

A KINECT-BASED AUGMENTED REALITY SYSTEM FOR INDIVIDUALS WITH AUTISM SPECTRUM DISORDERS

Xavier Casas, Gerardo Herrera, Inmaculada Coma and Marcos Fernández
Institute of Robotics, University of Valencia, Valencia, Spain

Keywords: Augmented Reality, Motion Capture Systems, Autism Spectrum Disorders.

Abstract: In this paper an Augmented Reality system for teaching key developmental abilities for individuals with ASD is described. The system has been designed as an augmented mirror where users can see themselves in a mirror world augmented with virtual objects. Regarding the content the tool is designed with the aim of facilitating the acquisition of certain skills in children with ASD. From a technical point of view, it has been necessary to solve two problems to develop the tool. The first one has been how to capture the user's movements, without wearing invasive devices, to animate their virtual avatar. A Kinect device has been chosen. The other problem has been to mix in the augmented virtual scene real objects located at different depths. An OSG shader has been developed, of which details will be given in this paper.

1 INTRODUCTION

Autism is a neuro-developmental disorder, and individuals with Autism Spectrum Disorder (ASD) have specific difficulties developing verbal and non-verbal communication. They also have difficulties socializing, a limited capacity for understanding mental states, and a lack of flexibility of thinking and behaving, especially when "symbolic play" is involved.

While Virtual Reality has proven effective in some people with autism (Stricklan, 1997; Herrera, 2008), there are still some more severely affected individuals who are unable to generalize their learning to new situations. In this sense, Augmented Reality is directly related to reality and mixes aspects of reality with computer-generated information, so it does not require as much capacity for abstraction as VR and people with autism who do not have abstraction capacity could benefit from their use (Herrera, 2006).

With regard to AR systems, researchers have proposed solutions in many different application domains. Some of them include applications intended for young children (Berry, 2006; Kerawalla, 2006), but a few of them are focused on disabled children (Richard, 2007).

In this paper we present "Pictogram room", a project developed by the Orange Foundation that intends to apply the latest technological advances of

Augmented Reality to benefit people with autism. The goal of the project is to help the child to improve critical abilities for his development. More specifically, the aim is to teach the individual with autism about self-awareness, body schema and postures, communication and imitation by means of an A.R. system.

Compared to other AR systems for individuals with ASD, this application takes the users' body movements as a method for interaction. Thus, moving their own body, children can play with the AR system. The purpose of the application is to teach the individual with ASD to develop his/her body schema and, after this, to go through a series of playful educational activities related to body postures, communication and imitation. These games show pictorial representations of various elements, including avatars representing the child and the teacher.

Therefore, our first task was to study existing MoCap technologies used to control virtual avatars with user's movements, looking for the one that fits better with our application, as described in section 2. Section 3 describes the system modules and how an augmented scene has been created by using Kinect data, and the implementation of an OSG shader that allows you to have virtual objects with real ones mixed at different depths is also described. Finally, testing results and conclusions are detailed.

2 STATE OF THE ART AND REQUIREMENTS.

Although traditionally MoCap systems for moving virtual actors have been used in Virtual Reality applications, with the growth of Augmented Reality (AR) applications, these systems are used to control a user in the real-world interacting with virtual objects in the augmented scene.

One of the earliest examples found is the ALIVE system (Maes, 1997), where a video camera captures images of a person in order to detect the user's movements but also to integrate the real image onto a virtual world. This can be considered a precursor of AR, where video images are integrated with 3D objects (like a dog) that are activated by the user's movements.

Several AR systems make use of markers (usually geometric figures) to detect positions (Mulloni, 2009). However, these systems do not provide enough accuracy to move a virtual character and all his joints. Sometimes optical systems with markers (Dorfmueller, 1999) are used to detect positions but, in other cases, more complex systems are used that are based on ultrasound or inertial sensors that detect movements in a wide range of use (Foxlin, 1998; Vlasic, 2007).

Regarding our system requirements, it is necessary to track the positions of different body parts in order to move an avatar representing the user. Moreover, taking into account the users to whom it is directed, who frequently experience sensory difficulties (Bogdashina, 2003) it is better that users do not wear complex devices. Thus, mechanical and electromagnetic and other invasive devices have been discarded from the beginning.

The release of the PrimeSense OpenNI for programming Kinect in December 2010, opened up new possibilities for us. Kinect incorporates an RGB camera (640x480 pixels at 30Hz) and a depth sensor. This is especially suitable to be used in an A.R. system where images from the real world are needed. The depth sensor provides distance information which is useful to create the augmented scene placing objects correctly.

Some researchers have started to use it in their applications to track people (Kimber, 2011), but there are no references of the use of this device in Augmented Reality applications to control avatars.

The PrimeSense OpenNI (OpenNI, 2011) provides information about positions and orientations of a number of skeleton joints, which can be used to control our virtual puppet. The first tests performed with this system were satisfactory

regarding motion capture. However, it has a negative aspect: the calibration requires the user to remain still for a few seconds in a certain posture in front of the camera. As the system is intended to train the individual with autism to match postures, as one of the final educational objectives, it makes no sense that he/she has to be able to copy a posture in order to enter the game. With the release of Microsoft's SDK the calibration problem was solved, this made us choose it as a solution for our system.

Thus, we have got a system that captures user positions in varying lighting conditions and without dress requirements or markers.

3 SYSTEM DESCRIPTION

After comparing the pros and cons of different motion capture devices a Pictogram Room using Kinect has been made. This system consists of a visualization screen measuring 3 x 2 meters, a projection or retro-projection system (depending on the room where it is to be installed), a PC, a Kinect device and speakers. Kinect is equipped with two cameras, an infrared camera and a video camera with a 640x480 pixel resolution, and a capture rate of 30 fps. So using Kinect it is possible to obtain not only a standard video stream, but also a stream of depth-images.

On the screen, images captured by Kinect are displayed mixed with virtual information, creating for users an augmented mirror where they can see themselves integrated onto the augmented scene.

The system is designed to be used by users playing with two different roles: child and educator. Each user is represented by a virtual puppet colored differently. Both users are located next to each other and they are tracked in the same space.

The task of the educator is to select exercises and activities that will be developed by the child, and after that to give him/her appropriate explanations. To achieve this, the teacher is provided with a menu system displayed on the screen which can be accessed by using the hand as a pointer. Once an exercise is selected, the child's and teacher's movements are captured and used as the interaction interface with the system.

This system has been implemented by creating a set of subsystems that deal with different tasks. On the one hand, the input system allows you to choose activities and capture the user's actions and movements. Depending on the activity, and the user's actions, the output system creates an augmented environment by integrating images from

the real world, video, sound and virtual elements. Finally, there is educational software to edit and manage exercises. Let's look at these three subsystems in more detail.

3.1 Input System

The input systems task is to collect information and it can work in two different modes. On the one hand, *edit mode* is intended to configure the system and to allow the educator to select educational activities. On the other hand, in *play mode* a child will develop the proposed activities. In both modes of functioning, there can be one or two users whose actions are detected by the system.

Kinect provides information regarding the positions of a number of user's joints. These positions are captured and transformed into positions and rotations applied to virtual skeletons.

In *edit mode*, Kinect provides the hand position which serves as an input pointer to select menus and push buttons.

In *play mode*, the user interacts with the system by moving his/her body and uses it for advancing throughout the series of educational activities.

3.2 Output System

This module is responsible for creating an augmented scene, taking input data coming from the capture devices and control messages sent by the exercise module.

In the augmented environment data coming from several sources are integrated: video images from the real world, virtual puppets representing users, virtual objects, recorded videos and music. Thus, users can see themselves in an augmented mirror projected in front of them, and they can interact with virtual objects.

Virtual avatars are controlled by means of joint positions acquired by the input system. As one of systems goals is to teach children to understand pictograms, these avatars have been designed as stick figures, and they are painted in the virtual world over user images. The augmented scene is created with OpenSceneGraph (OSG).

Depending on the exercise, bi-dimensional pictograms or even tri-dimensional objects can appear in the A.R. scene, and they can be located at different positions, even behind the user. With OSG it is easy to draw video images and virtual objects displayed over the video.

Nevertheless, if a virtual object must be placed at a certain depth, two things are needed: the z-position

of the users in the real world, and a mechanism to mix the real and virtual objects while maintaining the depth where they are located.

The problem of obtaining z-positions of users and objects in the real world is solved by Kinect. In addition to video images Kinect provides a depth map, which is a matrix where each pixel's depth is stored. With this map it is possible to create a drawing process maintaining the proper occlusion between real and virtual elements through the z-buffer.

The steps in this process (fig 1) are as follows:

1. Draw the image captured by the camera. This image reflects the real scene.
2. With the depth information obtained by the Kinect camera, it is possible to compute the z-buffer values related to each pixel of the real scene. This information will allow real objects to be placed in their proper position within the virtual world. To optimize the process a shader has been implemented, so operations are performed on the graphics card improving the process velocity.
3. Virtual objects are drawn and mixed with the real ones. To achieve this, a virtual camera must be configured with the same parameters as the real one. At this point, when drawing a virtual object, the z-buffer data written previously allows only those parts of the virtual objects that are in front of the real ones to be drawn.

Music Engine

Music therapy has been proved very useful for training individual with autism (Braithwaite, 1998). For this reason, in addition to the visual information, a music engine to generate musical sounds in the exercises has been made use of. This music engine can generate MIDI data in real time. In addition, MIDI stored files can also be modulated, so children's favorite songs can be loaded. With the information provided by Kinect about position and momentum, tempo and song volume can be modulated.

3.3 Educational Software

To configure the activities a management exercises module has been created. This module contains a set of classes to represent the different elements involved: virtual puppets, video images, sounds, virtual objects and pictograms and are described in an XML file.

The exercises are shown to the teacher through an interface where he/she can select the activities (see Figure 2).

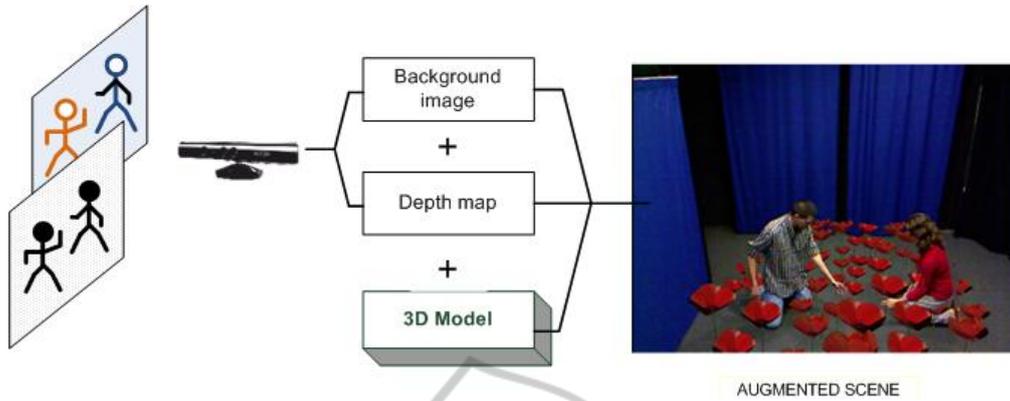


Figure 1: Scene composition using depth information.



Figure 2: Activities selection menus.

The activities are designed with the aim of facilitating the acquisition of certain skills in ASD children and the contents are classified into four sections with 20 activities each one, with two of them included in the first public version of our software: **body** and **postures**.

In the *learning about body* section, the activities are intended to develop the correspondence between oneself and the avatar schematic figure, to identify one own image, to discriminate between oneself and others and to identify different body parts. This has been developed overlaying the virtual avatar on the user's real image with different colors for each one.

Some of the activities designed involve users taking pictures of themselves and using these pictures later in the activity. Others consist of users touching pictograms that appear on the screen with different body parts (head, arms) (see Figure 3a). Other activities are designed to perform body movements. These movements are accompanied by musical notes dynamically generated by the MIDI engine, videos or lights appearing in the augmented scene always in a playful way.

The *body postures* section includes activities to

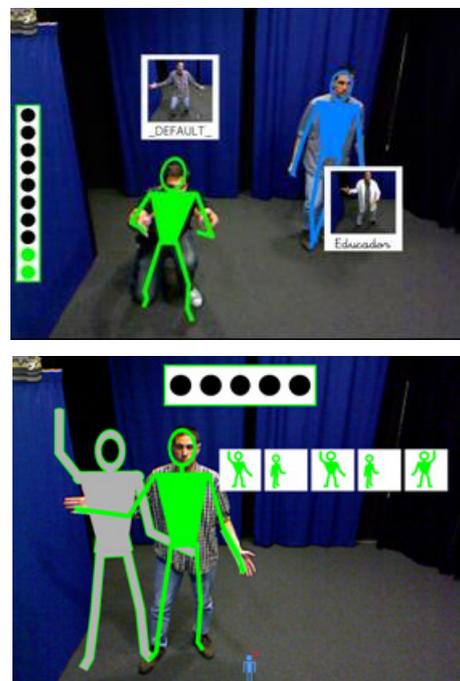


Figure 3: (a) Two users touching pictograms (b) Shadows guiding user.

train the user to match postures. For the design of these activities, in addition to the puppet representing the user, a shadow that guides the user has been created. This shadow is a grey colored stick figure and in some exercises it guides user's movements (See figure 3b).

Other times, instead of a full avatar, there are only certain parts of the body represented in the AR scene. This strategy is used to focus attention on those parts of the body. Some activities also make use of clipped images with a body outline in a posture that the user must imitate (see Figure 4).



Figure 4: Clipped images.

3.4 Interacting with the A. R. World

The system supports two users: the teacher and the student, who are represented in the augmented world by different virtual puppets. The activities are designed to be played by a child following the teacher's instructions and in some cases in a collaborative way. Moreover, users can enter and leave the scene at any time, being detected by the system without previous calibration.

In edit mode the teacher is responsible for selecting exercises to be executed. This can be done by means of a 2D menu system created with OpenSceneGraph and displayed on the screen (see Figure 2). To select menu options the teacher uses his/her hand to move a pointer represented with a hand icon. A button or menu option is activated when the hand is placed on it for a few seconds. This interaction with the 2D menu can also be done with a standard mouse.

Regarding the interaction in play mode, the exercises are designed to use body movements as a form of interaction. Thus, there are exercises that involve activities such as moving to different areas of the room, peering into the holes of the images or moving one's body into the indicated positions.

In this mode there are also two options to visualize the augmented world. One is showing the video images with the virtual objects superimposed on the display, and the other is eliminating the background and the user images and showing only the virtual figures and objects.

4 EVALUATION

In addition to functional testing, a set of assessment tests with users have been carried out.

The first tests consisted of a system assessment with 22 typically developed children, not having autism. These children were aged between 3 and 4.

For these tests, the system was installed in a school using a projection screen, a computer, speakers and a Kinect device. At the time the tests were conducted, Microsoft SDK was not yet released so they were done with OpenNI. As we have described in section 3, this library needs previous calibration, requiring the users to remain still in a certain position for several seconds (see Figure 5).



Figure 5: a) System calibration. b) Two users with their puppets.

Thus, as a first step, the children were asked to stay still for a few seconds in the calibration position. Then, the teacher gave explanations about the activities to be performed. Each child played for around 15 minutes, and he/she performed four different activities related to movements, touching objects or imitating postures.

When an activity had been completed the educator noted whether the child did it without support (3 points), with verbal (2 points) or physical (1 point) assistance from the educator, or was not able to do it (zero points). These tests with typically

developed children are useful to obtain reference values for each activity. With this normative information, each child with autism that uses the application can be compared with these reference values in order to have an idea of which skills have developed more and which have to be improved.

During the development of the activities the only problem that arose was the loss of calibration when a child went outside the system and re-entered. With regard to the activities, all the children were able to do them, although sometimes further explanations were necessary.

Following this, tests were performed in the same school with children with ASD. The tests were conducted with 5 children. One of them did not do any of the activities and could not even manage to do the posture calibration. Another child managed to do the calibration and began to play, but after a few minutes he stopped doing the activity and would not continue. The remaining three calibrated and played, some showing more skill than others. One of them played like the typically developed children (see Figure 6)

In future studies more test will be carried out with a great number of children with autism. Moreover, all the activities will be tested to find out if the activities are properly designed or if some aspects must be changed.

With regard to the calibration problems, as it was mentioned above, these have been solved with the use of Microsoft SDK which does not require calibration (i.e. it is not necessary for the child to match a given posture before entering the game).



Figure 6: System evaluation.

5 CONCLUSIONS

In this paper we have presented a system intended to improve the development of children with autism. The system uses Augmented Reality technologies.

The users are represented by means of virtual stick figures which have been implemented with a motion capture system that meets the two main requirements; it is a low cost system, and it does not require users to wear any device.

After the initial steps to test the feasibility of the system on a CAVE, the motion capture was developed based on vision algorithms for motion detection. Finally, the release of the Kinect SDK gave us a tool appropriate to our needs. From the technical point of view, the problem of integrating objects at different depths in the virtual world has been solved by programming an OSG shader that uses the depth map provided by Kinect.

With regard to the contents, a wide range of activities have been developed. These activities include images, videos, sounds and virtual objects integrated in the augmented world where children can play.

Therefore, the possibility of using Kinect to develop this type of application aimed at children with autism has been shown.

REFERENCES

- Berry R., Makino, M., Hikawa, N., Suzuki, M., Inoue, N., 2006. Tunes on the table. *Multimedia Systems*, 11 (3): 280-289.
- Bogdashina, O. 2003. Sensory Perceptual Issues in Autism and Asperger Syndrome Different Sensory Experiences Different Perceptual Worlds. Jessica Kingsley Publishers: London UK.
- Braithwaite, M., & Sigafos, J. 1998. Effects of social versus musical antecedents on communication responsiveness in five children with developmental disabilities. *Journal of Music Therapy*, 35(2), 88-104.
- Dorfmueller, K., 1999. Robust tracking for augmented reality using retroreflective markers. *Computer & Graphics*. Volume, 23, 6, 795-800.
- Foxlin, E., Harrington, M., Pfeifer, G., 1998. Constellation: A Wide-Range Wireless Tracking System for Augmented Reality and virtual set applications. *Proceedings of SIGGRAPH '98*.
- Vlasic D. et al., 2007. Practical motion capture in everyday surroundings. *ACM Transactions on Graphics*, Vol. 26, 3.
- Herrera, G; Jordan, R; Gimeno, J, 2006. Exploring the advantages of Augmented Reality for Intervention in ASD. *Proceedings of the World Autism Congress, Southafrica*.
- Herrera, G., Alacantud, F., Jordan, R., Blanquer, A., Labajo, G., 2008 Development of symbolic play through the use of virtual reality tools in children with autistic spectrum disorders *SAGE Publications and The National Autistic Society*, Vol 12(2) 143-157.
- Kerawalla, L., Luckin, R., Seljeflot, S., Woolard, A.,

2006. Making it real: exploring the potential of augmented reality for teaching primary school science. *Virtual Reality*, 10, 163-174.
- Kimber, D., Vaughan, J., Rieffel, E., 2011. Augmented Perception through Mirror Worlds. Conference AH'11.
- Maes, P., Darrell, T., Blumberg, B., Pentland, A., 1997. The ALIVE system: wireless, full-body interaction with autonomous agents. *Multimedia Systems* 5: 105-112.
- Mulloni, A. 2009. Indoor Positioning and Navigation with Camera Phones. *Pervasive Computing*, 8, 2, 22-31.
- OpenNI TM. <http://www.openni.org/> Visited: 2011.
- Richard, E., 2007 Augmented Reality for Rehabilitation of Cognitive Disabled Children: A Preliminary Study. *Virtual Rehabilitation*. pp.102-108.
- Strickland, D. 1997. Virtual reality for the treatment of autism. In *Virtual Reality in Neuro-PsychoPhysiology*. G. Riva, Ed. IOS press. Chapter 5, pp. 81-86.

