Proceedings of the Twenty-Ninth International
Florida Artificial Intelligence Research Society Conference

Intelligent Context-Aware Augmented Reality to Teach Students with Intellectual and Developmental Disabilities

Christopher Reardon[1], Rachel Wright[2], David Cihak[2], Lynne E. Parker[1]
[1]Department of Electrical Engineering and Computer Science and
[2]Department of Theory and Practice in Teacher Education
University of Tennessee, Knoxville, Tennessee 37996
{creardon, rwrigh15, dcihak, leparker}@tennessee.edu

Abstract

There is a compelling need to develop tools and strategies for people with intellectual and developmental disabilities (I/DD) in order to facilitate independence, self sufficiency, and address poor employment outcomes in adulthood. Through use of augmented reality (AR) and machine learning methods, we create an intelligent, contextually aware instructional system for persons with I/DD. We present results that demonstrate our system can be used independently by students with I/DD to quickly and easily acquire the skills required for performance of three relevant vocational tasks.

Introduction

Today's young people with I/DD face harsh realities as they enter enter adulthood, such as low employment rates, poor wages and benefits, limited community supports, and low rates of independent living (Grigal and Hart 2010). There is therefore a strong motivation to provide support for this population to increase their independence in performance of vocational tasks. One method of providing vocational support to people with I/DD is prompting, which refers to any type of assistance provided to help an individual perform a given skill. Prompting supports include modeling via job coaches, or printed or digitized materials. Because success in any job requires the performance of multiple skills, persons with I/DD can quickly acquire a plethora of such supports. Though portability of these supports is increased as they are provided through digital storage and mobile devices, the ability to quickly retrieve contextually correct support is extremely challenging, in that it remains reliant on the organizational abilities of the user, a deficit in which is inherently associated with I/DD. Therefore, additional assistance is still needed to ensure that the individual will learn to perform the skill steps correctly and with the greatest level of independence possible (Lancioni and O'Reilly 2001). Ironically, technology based approaches are often completely dependent on another person to set up and manage the content, as well as initiate all operations of the support devices. While these types of supports can be successful in training situations, because their effectiveness is contingent on the actions of another person, this creates user dependence, rather than independence, resulting in frustration, disillusionment, and device abandonment.

In this work, we describe an approach that leverages modern assistive technologies to create a learning experience superior to current instructional approaches. The main contribution of this paper is an application using computer vision and machine learning methods to provide context-aware instructional prompts delivered via a wearable, AR interface, to teach students with I/DD vocational skills. Our combination of established artificial intelligence and computer vision approaches with an easily usable and non-cumbersome AR experience to provide intelligent instruction yields a unique and powerful application to address a compelling need for an important part of our population. We detail our approach and present results that show students with I/DD successfully learned three different skills using our application. We believe this has great potential to empower persons with I/DD to live fuller, more independent lives.

Background

Augmented Reality in I/DD Education

Augmented reality allows the user to perceive and interact with the real world while simultaneously receiving additional information virtualized into their field of perception and has great potential in education (Bower et al. 2014). AR and wearables eliminate the number of steps needed to access information and resolve the deficiencies in the skills required for basic operation of devices. Of the limited studies involving AR and students with I/DD, all have been highly successful in teaching academics (Richard et al. 2007), vocational skills (Gómez et al. 2014), and independent navigation skills (McMahon, Cihak, and Wright 2015).

Context awareness is the ability to provide information appropriate to the situation. The motivating strength of context awareness is precision: it enables only the correct information to be displayed the appropriate time. Research involving context awareness typically uses tags to provide the context, e.g. (Gómez et al. 2014) uses QR tags and user-entered context information coupled with an intelligent middlewear layer, and (Gómez et al. 2011) teaches vocational skills with smartphones. Context awareness was provided by bluetooth tags in (Chang and Wang 2010) to teach wayfind-

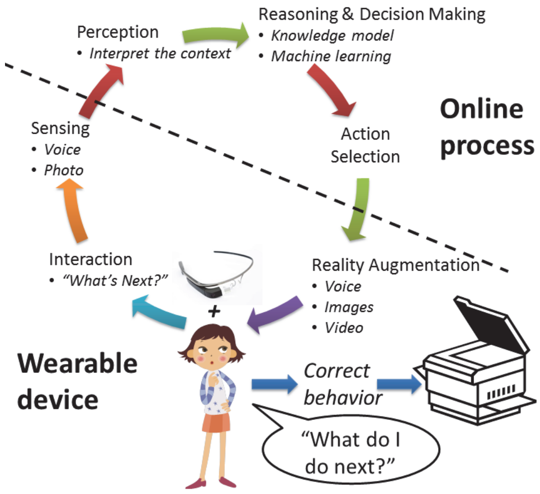


Figure 1: System overview. Process begins with user asking for assistance. A picture taken from the wearable's camera is classified and an augmented image and audio containing the proper instructional prompt for the next step is delivered.

ing using a PDA. By using computer vision and machine learning methods, our system is able to extract the context from the scene without the use of visual codes, thereby further unencumbering the student from dependency on support persons and technology and increasing the real-world generalizability of the learning process.

Context-Aware Augmented Reality

Figure 1 shows an overview of the system. A student with an AR wearable, when learning to perform a new task, can ask for the next step at any point in the task sequence. Our approach takes an image from the user's point of view and uploads it to an online server, where the image is processed and an appropriate instructional prompt is pushed back to the user's device. In our applications, the prompt is both audio and an augmented view of the image uploaded, and can include highlighted objects, buttons, or points of interaction, correct models of the solution, and/or text.

Contextual awareness is achieved through perceiving the content of the image and deciding which instructional prompt to present. Our approach first parses the image for relevant information, then uses supervised learning to solve the problem of identifying the correct context of the image. Using the classifier output and the known ground truth, the proper visual and audio prompts for the next step in the task are selected from the knowledge model, which is specific to the task, for example in the form of a lookup table or decision tree. The prompt is then delivered seamlessly to the user through the wearable AR interface.

The key differentiating feature of the context awareness is the ability to only provide prompts for the steps that the user cannot remember. This is especially significant when providing instructional support to this population in that it enables them to quickly learn to perform tasks independently. Such self-directed learning gives the student complete control and propels them towards independence.

Technical Discussion

Because of our user population's deficit in organizational skills and technological aptitude, seamless integration of the various system components and technologies is essential for an efficient user experience and successful learning.

On the wearable side of Figure 1, we use a Google Glass as our AR device. Being a first-to-market device, the Glass has some limitations. The processing capabilities are limited, particularly in terms of heat dissipation, which under heavy load requires the Glass to automatically shut down. Because of this issue, a client-cloud configuration was selected for the system, where the client is responsible for the user interface, and the artificial intelligence (perception, reasoning, and decision making) resides on the cloud side.

An application was developed for the Glass in Android 4.4.2. The user experience was streamlined: the simple audio command, "Okay Glass, what's next?" triggers the app. The user takes a picture, which is uploaded to the cloud server, and an image and audio instructional prompt is provided via the Glass display and built-in speaker within 5-10 seconds.

On the cloud side, intelligent instruction is made possible using OpenCV for image processing and Support Vector Machine (SVM) classifier implementation. We simplify the image processing problem slightly by using known information about the environment such as known color contours to segment and subselect the relevant areas of the image. After preprocessing images, Histogram of Oriented Gradient (HOG) features are extracted, and SVMs are used to classify the image. In the event of a failed or low-probability classification, the user is presented with a prompt to try again.

SVMs are trained a priori for the task steps using images of each step taken by the experimenters. A time-consuming challenge was to identify the correct number of images that are able to generalize to sufficiently represent the images the students will take, while simultaneously tuning the SVM parameters to achieve a high level of accuracy without overfitting. Future work could allow for better insight and automation into this process. Prior to the training phase of each experiment (see Experiments below), the students were instructed on the use of Glass. Classifiers were re-trained using images captured by the users in between trials to decrease failed classifications for each individual user.

A major goal of this research was to create a technical framework that allows for multiple different experiments. With modular software development, we easily incorporate three different decision workflows, as described in the Experiments section. Future work includes enabling the correct decision workflow to be selected on the fly, thus allowing the user to receive training or assistance for multiple tasks simultaneously, which would further increase independent use of the system.

Evaluation

Single-Case Experimental Design

Single case experimental design (SCED) is commonly used in special education research. As opposed to comparison between groups or subjects, participants serve as their own control data for the purpose of comparing performances between at least two experimental phases (Gast and Spriggs 2010). To determine if a causal or functional relation exists between the delivery of the independent variable (IV), the instruction system, and significant increases in the dependent variable (DV), the acquisition and maintenance of the skills required, a combined multiple probe design across participants and tasks was employed to permit intra-subject and inter-subject replication. This design allows for evaluation of intervention effects while controlling for threats to internal validity (i.e., that the learning is due to the instructional intervention) in situations where withdrawal of skill knowledge is not possible. By introducing the intervention subsequently across a minimum of three replications of skills or tasks, the possibility of any observed change occurring due to extraneous factors (e.g., practice or history effects) is eliminated, which allows for experimental control and the establishment of a causal relationship (Horner et al. 2005).

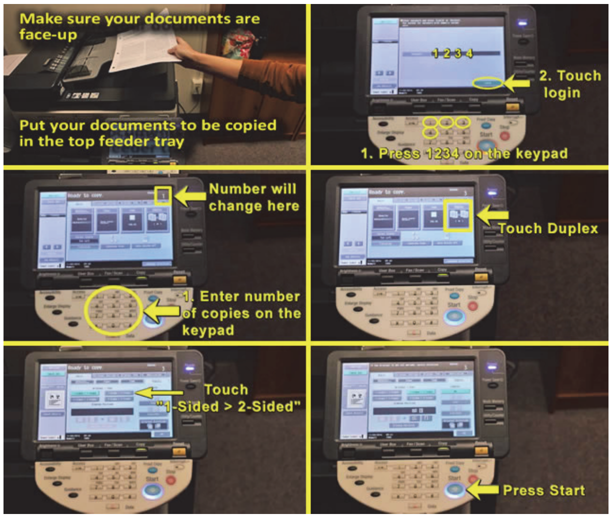


Figure 2: Experiment 1 images annotated with instructions.

Experiments

Three different skills were selected to be taught using our intelligent AR system. Three participants participated in all three experiments: two males and one female ages 19-29 and IQ scores ranging between 57 and 63. All attend a postsecondary education program for students with I/DD, regularly participate in technology in education research with IRB approval, and were unable to perform the tasks independently.

Prior to instruction, baseline measurements ("probes") were collected for all three participants on each of the three targeted tasks. Upon data stability, the intervention was introduced for Task 1 while Task 2 and Task 3 remained in baseline. Once criterion was reached in Task 1, probe data were collected for remaining tasks. Upon stable probe data, intervention was introduced to Task 2 while Task 3 remained in baseline. Once criterion was reached in Task 2, the intervention was introduced to Task 3. This pattern of introducing the independent variable and collecting probe data continued until each participant reached criteria across all three tasks. Experiments were comprised of three phases: (a) *Baseline*: probe sessions were performed to collect baseline data on each participant's performance of each target task; (b) *Instruction*: our system was used to provide instruction for correct performance of each step in Task 1 and Task 2, as well as familiarize the students with the process of asking the device for assistance; and (c) *Intervention*: students were instructed to perform the task independently, and access the system for assistance with any step of the task as needed.

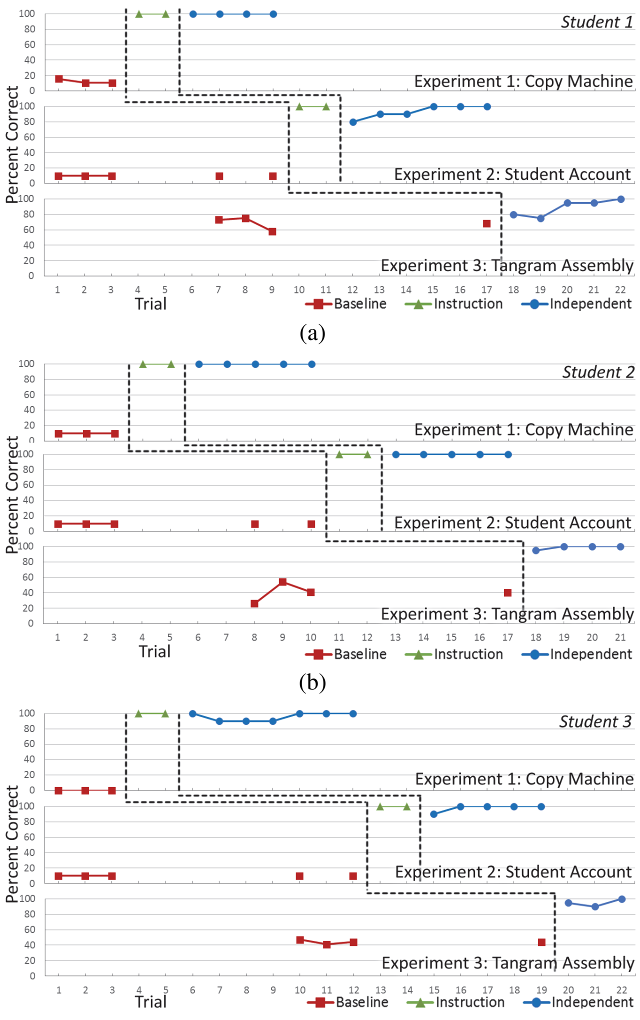
Experiment 1: Copy Machine The copy machine task consists of making the correct number of double-sided copies of a document, which is useful in office vocational work. The menus of the commercial Konica-Minolta Bizhub 363 copy machine used for this task were sufficiently complex to make using them non-intuitive for the population in this study. This task was split into 6 steps (Figure 2). Image inputs were segmented into two subselected regions: the copy machine image and the region of the image that contains the number of copies entered. Two linear SVM classifiers were trained, one for each subselection, with 100 representative images for each step / copy count, and used in parallel. The combined results of the two classifications were then used to decide which step was currently shown. Testing cross-validation showed accuracy above 95% with SVM cost $C = 2.6$ tuned manually via grid search.

Experiment 2: Student Account Statement Students learned to download a copy of their prepaid student account statement; retrieving electronic account information is a critical skill in both employment and independent living. The input for the context aware classification for this task was the section of the image containing the computer screen. A single linear SVM classifier was used, trained with 20 images from each of 9 steps. The SVM cost was set to $C = 5.9$ with a cross-validation testing accuracy of 93%.

Experiment 3: Geometry with Tangrams For the third experiment, a geometric and assembly task was selected. The ability to manipulate objects, understand relative object placement and orientation, and use the same set of parts, shapes, or items to follow novel assembly instructions is important in many job settings. Students were asked to identify, place, and rotate colored geometric shapes on a table. This task has 20 steps. The difficulty of the steps escalated from simple shape identification to relative placement and rotation, to assembling more complicated shapes. Because this task involves complicated, fine manipulation, image inputs were processed by first finding color contours, then classification was performed with a decision tree to determine the correct step and corresponding instructional prompt.

Results

The results show that with the provided intervention all students were able to achieve skill acquisition to mastery for all three tasks, defined as independent performance at 100% correct. As shown in Figure 3, in the baseline phase all students were confirmed unable to perform the task indepen-



(a)

(b)

(c)

Figure 3: Results for Experiments 1-3 for each Student 1-3 ((a)–(c)). Data is collected in three phases, shown separated by dashed lines, as described in Experiments.

dently. After several iterations of the training phase, in which the students were required to use the system for all steps (and therefore performed at 100% correct), students showed rapid acquisition of the skills in the independent phase, where they only used the system when they could not remember the next correct step. Note that in Experiment 3, the format of the skill instruction required no training phase.

The authors hypothesize that the rapid gains in skill acquisition can be attributed to the self-directed nature of the intervention system; providing the participants with the ability to control their own prompts, combined with the intelligent AR environment may have contributed to increased engagement and efficiency in learning the tasks and warrants further study. A genuine level of satisfaction with the experience on the part of the students was observed, which we also attribute to the efficiency of the experience.

Conclusion

In this work, we have shown that an intelligent, contextually-aware system using AR can successfully teach students with I/DD vocational tasks. Importantly, we have observed that a system that combines machine learning and computer vision techniques can provide an in vivo learning experience through a wearable, AR device that allows users to have control of their own learning. This shows promise for the future incorporation of intelligence into assistive technologies that will lead to increased independence for people with I/DD.

References

Bower, M.; Howe, C.; McCredie, N.; Robinson, A.; and Grover, D. 2014. Augmented reality in education–cases, places and potentials. *Educational Media International* 51(1):1–15.

Chang, Y.-J., and Wang, T.-Y. 2010. Comparing picture and video prompting in autonomous indoor wayfinding for individuals with cognitive impairments. *Personal and Ubiquitous Computing* 14(8):737–747.

Gast, D. L., and Spriggs, A. D. 2010. Visual analysis of graphic data. *Single subject research methodology in behavioral sciences* 199–233.

Gómez, J.; Montoro, G.; Haya, P. A.; García-Herranz, M.; and Alamán, X. 2011. Distributed schema-based middleware for ambient intelligence environments. *Ubiquitous Developments in Ambient Computing and Intelligence: Human-Centered Applications* 205–218.

Gómez, J.; Alamán, X.; Montoro, G.; Torrado, J. C.; and Plaza, A. 2014. Amicog–mobile technologies to assist people with cognitive disabilities in the work place. *Advances in Distributed Computing and Artificial Intelligence Journal* 2(7):9–17.

Grigal, M., and Hart, D. 2010. *Think College! Postsecondary Education Options for Students with Intellectual Disabilities.* Brookes Publishing Company.

Horner, R. H.; Carr, E. G.; Halle, J.; McGee, G.; Odom, S.; and Wolery, M. 2005. The use of single-subject research to identify evidence-based practice in special education. *Exceptional Children* 71(2):165–179.

Lancioni, G. E., and O'Reilly, M. F. 2001. Self-management of instruction cues for occupation: Review of studies with people with severe and profound developmental disabilities. *Research in Developmental Disabilities* 22(1):41–65.

McMahon, D.; Cihak, D. F.; and Wright, R. 2015. Augmented reality as a navigation tool to employment opportunities for postsecondary education students with intellectual disabilities and autism. *Journal of Research on Technology in Education* 47(3):157–172.

Richard, E.; Billaudeau, V.; Richard, P.; and Gaudin, G. 2007. Augmented reality for rehabilitation of cognitive disabled children: A preliminary study. In *Virtual Rehabilitation, 2007*, 102–108. IEEE.