

Augmented Reality Object Selection User Interface for People with Severe Disabilities

Julius Gelšvartas¹, Rimvydas Simutis¹ and Rytis Maskeliūnas²

¹*Automation Department, Faculty of Electrical and Electronics Engineering, Kaunas University of Technology, Studentų st. 50-154, Kaunas, Lithuania*

²*Centre of Real Time Computer Systems, Kaunas University of Technology, Baršausko st. 59-338a, Kaunas, Lithuania*

Keywords: Augmented Reality, Projection Mapping, Calibration, Structured Light, Projector.

Abstract: This paper presents a user interface that was designed for people with severe aging related conditions as well as mobility and speech disabilities. Proposed user interface uses augmented reality to highlight objects in the environment. Augmented reality is created using a projection mapping technique. Depth sensor is used to perform object detection on a planar surface. This sensor is also used as part of a camera-projector system to perform automatic projection mapping. This paper presents the user interface system architecture. We also provide a detailed description of the camera-projector system calibration procedure.

1 INTRODUCTION

People suffering from severe age related conditions such as dementia can experience mobility problems and speech pathologies. People suffering from tetraplegia can also experience similar symptoms. These conditions drastically impact patient's ability to communicate. Moreover, these patients require some form of constant care that is both time consuming and expensive. Often assistive technology is the only way these patients can communicate. Efficient assistive technologies have the potential of improving patient's quality of life and sometimes even reducing the need for care.

Assistive technology is a very broad term covering adaptive, rehabilitative and assistive devices that help people perform tasks that they were formerly unable to accomplish. Such system usually consist of two parts, namely assistive devices and human computer interaction (HCI) interface. Unfortunately, assistive devices that are on the market today are not very efficient. The efficiency of such devices can be improved by developing specialized user interface (UI).

The primary use case of our system is selection and manipulation of objects that are placed on a tabletop in front of the user. UI presented in this paper consists of projector, depth camera (such as Microsoft Kinect) and user action input device. Our system can be used with several user action input devices, namely sip/puff, consumer grade brain computer in-

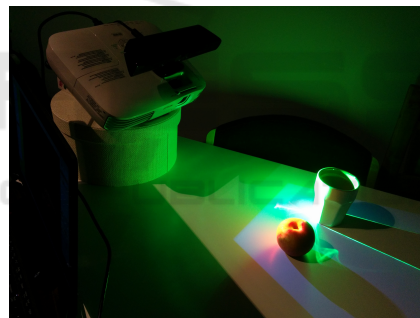


Figure 1: Experimental setup and the example of high-lighted object.

terface (BCI) and eye tracker devices. Depth camera is used to automatically detect tabletop plane and objects positioned on that plane. Finally, projector is used together with depth camera to create a camera-projector system that performs automatic projection mapping. This makes it possible to highlight all detected objects. The user then selects the desired object and it can be manipulated either by care giver or a robotic arm. For example, when the user selects a glass of juice it is grabbed by a robotic arm and brought to users mouth so that he can drink from it.

The main goal of the assistive technology is to either restore or substitute an ability that a patient has lost. BCI is the best technology for restoring natural abilities, because HCI is performed by thinking. BCI research field has had significant advances in recent years. It has been shown in (Hochberg et al.,

2012) that BCI systems can be used to directly control robotic arms. Similar system has been used for patient's limb control (Ajiboye et al., 2017). These advances are, however, made with expensive and usually invasive BCI devices. Consumer-grade BCI devices still have very limited capabilities (Maskeliunas et al., 2016).

Systems that do not use direct limb or robotic arm control require UI to operate. HCI efficiency directly affect system usability. A standard way of presenting UI to the users is by using conventional displays. This approach is inefficient when the user is communicating or interacting with the environment. Augmented and virtual reality have successfully been used in assistive technologies and rehabilitation (Hondori et al., 2013). Augmented reality is better suited for information presentation, because virtual reality headset would further separate locked in patient from the environment. Projection mapping is one of the best ways to display augmented reality UI.

Projection mapping can either be manual or automatic. Manual projection mapping is mostly used in entertainment and art industries where the scene is static. The system presented in this paper uses automatic projection mapping, because the environment is dynamic and objects can change their positions. Automatic projection mapping can be performed when the transformation between depth camera and projector optical frames is known. This transformation is obtained by calibrating camera-projector system. One way to calibrate camera-projector system is by using structured-light (Moreno and Taubin, 2012). Alternatively, method proposed in (Kimura et al., 2007) can be used when the camera is already calibrated. This paper utilized a practical calibration method proposed in (Yang et al., 2016).

The system presented in this paper is similar to (Benko et al., 2012), but without the accounting for deformations caused by physical objects. More advanced dynamic projection mapping methods have been created in recent years (Sueishi et al., 2015). Such systems require more expensive hardware setup. Figure 1 shows the experimental setup of the presented system.

The remaining paper is structured as follows. Section 2 describes the proposed projection mapping based system. The results and discussions are presented in section 3. Section 4 is the conclusion.

2 MATERIALS AND METHODS

The augmented reality UI is constructed and presented using a camera-projector system. The work

flow for setting up and calibrating the projection mapping system is as follows:

1. The depth camera and projector are setup in front of the scene. The projector and camera should be fixed sturdily to each other. Ideally the camera and projector should be fixed to a common metal frame or integrated into one housing. This is necessary so that the extrinsic parameters calculated during calibration do not change when the system is operating. If the camera-projector system would be integrated into a single device the calibration process could be performed only once i.e. factory calibration.
2. Calibration of projector-camera system is performed by placing a board with circular black dot pattern in front of projector and camera. The projector is used to show a similar white dot pattern that appears on the same board. The camera image of the board is recorded and positions of projected and real dots are estimated. To correctly estimate the system parameters several images with varying board position and orientation have to be captured. The estimates are used to calculate the intrinsic and extrinsic parameters of the projector-camera system.

In the camera-projector system a projector is treated as a virtual camera device. We use a pinhole camera model to describe both the camera and the projector (i.e. a virtual camera). The pinhole camera intrinsic parameters consist of a 3×3 camera matrix C and a 1×5 distortion coefficients matrix D . Intrinsic parameter matrices C and D can be combined to create a 3×4 camera projection matrix P (Hartley and Zisserman, 2003). Matrix P can be used to project 3D world points in homogeneous coordinates into an image. During calibration we obtain two camera projection matrices P_c for camera and P_p for projector.

Camera-projector system also has extrinsic camera parameters. Extrinsic camera parameters consist of translation vector T and rotation matrix R . In our case T and R define the translation and rotation of the projector optical origin in the camera origin coordinate system. After calibration camera-projector system intrinsic and extrinsic parameters are obtained. These parameters are used to perform automatic projection mapping during system operation.

During the system operation depth sensor is acquiring depth images of the scene in camera optical frame coordinate system. These images have to be transformed into a projector optical frame coordinate system. This transformation consists of the following

three steps:

1. The depth image is converted to a 3D point cloud. This is achieved by re-projecting each depth image pixel (u, v, Z) to 3D point $(X, Y, Z, 1)$ in depth camera frame. Here u and v are pixel coordinates along image rows and columns. X and Y are 3D point coordinates along X and Y axis in meters. Depth image pixel values are already in meters and are copied to 3D point Z axis coordinate. The transformation is performed using the camera intrinsic parameter matrix P_c obtained during system calibration.
2. 3D point cloud transformation. Each 3D point is transformed to projector frame by multiplying this point by transformation matrix that is constructed from camera-projector system extrinsic parameters i.e. rotation matrix R and translation T .
3. Projecting point cloud into projector frame. Every transformed 3D point is projected into projector frame by multiplying them with projection matrix P_p . P_p is the projector device projection matrix that is obtained after calibrating camera-projector system.

The transformed depth images are used for object detection. After object detection pipeline finishes the transformed depth image pixels that belong to a detected object are used in projection mapping.

3 RESULTS AND DISCUSSION

The camera-projector system described in the previous section has been used to create an object selection UI. The system architecture of this UI application can be seen in Fig. 2. UI implementation contains the following components:

1. **Depth sensor driver** is used to acquire depth images from Kinect sensor. Note that other depth sensors or stereo cameras can easily be integrated into our system.
2. Tabletop detector from **Object Recognition Kitchen** package (Willow Garage, 2017) is used to perform object detection. The detector uses depth image that has been transformed to the projector optical frame. The object detector used in this paper has two parts, namely a table finder and an object recognizer. The tabletop is detected as a dominant plane in the depth image using the algorithm presented in (Poppinga et al., 2008). All points that are above the tabletop plane are used for clustering to identify individual objects. The

remaining points are discarded. Object point clusters are used to highlight the objects with projector. Each cluster is also compared to object model database to determine the type of object. At the moment we are not using object type and database mesh, however this information could be used to render more accurate object highlights.

3. **User action input modality** is a package that provides an interface used to abstract various action input devices. This interface currently supports two actions, namely selecting current object and moving to next object in the list. Some devices only have one action, in this case the moving to next object is performed automatically at given intervals and the user only has to select the desired object.

This package has a well defined interface that can be used to integrate additional assistive devices into the system. Current system supports sip/puff devices, eye-trackers and consumer grade BCI devices.

4. **Object selector** keeps track of all detected objects. The object list is updated each time a new depth image is processed. The update process removes objects that have been removed from the scene and adds new object when they appear in the scene.

This package also maintains an index of the current object. The current object index is updated by actions from **User action input modality** package.

Finally, this package can also generate external signals, when a particular object is selected. The external signal can be used to display information for caregiver. This can be done by changing the color of highlighted object or showing notifications on a separate screen. The system could also be integrated with additional assistive technologies. For example, generated external signal could be used by a robotic arm controller that grasps an object and gives it to the user.

5. **Object highlight projector** receives a list of points belonging to each detected object. These points are used to create a projection mapping image that is shown by the projector. The index of current object is received from **Object selector** and this object is highlighted in different color. The detailed description of projection mapping algorithm can be found in section 2.

The proposed system could further be improved by calculating object highlights from 3D models. The highlight visibility could be improved by projecting

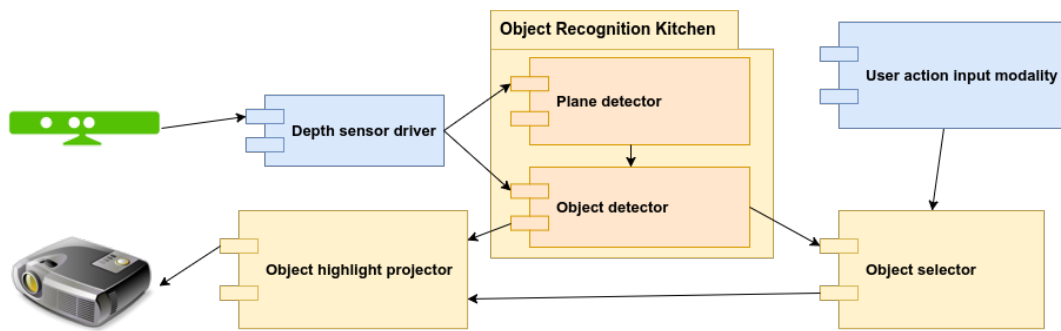


Figure 2: Object selection user interface system diagram.

animated highlights. Another option is to create highlights that account for object texture instead of using single color highlight for the whole object.

Automatic projection mapping produces accurate object highlights. Highlight visibility can sometimes be reduced when the scene contains shiny, reflective or transparent objects. The easiest way to overcome these problems is to make sure that the scene contains only diffuse objects. The proposed system will operate in a controlled environment and the correct object choice is performed by the system setup personnel.

The proposed system could further be improved by using a more advanced object detector. The main requirement for an object detector is that it should detect objects in disparity maps or 3D point clouds produced by depth camera.

4 CONCLUSIONS

This paper presents a UI that uses projection mapping to highlight objects in a scene. Our system is specifically designed for patients with severe disabilities. Detailed description of the system architecture is provided. Projecting object highlights onto real objects creates a more natural user experience, because the user can interact with assistive UI by looking directly into a scene. The main disadvantage of using projection mapping system is that projected light might not be visible well in very bright environments. This problem can be partially solved by having more powerful projector.

The ideas presented in this paper could also be used with head-up display (HUD) systems. HUD would provide a more natural UI and would help to avoid problems in bright environments. Available consumer grade HUD systems, however, are very expensive, whereas projectors are widespread and affordable.

ACKNOWLEDGEMENTS

This research has been supported by European cooperation in science and technology COST Action CA15122, Reducing Old-Age Social Exclusion: Collaborations in Research and Policy (ROSEnet).

REFERENCES

- Ajiboye, A. B., Willett, F. R., Young, D. R., Memberg, W. D., Murphy, B. A., Miller, J. P., Walter, B. L., Sweet, J. A., Hoyen, H. A., Keith, M. W., et al. (2017). Restoration of reaching and grasping movements through brain-controlled muscle stimulation in a person with tetraplegia: a proof-of-concept demonstration. *The Lancet*, 389(10081):1821–1830.
- Benko, H., Jota, R., and Wilson, A. (2012). Miragetable: freehand interaction on a projected augmented reality tabletop. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 199–208. ACM.
- Hartley, R. and Zisserman, A. (2003). *Multiple view geometry in computer vision*. Cambridge university press.
- Hochberg, L. R., Bacher, D., Jarosiewicz, B., Masse, N. Y., Simeral, J. D., Vogel, J., Haddadin, S., Liu, J., Cash, S. S., van der Smagt, P., et al. (2012). Reach and grasp by people with tetraplegia using a neurally controlled robotic arm. *Nature*, 485(7398):372.
- Hondori, H. M., Khademi, M., Dodakian, L., Cramer, S. C., and Lopes, C. V. (2013). A spatial augmented reality rehab system for post-stroke hand rehabilitation. In *MMVR*, pages 279–285.
- Kimura, M., Mochimaru, M., and Kanade, T. (2007). Projector calibration using arbitrary planes and calibrated camera. In *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*, pages 1–2. IEEE.
- Maskeliunas, R., Damasevicius, R., Martisius, I., and Vasiljevas, M. (2016). Consumer-grade eeg devices: are they usable for control tasks? *PeerJ*, 4:e1746.
- Moreno, D. and Taubin, G. (2012). Simple, accurate, and robust projector-camera calibration. In *3D Imaging*,

- Modeling, Processing, Visualization and Transmission (3DIMPVT), 2012 Second International Conference on*, pages 464–471. IEEE.
- Poppinga, J., Vaskevicius, N., Birk, A., and Pathak, K. (2008). Fast plane detection and polygonalization in noisy 3d range images. In *Intelligent Robots and Systems, 2008. IROS 2008. IEEE/RSJ International Conference on*, pages 3378–3383. IEEE.
- Sueishi, T., Oku, H., and Ishikawa, M. (2015). Robust high-speed tracking against illumination changes for dynamic projection mapping. In *Virtual Reality (VR), 2015 IEEE*, pages 97–104. IEEE.
- Willow Garage, R. c. (2017). Ork: Object recognition kitchen. https://github.com/wg-perception/object_recognition_core.
- Yang, L., Normand, J.-M., and Moreau, G. (2016). Practical and precise projector-camera calibration. In *Mixed and Augmented Reality (ISMAR), 2016 IEEE International Symposium on*, pages 63–70. IEEE.

