BulkScreen: Saliency-Based Automatic Shape Representation of Digital Images with a Vertical Pin-Array Screen

Riku Arakawa*
Yudai Tanaka*
Hiromu Kawarasaki
Kiyosu Maeada
{arakawa-riku428,tanaka-yudai327,rmhkwar0518,kiyosu775}@g.ecc.u-tokyo.ac.jp
The University of Tokyo

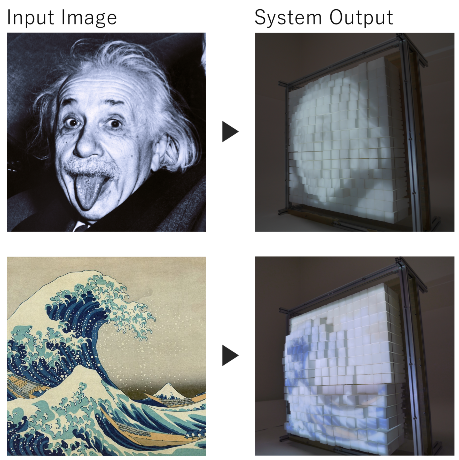

Figure 1: Image projection with BulkScreen. Upper: The portrait of Albert Einstein. Lower: The Great Wave by Hokusai.

*Both authors contributed equally to this research.

ABSTRACT

Digital images appearing on displays in everyday activities (e.g., photos on a smartphone) are automatically and instantly rendered without manual intervention such that we can seamlessly appreciate them. In contrast, shape displays require manual designs of outputs upon actuation of input images to render 3D shapes. In this work, we aim to achieve automatic and on-the-spot actuation of digital images so that we can seamlessly see 3D physical images. To this end, we developed *BulkScreen*, an image projection system that can automatically render 3D shapes of input images on a vertical pin-array screen. Our approach is based on a deep-neural-network saliency estimation coupled with our post-processing algorithm. We believe this spontaneous actuation mechanism facilitates applications

TEI ’20, February 9–12, 2020, Sydney, NSW, Australia

ACM ISBN 978-1-4503-6107-1/20/02.
https://doi.org/10.1145/3374920.3374973

with shape displays such as real-time picture browsing and display advertisement, building on the benefit of representing physical shapes; tangibility.

CCS CONCEPTS

• **Human-centered computing** → **Displays and imagers**; • **Hardware** → *Emerging interfaces*; • **Computing methodologies** → Shape representations.

KEYWORDS

shape display, image projection, saliency detection

ACM Reference Format:

Riku Arakawa, Yudai Tanaka, Hiromu Kawarasaki, and Kiyosu Maeada. 2020. BulkScreen: Saliency-Based Automatic Shape Representation of Digital Images with a Vertical Pin-Array Screen. In *Fourteenth International Conference on Tangible, Embedded, and Embodied Interaction (TEI '20), February 9–12, 2020, Sydney, NSW, Australia.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3374920.3374973

INTRODUCTION

Shape display, dynamic actuation of digital contents with physical objects such as square pins [3, 11] enhances presence of digital matters by providing their 3D physical existence, enabling tangible interaction (e.g., touching the shapes of rendered images). However, while we can seamlessly see images on our smartphones or laptops with autonomous and instant GUI rendering process, shape displays require manual designs of output shapes consisting of a large number of 3D pixels before showing images. Since this process is often time consuming, the displays lack the flexibility of rendering varying input images on the spot. In addition, previous shape displays were table-top surfaces, limiting its use cases to scenarios such as 3D map manipulation and CAD modelling. This restriction hinders further applications of shape displays requiring vertical placement such as image showcasing.

To address the above issues, we propose BulkScreen, a vertical screen that can automatically render 3D shapes with 16×16 pin array. A deep learning model first captures the saliency feature of an input image. Subsequently, our post-processing algorithm convert the estimated saliency into the depth information for each pin. Then, based on the depth data, 256 motorized pins are actuated dynamically to depict the 3D shape, while the original image is projected onto the pin-array screen. Our system has two novelties: i) automatic processing of 2D images to render 3D voxelized shapes based on a deep learning model; ii) a mechanical structure of motor-actuated pins to enable a vertically transforming flat to exist. Our automatic rendering system enables user experience of appreciating various physically shaped digital images on the vertical surface without manually designing output shapes.

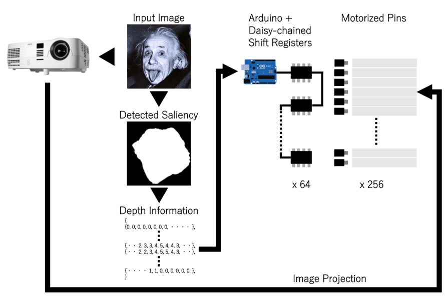

Figure 2: Architecture of the system.

BACKGROUND

To enhance the presence of 2D visual contents, technologies for rendering 3D images have been widely investigated. Stereoscopic rendering is by far the most established and widespread approach to realize 3D images in the physical environment. There are some off-the-shelf holographic displays based on the technology [2, 6]. As a different method to render shapes of digital visual contents, “shape displays” have been actively investigated. Though their image resolution is low, those devices enable tangible interaction such as touching and directly manipulating 3D images, which stereoscopic methods cannot achieve. A famous early work in this literature is FEELEX by Iwata et al. [7], where 36 motorized pins are used to actuate the shapes of graphics projected onto the soft surface. Ishii et al. developed a high resolution (30×30) shape display that is fast and accurately actuated [3]. However, these devices cannot actuate the shapes of projected visual images without preparation in advance, lacking the ability to automatically detect the shape features of input data in real-time. On the other hand, methods for real-time depth estimation of 2D images have been widely explored in the computer vision field [4, 9]. As a first step to combine such a computational algorithm with a shape display and to explore potential applications, we develop an automatic rendering and actuation mechanism for the shape display based on a deep learning model.

AUTOMATIC DESIGN OF OUTPUT SHAPES

As discussed in the “BACKGROUND” section, what distinguishes BulkScreen from other shape displays is the automated pipeline for determining each pin’s depth. This is achieved through a salient object detection process which is used for identifying visually distinctive areas in an image as a branch of image segmentation [1]. In our system, a pre-trained Convolutional Neural Network [5] converts an input image into a saliency map so that each pixel value represents an estimated saliency level (Figure 2). Indeed, there is an assumption that making those salient pixels 3D is a suitable approach for reflecting the characteristic of an input image in the 3D shape.

Once a saliency map is obtained, it is binarized and resized to 16×16 pixels, each of which corresponds to a pin of the screen. The discrete depth of each pin is then defined so that the center part of highly salient pixels becomes convex (Figure 2). In detail, the depth of the edge pixels of the salient part is set to be level 1 and an inner part’s depth is incrementally increased up to level 8 as its location becomes center. This heuristic is derived from an intuition that people’s attention is likely to turn to the central part of each object and therefore we posit that making those pixels close to users could enhance presence of an image. Note that all of the above processes run without delay in a conventional CPU environment, enabling BulkScreen’s prompt response to input images.

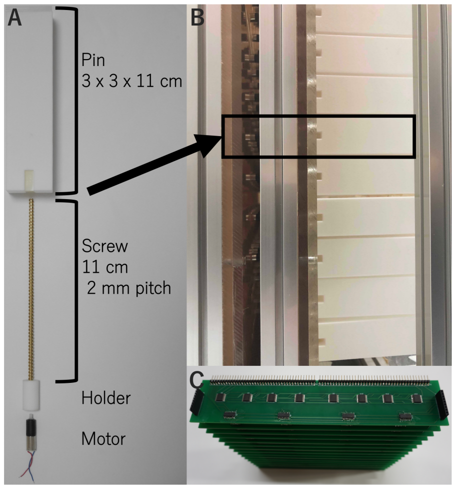

Figure 3: A: A motorized pin. B: Motorized pins embedded in the screen frame. C: The circuit server with 16 PCBs.

HARDWARE IMPLEMENTATION

To achieve autonomous actuation of the vertical screen, we developed a controlling and supporting structure for 16 × 16 arrayed pins. Referring to the method depicted in Dynablock [10], we designed a motorized pin where each pin is connected to an actuation unit (Figure 3 (A)) consisting of a DC motor (TGPP06-D136: 240 rpm with 1:136 gear ratio), a 3D printed motor holder, and an off-the-shelf lead screw (2 mm pitch, 1 start, 11 cm length). All units are embedded in the 60 × 60 cm vertical supporting grid comprising aluminum frames and acrylic boards (Figure 3 (B)). To control all units simultaneously, we designed a circuit server with 16 PCBs (Figure 3 (C)). Each PCB comprises four shift registers (74HC5959) and eight motor drivers (TB6612FNG). Since one driver controls two motors and one shift register sends individual signal to two drivers, a PCB unit is capable of controlling 16 motors simultaneously. Thus, a circuit server is able to actuate 256 pins.

As illustrated in Figure 2, in each image rendering process, the Arduino UNO transmits a serial signal, conveying the arrayed depth data from the algorithm to the shift registers. Through the shift registers, the serial input is transformed to each motor driver in parallel, which controls motors through pulse-width modulation (PWM). Since the lead screw travels roughly 8 mm per second based on the dc motor’s rotation speed of 240 rpm and the eight discrete depth levels are arranged in equal intervals of 10 mm, each pin approximately takes 1.25 × (depth level) seconds to reach to the designed location. After the actuation of one image, the system sends the signal to move all pins to the flat position, preparing for the next rendering process.

FUTURE WORK

The presented system is work-in-progress. In this section, we discuss the expected future work in two directions: the evaluation of users’ perception toward BulkScreen’s automatically-designed shapes and the exploration of its potential applications.

Evaluation Direction

The design of BulkScreen, especially its saliency-based automatic shape representation of 2D digital images involves an important but still unexplored topic in the shape display research field: *What is the best 3D shape representation for 2D visual contents?* Regarding this point, there are two assumptions in the current implementation, which we mentioned in the “AUTOMATIC DESIGN OF OUTPUT SHAPES” section. One is the use of saliency models in order to reflect image characteristics and the other is the heuristic of making the salient pixels convex in pin-array screen. These approaches, however, cannot be applied universally to a wide variety of visual contents. For example, in our pilot demonstration of BulkScreen at South by Southwest 2019 (https://www.sxsw.com/), we observed that some visitors responded positively toward a photo of human face voxelized in this manner by saying something like “The face of a person can be seen floating with a sense of reality”. On the other hand, there are a few visitors who pointed out its strange perspective when watching the output of a landscape picture. To further dig into this topic, it is necessary to apply different types of representation rules on a variety of visual contents (human face, landscape, geometric shape, etc.), the perception of which shall be measured through user study. We believe BulkScreen’s framework will enable this investigation.

Application

We believe BulkScreen can enhance our interaction with visual contents in various applications, leveraging its automatic shape representation and vertical screen. Currently, we have developed a dynamic 3D picture frame as a first application. Figure 1 depicts the system rendering 3D shape features of various types of images with original images projected (portrait: Einstein (Upper) and drawing: Hokusai (Lower)). As the system is capable of actuating a variety of input images with automatic designing of the shapes, users can enjoy a transition of their favorite pictures projected onto BulkScreen with their tangibility and enhanced presence.

For a potential application, by expanding its screen spatially with more pin components, BulkScreen could be installed as a display advertisement as the system can contribute to the improvement of exhibited products’ presence with real 3D pixels. Another direction is to improve the speed for actuating each pin. This could be realized by replacing the actuators with more powerful ones, such as those used in [3]. In this manner, BulkScreen can be utilized for situations requiring quick interactions. For example, we envision a future video calling app using BulkScreen where the display’s shape will be transformed in accordance with the movements of the other user in real-time, anticipating remote conversations [8] with an enhanced sense of presence.

CONCLUSION

In this work, we proposed BulkScreen which combines a saliency estimation algorithm and a vertical pin-array shape display. The system achieves automatic design of the 3D shape and its physical actuation after loading an input digital image. We believe our automatic shape rendering system enables spontaneous manner of seeing represented shapes of digital images where the rendering process does not involve human workload. The system will broaden interaction with 3D images rendered by shape displays, which we will explore in the future.

ACKNOWLEDGEMENT

The authors are grateful to Hiro Watanabe for his insightful discussions and financial support on this project. The authors also thank Rina Shinohara for her help with assembling the proposed device.

REFERENCES

[1] Radhakrishna Achanta, Francisco Estrada, Patricia Wils, and Sabine Süsstrunk. 2008. Salient region detection and segmentation. In *International conference on computer vision systems*. Springer, 66–75.

[2] LOOKING GLASS FACTORY. 2018. The Looking Glass. Retrieved July 6, 2019 from https://lookingglassfactory.com/

[3] Sean Follmer, Daniel Leithinger, Alex Olwal, Akimitsu Hogge, and Hiroshi Ishii. 2013. inFORM. In *Proceedings of the 26th annual ACM symposium on User interface software and technology - UIST '13*. ACM Press. https://doi.org/10.1145/2501988.2502032

[4] Clement Godard, Oisin Mac Aodha, and Gabriel J. Brostow. 2017. Unsupervised Monocular Depth Estimation with Left-Right Consistency. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE. https://doi.org/10.1109/cvpr.2017.699

[5] Qibin Hou, Ming-Ming Cheng, Xiaowei Hu, Ali Borji, Zhuowen Tu, and Philip H. S. Torr. 2019. Deeply Supervised Salient Object Detection with Short Connections. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 41, 4 (April 2019), 815–828. https://doi.org/10.1109/tpami.2018.2815688

[6] H+Technology. 2015. Holus. Retrieved July 6, 2019 from https://hplustech.com/

[7] Hiroo Iwata, Hiroaki Yano, Fumitaka Nakaizumi, and Ryo Kawamura. 2001. Project FEELEX. In *Proceedings of the 28th annual conference on Computer graphics and interactive techniques - SIGGRAPH '01*. ACM Press. https://doi.org/10.1145/383259.383314

[8] Abigail Sellen. 1995. Remote Conversations: The Effects of Mediating Talk With Technology. *Human-Computer Interaction* 10, 4 (Dec. 1995), 401–444. https://doi.org/10.1207/s15327051hci1004_2

[9] Shuran Song and Jianxiong Xiao. 2016. Deep Sliding Shapes for Amodal 3D Object Detection in RGB-D Images. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE. https://doi.org/10.1109/cvpr.2016.94

[10] Ryo Suzuki, Junichi Yamaoka, Daniel Leithinger, Tom Yeh, Mark D. Gross, Yoshihiro Kawahara, and Yasuaki Kakehi. 2018. Dynablock. In *The 31st Annual ACM Symposium on User Interface Software and Technology - UIST '18*. ACM Press. https://doi.org/10.1145/3242587.3242659

[11] Junichi Yamaoka and Yasuaki Kakehi. 2017. ProtoMold. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI '17*. ACM Press. https://doi.org/10.1145/3025453.3025498