniques, the system sends data to the detected person or object when the sender performs a sending gesture.

When the user is facing another user, each can see the other user's files that have been made public. These files appear side by side to the user's face. User files can be distinguished from other user files by using face identification as every user has their own attached file according to their identity. When the camera detects a face, the system runs the identification algorithm on it. Once the user is identified a request is sent to receive this user's public files; next, these files are sent to the user facing him and these files are sent to him and can be shown augmented beside the user's face.

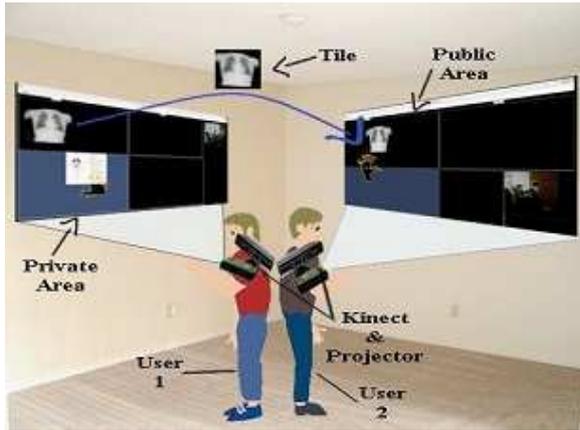


Fig. 1. The image shows the system overview, Interactive Projected Interface

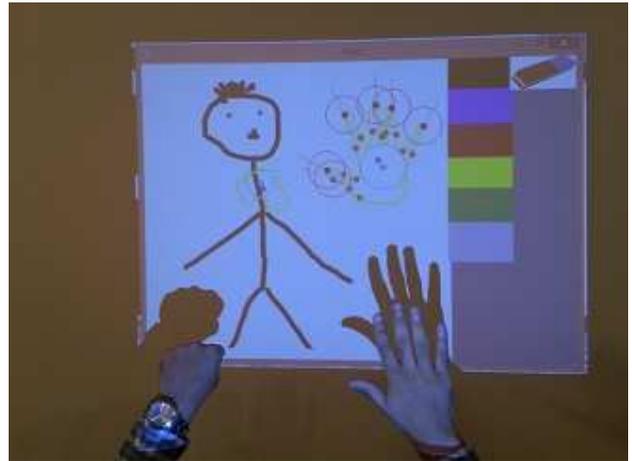


Fig. 4. User 1 drawing a man using his/her finger, their projection surface is a wall. Then, making the merging gesture

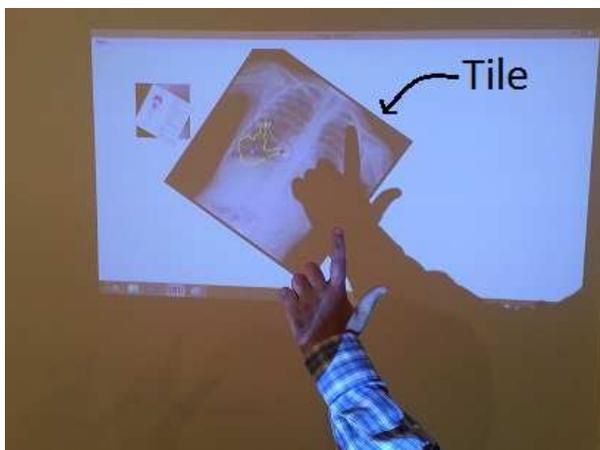


Fig. 2. Interactive Projected Interface: User creating live tiles and manipulating images within, using Mid-Air Interaction



Fig. 5. User 2 drawing an umbrella using his/her finger, projection surface is a table

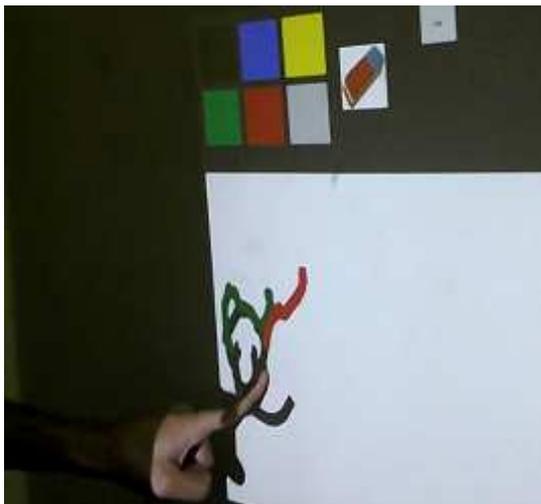


Fig. 3. Interactive Projected Interface: User drawing a diagram, using Touch Interaction



Fig. 6. The two drawings are merged, after users made the merging gesture

Communication through Distance

RemoAct makes sure that the system users can communicate not just with one another but also with the surrounding devices. In this module, users can send each other files by making a pointing gesture at the designated user. The user can also send files to other devices by making a pointing gesture towards the designated device. For example, if the user points towards a display screen, this screen would display what is currently being projected by the user; if the user points to a printer, the printer would print the document that is currently being projected by the user. The user can also send to other devices like a projector and so on. The user can send a file to another device or user by only pointing at them.

Proposed System

Mid-Air Interaction

In a similar manner to Li (2012), we use Kinect for hand gesture interaction. As stated previously, using 2D video-based algorithms in tracking hands leads to confusion when there is occlusion. Instead, if color-based segmentation is used, the hands are not easy to be separated from the face due to their similar skin colors; furthermore, the performance suffers when the lighting conditions change or the background is non uniform. Data glove devices that provide hand positions may be a good way to augment color-based segmentation, but they cause uncomfortable User Experience (UE) because of the physical contact. Fortunately, a new contact-less approach came out in the past ten years: Using depth discontinuities to separate hands from the background. In this way, depth-based systems avoid all these problems mentioned above. Moreover, the private area is protected from being accessed by other users as the main controller is gained by portable shoulder Microsoft Kinect.

Touch Interaction

Revise on DigitalDesk (Wimmer *et al.*, 2010) used a technique to track hand interactions. We first preprocess the depth image by computing its horizontal derivative using a 5×5 Sobel filter. The window size is chosen according to the depth image resolution, 640×480. We then traverse the derivative curve of each scan line looking for a specific oscillatory pattern (derivative is first positive, then oscillates near zero and then is predominantly negative). We label this 1-D range as a finger strip, a candidate that might become part of a finger. The width of the strip is constrained by the range of typical finger widths (between 5mm and 25mm). Next, vertically adjacent strips are grouped together to form a finger candidate. If the length of the candidate satisfies the length constraint (25mm-150mm), it is regarded as a finger, with the fingertip at the center of the topmost strip. Since noise can cause strips of a real

finger to exceed the width constraint and as a result a finger is cut off, we add a noise tolerance strategy: If there are two strip clusters that are vertically closed to each other ($j - 10$ pixels) a linear interpolation is applied to the area between them to connect the two parts.

In order to identify touch points, we use the BFS flooding technique. Starting from the fingertip in the depth map, we flood fill in all directions but downwards, until the number of flooded pixels exceeds a threshold (500 pixels). The flooding is performed by a breadth-first search. If the finger is touching a surface, the flooded pixels will form a triangle. This technique allows us to discriminate between hovering on top of the surface or touch based interaction on the surface or on any other surface (e.g., a book or a sheet of paper in mid-air).

Gesture Recognition

In *RemoAct*, single finger gestures are used to give simple commands to the system. 3D coordinates extracted from the Kinect depth camera are used for gesture recognition but after a preprocessing procedure. The preprocessing procedure is based mainly on the 1D Uni-stroke and Protractor preprocessing algorithms (Li, 2010; Wobbrock *et al.*, 2007).

These gestures were classified using Support Vector Machine (SVM) learning algorithm, by applying LIBSVM: A Multi-Class Support Vector Machine (SVM) tool for classification and prediction. To define the SVM parameter, we use Radial Basis Kernel Function (RBF) with parameters Cost and Gamma automatically predicted using a Grid Parameter Search tool (grid.py), which uses v-fold cross validation to estimate the accuracy of each parameter combination and decide the best parameter combination for the problem given.

In v-fold cross-validation, the training set is divided into v subsets of equal size. Sequentially, one subset is tested using the classifier trained on the remaining (v-1) subsets. The cross-validation accuracy is the percentage of data, which are correctly classified. The grid search tool (Grid.py), which is written in Python, takes three main parameters for prediction:

- (V n): n equally divided subsets of the training data set for V-fold Cross Validation
- $\log_2 C$ Begin, end, step: The ranges of Cost, its beginning, end and the number of steps to form the sequence of the tried Cost parameter
- $\log_2 \gamma$ Begin, end, step: The ranges of Gamma, its beginning, end and the number of steps to form the sequence of the tried Gamma parameter

According to the tool's documentation, trying exponentially growing sequences of C and γ is a practical method to identify good parameters. We have done an empirical study using different parameters' values to get the C and γ values which achieve the highest cross-validation accuracy as shown in Table 2.

Table 2. SVM parameters experiment

$\log_2 C$ (begin, end, step)	$\log_2 g$ (begin, end, step)	C	γ	Cross validation
$\log_2 C$ (-5, 15, 2)	$\log_2 g$ (3, -15, -2)	2^9	2^4	Accuracy rang V=3, 4, 5,..., 9, 10
$\log_2 C$ (-10, 30, 10)	$\log_2 g$ (3, -15, -2)	2^9	2^3	87.1%-88.9%
$\log_2 C$ (-5, 15, 2)	$\log_2 g$ (-10, 30, 10)	2^{11}	2^0	81.5%-86.1%
$\log_2 C$ (5, 10, 2)	$\log_2 g$ (-10, 30, 10)	2^5	2^{10}	80.5%-81.5%
$\log_2 C$ (-10, 30, 10)	$\log_2 g$ (5, 9, 2)	2^{10}	2^5	89.8%-91.67%

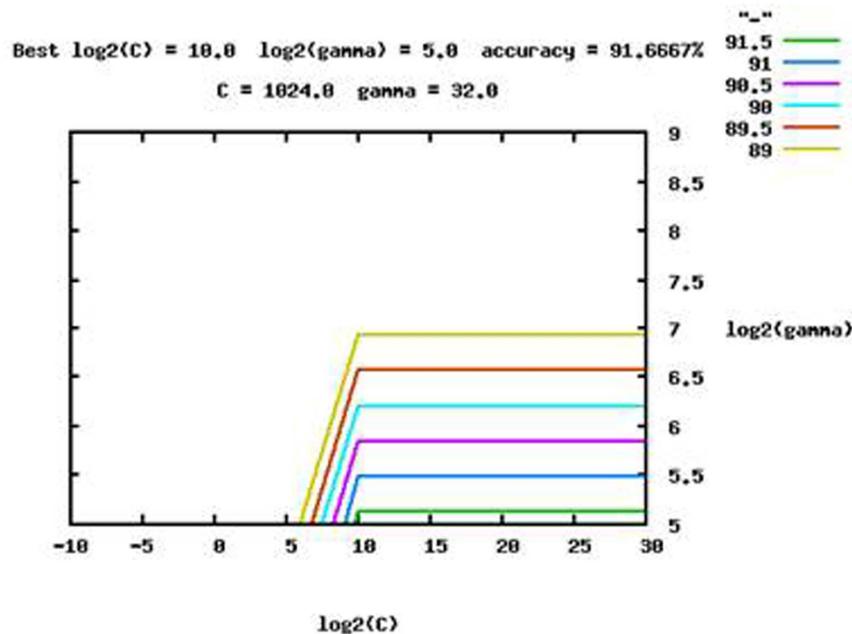


Fig. 7. SVM Parameters graph

From the previous experiment, we concluded that the acceptable (C, γ) were (210, 25) with the cross-validation rate 91.667% Fig. 7.

Various pairs of (C, γ) values on the x-axis and y-axis respectively is tried and the one with the best cross-validation accuracy is chosen. We have used two gestures, (Check) and (Click), as shown in Fig. 8 and 9 respectively.

For resources management, the gesture recognizer prediction only starts when one finger is recognized by the depth camera.

Object Detection and Face Identification

In *RemoAct* Cascade classifier is used for object detection, using Haar like features and a boosting classifier (AdaBoost) applying it to LCD screen on the cascade classifier, It is trained using 2000 positive image and 1000 negative image.

Face identification is applied in order for the system to know whom to send the data to. We used Emgu CV image processing library. Haarcascades is used for face detection and Eigen Object Recognizer for identification. To increase the performance and speed, parallel programming and tree

structure are used for face classification. The tree structure is more efficient for comparing and eliminating the face feature, while parallel programming is used to create several threads to retrieve the correct face.

Finger Tracking

Finger tracking was accomplished using image processing on an RGB image stream, but it was not accurate to the needed extent, in order to accomplish a successful tracking a white background should be present, at the same time different lighting conditions affected the accuracy to a great extent. Another method that seemed to help a lot and increased the accuracy dramatically was to use the depth data that is coming from the depth image produced by Microsoft Kinect as shown in Fig. 10 and using Open NI to process the raw data coming from the kinect.

The technique was to track the user's hands first. To detect the hands, we collect the hands point's coordinates then using K-mean classifier and using ascertain distance value of 50 pixels, it then classifies the hands as two separate hands or one hand.

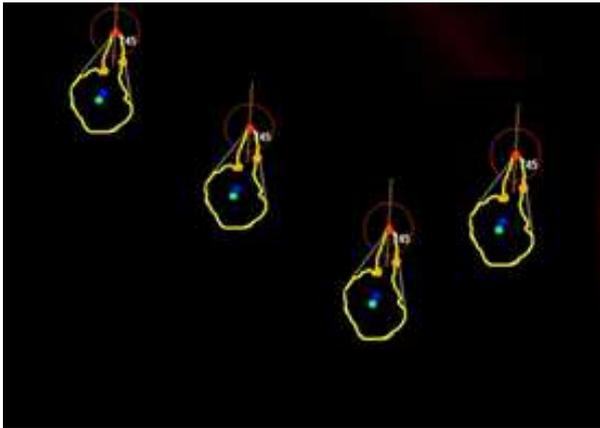


Fig. 8. Check gesture

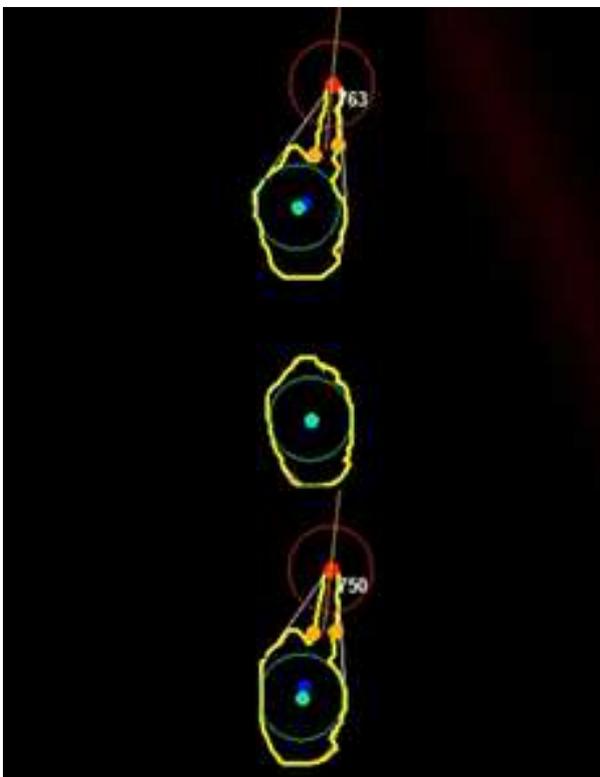


Fig. 9. Click gesture

The next stage is to track the user fingers; this is applied by developing a contour around the hand and then a convex hull is drawn that covers the hand using Graham scan algorithm. Finally finger points are gathered by searching for the common points that are located on both the contour and the convex hull of the hand where the contour is curved enough.

Gesture Start\End

To perform a certain gesture, coordinates are recorded and then compared to the model produced by the SVM.

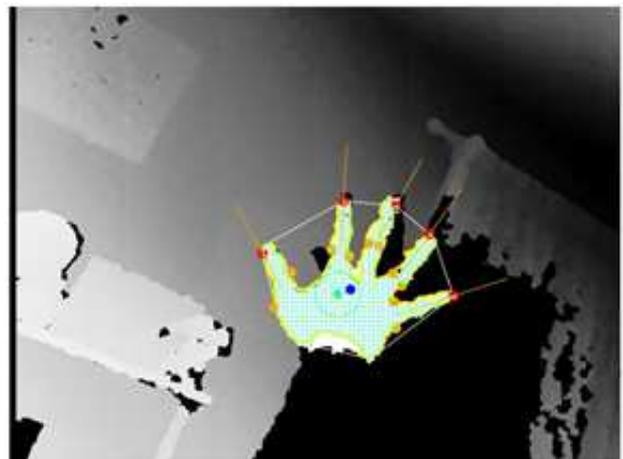


Fig. 10. Depth image

The trigger to start recording is to show only one finger out of the featured hand. To signal the end of the gesture all five fingers of the hands must be shown to stop tracking and start processing.

Experimental Studies

In order to evaluate the key characteristics of our system, we conducted two experimental studies to test these characteristics.

Experiment I

This user study tests the ability of RemoAct to capture gestures on projected surfaces; it also tests whether the user can create several tiles and whether the system can merge two projected surfaces from two users and send the result correctly to different objects and other system users. Furthermore, it verifies whether the system can identify the touch interaction.

Participants

We recruited 10 participants ranging in age from 19 to 59, males and females. Nine participants were right-handed and 1 left-handed.

Procedure

Participants interacted with the projected display on different surfaces. They were allowed to draw some paintings on a paint-like application using their finger to test the finger tracking performance. Before starting the experiment, Participants were guided on how to draw using fingers and how to use different gestures to interact with the system. Two minutes of training were sufficient for the users to start interacting with the system conveniently.

The participants were allowed to test the projected image manipulation through three gestures (1) creating surface gesture (2) drag and drop (3) zoom in/out. Before

testing, participants were trained to use each gesture for 5 times before getting to use the system. Each participant repeated the experiment two times, in order to test the mid-air interaction and the touch interaction.

RemoAct enables the user to merge two projected surfaces of two different users to work simultaneously. In order to test this feature, we performed an experiment using our Paint-like application, where each of the two users was allowed to draw a painting on his/her own. After drawing the painting, the user performed a merging gesture to have the two paintings merged together on one surface. The experiment was repeated three times to record the accuracy.

After each user has created their own painting, (s) he sends their own painting to LCD Screen. This experiment mainly aims to test object detection accuracy and data transferring. In the experiment, the participant performs the sending gesture while facing the device towards the object he aims to send to, so that the kinect can see the target. Participants will repeat the experiment two times on each object for accuracy measures.

One key feature of *RemoAct* is that each user can have multiple surfaces projected from a single projector, where these surfaces can be set to private or public. To test this feature, we created an experiment to test the collaboration between users where each participant created two surfaces, one is set to private and the other to public. The participant sent the public surface using a sending gesture to another participant by facing the device towards this participant, so that the kinect can see the target; through the face detection techniques used by *RemoAct*, the public surface is received by the detected participant. Before performing the experiment, each involved participant's face is captured three times for training purposes. The aim of the experiment is to test face detection and file transfer features.

Experiment II

This experiment tests how the systems captures different gestures, recognize objects and identifies users.

Participants

We recruited 5 participants ranging in age from 20 to 25, males and females. Three participants were right-handed and 2 left-handed.

Procedure

Each participant was first trained to use the gestures. Each user then made each gesture 10 times and the correctly recognized gestures were counted. Then we tested the accuracy for face identification. First each user is tested separately, then more users are added together, until all five users are together in the same capture. These procedures are conducted 3 times in 3 different lightings.

Finally, we compared *RemoAct* to other conventional methods. To compare file sharing, the participants dragged a tile to the public area to share it, sent the file using shared folder on the cloud and sent a file using flash memory. Next, we tested merging two drawings. One participant drew a painting then merged with the other.

Results and Discussion

After running ANOVA test, we observed that there were no significant performance differences between participants, thus, results were combined. Significant performance difference appeared between the training phase and the actual testing. The following results are for the testing phase after the participants finishes training. For gesture recognition, accuracy reached 90-95%. Object recognition and face identification accuracy varied with the variation of light. In well-lit places the accuracy reached over 85% while in darker places, the accuracy decreased. For finger tracking, the participants faced inconvenience at first and needed more training.

After testing different environments, we reached the conclusion that fixing the kinect to a distance range 50-80 cm from the user hands, whether on their shoulder or on a chair back, will result in more accuracy for finger tracking. We can use the projector as a light source, by displaying a bright screen when the face identification or object recognition is triggered. Having distance limitation and light brightness considered to be one of the main challenges facing *RemoAct* that need to be solved in the future.

After testing different conventional systems and *RemoAct*, we found that *RemoAct* takes less time for two users to draw one chart. As the system enables the users to work simultaneously, it reduces the needed time, short compared to successive drawing. Comparing sharing, we found that the system is faster and uses fewer steps than sharing with flash drive. Using *RemoAct* can take a slightly longer time than using a shared folder on the cloud, if the users are not properly trained. However, *RemoAct* offers the advantage of creating many tiles on the fly.

Future Work

Although this work has derived the implications of the design guidelines and demonstrated appropriate scenarios, future work on *Remoact* project will continue to add predefined shapes, (square, rectangle, triangle etc.) so it is easier to draw UML diagrams, to the paint application to help user to draw flowcharts, class diagrams, sequence diagrams, etc... For tracking user fingers, we intended to develop an enhanced algorithm to increase the accuracy and to increase intuition movements. Improving the interaction with the projected interface, by using finger tracking algorithm to make it touch like. We also want to enhance the calibration between the user's projected interface and

the surfaces projected on, so the user can interact freely with their interface on different surfaces.

Conclusion

In this study, we introduced and evaluated the novel interaction system of *RemoAct*, in which a wearable system that provides a new interaction technique for dealing with devices through projected display and hand gestures in an everyday environment. *RemoAct* combines a depth camera and a pocket projector to deliver touch experience on displays projected on everyday surfaces. In addition, we presented the ability of merging two projected displays on the same surface to enable better sharing and visualization of data. We showcased a feature where a user can send data to people or objects detected by the depth camera through simple hand gestures. *RemoAct* also integrates augmented reality, which is used to add more intuitive experience to the users.

In *RemoAct*, we aimed to show how depth cameras and projectors can be used to offer free interaction experience on everyday surfaces and offer easier sharing of data with people and objects.

Acknowledgement

We thank HCI lab Helwan University for supervision, Microsoft Advanced Technology Lab in Cairo (ATLC) and Information Technology Industry Development Agency (ITIDA) for sponsorships.

Funding Information

This project was self-funded, by the team members. We received sponsorship amount from Microsoft Advanced Technology Labs in Cairo and Information Technology Industry Development Agency.

Author's Contributions

All authors equally contributed in this work.

Ethics

This article is original and contains unpublished material. The corresponding author confirms that all of the other authors have read and approved the manuscript and no ethical issues involved.

References

- Avrahami, D., J.O. Wobbrock and S. Izadi, 2011. Portico: Tangible interaction on and around a tablet. Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology, Oct. 16-19, ACM, New York, pp: 347-356. DOI: 10.1145/2047196.2047241
- Azuma, R.T., 1997. A survey of augmented reality. Presence: Teleoperators Virtual Environ., 6: 355-385.
- Boring, S., D. Baur, A., S. Gustafson and P. Baudisch, 2010. Touch projector: Mobile interaction through video. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Apr. 10-15, ACM, New York, pp: 2287-2296. DOI: 10.1145/1753326.1753671
- Comport, A.I., E. Marchand, M. Pressigout and F. Chaumette, 2006. Real-time markerless tracking for augmented reality: The virtual visual servoing framework. IEEE Trans. Visualization Comput. Graphics, 12: 615-628. DOI: 10.1109/TVCG.2006.78
- Francese, R., I. Passero and G. Tortora, 2012. Wiimote and kinect: Gestural user interfaces add a natural third dimension to HCI. Proceedings of the International Working Conference on Advanced Visual Interfaces, May 22-25, ACM, New York, pp: 116-123. DOI: 10.1145/2254556.2254580
- Fрати, V. and D. Prattichizzo, 2011. Using Kinect for hand tracking and rendering in wearable haptics. Proceedings of the IEEE World Haptics Conference, Jun. 21-24, IEEE Xplore Press, Istanbul, pp: 317-321. DOI: 10.1109/WHC.2011.5945505
- Gustafson, S., D. Bierwirth and P. Baudisch, 2010. Imaginary interfaces: Spatial interaction with empty hands and without visual feedback. Proceedings of the 23rd Annual ACM Symposium on User Interface Software and Technology, Oct. 3-6, ACM, New York, pp: 3-12. DOI: 10.1145/1866029.1866033
- Harrison, C., H. Benko and A.D. Wilson, 2011. OmniTouch: Wearable multitouch interaction everywhere. Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology, Oct. 16-19, ACM, New York, pp: 441-450. DOI: 10.1145/2047196.2047255
- He, L., H. Wang and H. Zhang, 2011. Object detection by parts using appearance, structural and shape features. Proceedings of the International Conference on Mechatronics and Automation, Aug. 7-10, IEEE Xplore Press, Beijing, pp: 489-494. DOI: 10.1109/ICMA.2011.5985611
- Izadi, S., D. Kim, O. Hilliges, D. Molyneaux and R. Newcombe *et al.*, 2011. Kinectfusion: Real-time 3d reconstruction and interaction using a moving depth camera. Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology, Oct. 16-19, ACM, New York, pp: 559-568. DOI: 10.1145/2047196.2047270
- Jin, C., S. Takahashi and J. Tanaka, 2006. Interaction between small size device and large screen in public space. Proceedings of the 10th International Conference on Knowledge-Based Intelligent Information and Engineering Systems, Oct. 9-11, Springer Berlin Heidelberg, UK, pp: 197-204. DOI: 10.1007/11893011_25

- Kane, S.K., D. Avrahami, J.O. Wobbrock, B. Harrison and A.D. Rea *et al.*, 2009. Bonfire: A nomadic system for hybrid laptop-tabletop interaction. Proceedings of the 22nd Annual ACM Symposium on User Interface Software and Technology, Oct. 4-7, ACM, New York, pp: 129-138. DOI: 10.1145/1622176.1622202
- Kasahara, S., V. Heun, A.S. Lee and H. Ishii, 2010. Second surface: Multi-user spatial collaboration system based on augmented reality. Proceedings of the SIGGRAPH Asia 2012 Emerging Technologies, Nov. 28-Dec. 1, ACM, New York, pp: 1-20. DOI: 10.1145/2407707.2407727
- Kim, S., J. Chung, A. Oh, C. Schmandt and I.J. Kim, 2010. Ilight: Information flashlight on objects using handheld projector. Proceedings of the Extended Abstracts on Human Factors in Computing Systems, Apr. 10-15, ACM, New York, pp: 3631-3636. DOI: 10.1145/1753846.1754030
- Lee, J., S.H. Lim, J.W. Yoo, K.W. Park and H.J. Choi *et al.*, 2007. A ubiquitous fashionable computer with an i-throw device on a location-based service environment. Proceedings of the 21st International Conference on Advanced Information Networking and Applications Workshops, May 21-23, IEEE Xplore Press, Niagara Falls, pp: 59-65. DOI: 10.1109/AINAW.2007.63
- Leithinger, D., D. Lakatos, A. DeVincenzi, M. Blackshaw and H. Ishii, 2011. Direct and gestural interaction with relief: A 2.5d shape display. Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology, Oct. 16-19, ACM, New York, pp: 541-548. DOI: 10.1145/2047196.2047268
- Li, Y., 2010. Protractor: A fast and accurate gesture recognizer. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Apr. 10-15, ACM, New York, pp: 2169-2172. DOI: 10.1145/1753326.1753654
- Li, Y., 2012. Hand gesture recognition using Kinect. MSc Thesis, Department of Computer Engineering and Computer Sciences, University of Louisville, Louisville, Kentucky.
- Lin, J., H. Nishino, T. Kagawa and K. Utsumiya, 2011. A method of two-handed gesture interactions with applications based on commodity devices. Proceedings of the International Conference on Complex, Intelligent and Software Intensive Systems, Jun. 30-Jul. 2, IEEE Xplore Press, Seoul, pp: 697-702. DOI: 10.1109/CISIS.2011.117
- Mao, G.Z., Y.L. Wu, M.K. Hor and C.Y. Tang, 2009. Real-time hand detection and tracking against complex background. Proceedings of the 5th International Conference on Intelligent Information Hiding and Multimedia Signal Processing, Sep. 12-14, IEEE Xplore Press, Kyoto, pp: 905-908. DOI: 10.1109/IIH-MSP.2009.133
- Mariano, V., J. Min, J.H. Park, R. Kasturi, D. Mihalcik and H. Li *et al.*, 2002. Performance evaluation of object detection algorithms. Proceedings of the 16th International Conference on Pattern Recognition, Dec.11-15, IEEE Xplore Press, pp: 965-969. DOI: 10.1109/ICPR.2002.1048198
- Mistry, P., P. Maes and L. Chang, 2009. Wuw-wear ur world: A wearable gestural interface. Proceedings of the Extended Abstracts on Human Factors in Computing Systems, Apr. 4-9, ACM, New York, pp: 4111-4116. DOI: 10.1145/1520340.1520626
- Pinhanez, C.S., 2001. The everywhere displays projector: A device to create ubiquitous graphical interfaces. Proceedings of the 3rd International Conference on Ubiquitous Computing, Sep. 30-Oct. 2, Springer Berlin Heidelberg, USA, pp: 315-331. DOI: 10.1007/3-540-45427-6_27
- Rekimoto, J. and M. Saitoh, 1999. Augmented surfaces: A spatially continuous work space for hybrid computing environments. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, May 15-20, ACM, New York, pp: 378-385. DOI: 10.1145/302979.303113
- Shilkrot, R., S. Hunter and P. Maes, 2011. Pocomo: Projected collaboration using mobile devices. Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services, Aug. 30-Sep. 2, ACM, New York, pp: 333-336. DOI: 10.1145/2037373.2037424
- Soler, L., S. Nicolau, J. Schmid, C. Koehl and J. Marescaux *et al.*, 2004. Virtual reality and augmented reality in digestive surgery. Proceedings of the 3rd IEEE and ACM International Symposium on Mixed and Augmented Reality, Nov. 2-5, IEEE Xplore Press, Washington, DC, pp: 278-279. DOI: 10.1109/ISMAR.2004.64
- Wachs, J.P., M. Kolsch, H. Stern and Y. Edan, 2011. Vision-based hand-gesture applications. Commun. ACM, 54: 60-71. DOI: 10.1145/1897816.1897838
- Wang, R., S. Paris and J. Popovic, 2011. 6d hands: Markerless hand-tracking for computer aided design. Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology, Oct. 16-19, ACM, New York, pp: 549-558. DOI: 10.1145/2047196.2047269
- Wimmer, R., F. Hennecke, F. Schulz, S. Boring, A. Butz and H. Hußmann, 2010. Curve: Revisiting the digital desk. Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries Oct. 16-20, ACM, New York, USA, pp: 561-570. DOI: 10.1145/1868914.1868977.
- Willis, K.D., I. Poupyrev, S.E. Hudson and M. Mahler, 2011a. Sidebyside: Ad-hoc multi-user interaction with handheld projectors. Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology, Oct. 16-19, ACM, New York, pp: 431-440. DOI: 10.1145/2047196.2047254

- Willis, K.D., I. Poupyrev and T. Shiratori, 2011b. Motionbeam: A metaphor for character interaction with handheld projectors. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, May 7-12, ACM, New York, pp: 1031-1040. DOI: 10.1145/1978942.1979096
- Wilson, A.D. and H. Benko, 2010. Combining multiple depth cameras and projectors for interactions on, above and between surfaces. Proceedings of the 23rd Annual ACM Symposium on User Interface Software and Technology, Oct. 3-6, ACM, New York, pp: 273-282. DOI: 10.1145/1866029.1866073
- Wobbrock, J.O., A.D. Wilson and Y. Li, 2007. Gestures without libraries, toolkits or training: A \$1 recognizer for user interface prototypes. Proceedings of the 20th Annual ACM Symposium on User Interface Software and Technology, Oct. 7-10, ACM, New York, pp: 159-168. DOI: 10.1145/1294211.1294238
- Yin, X. and M. Xie, 2001. Hand gesture segmentation, recognition and application. Proceedings of the IEEE International Symposium on Computational Intelligence in Robotics and Automation, Jul. 29-Aug. 1, IEEE Xplore Press, pp: 438-443. DOI: 10.1109/CIRA.2001.1013240
- Zafrulla, Z., H. Brashear, T. Starner, H. Hamilton and P. Presti, 2011. American sign language recognition with the kinect. Proceedings of the 13th International Conference on Multimodal Interfaces, Nov. 14-18, ACM, New York, pp: 279-286. DOI: 10.1145/2070481.2070532
- Zhi-gang, X. and H. Lei Zhu, 2009. Vision-based detection of dynamic gesture. Proceedings of the International Conference on Test and Measurement, Dec. 5-6, IEEE Xplore Press, Hong Kong, pp: 223-226. DOI: 10.1109/ICTM.2009.5412956