Real-time, anthropomorphic 3-D scanning and voxel display system using consumer depth cameras as an interactive means of individual artistic expression through light

Andrei-Ducu Predescu[1] and *Georgios Triantafyllidis*[1]

[1]AAU Copenhagen, Department of Architecture Design and Media Technology, 2450 A.C Meyer Vænge 15, Copenhagen SV, Denmark

Abstract. The research presented in this paper revolves around the development of an interactive light installation called NEO-David. The focus is on the development of kinetic light within the boundaries of real-time generated anthropomorphic form. The case study seeks to address the issues related to democratic aspects of art and participatory artistic development. The paper presents the setup of such a system and explores different technical development challenges of the design.

1 Introduction

The paper aims to establish a theoretical framework for the development of a system capable of recreating three-dimensional properties of an individual within the boundaries of a three-dimensional display area. Furthermore, the system's goal is to create real time content as opposed to already available pre-generated information. The paper describes the research, development methodology and prototyping procedure of the NEO-David light art installation, which main goal is the development of a possibility in representing an anthropomorphic abstraction of an individual, while at the same time being capable of "mirroring" the movement generated by the user interacting with the installation.

1.1 Concept

The idea behind the installation is based on an architectural concept known as the Musicon Masterplan Development. The flagship element of the masterplan is represented by the Ragnarock Museum for Pop, Rock and Youth Culture. The museum has tried to embody the metaphor of a rockstar's lifestyle [1]. However, certain elements of this metaphor have not managed to survive the cruel reality of budget austerity. The NEO-David Installation aimed to be a conceptual prosthesis, trying to complete the original concept idea. The NEO-David has thus become part of the architectural ensemble by providing support in "celebrating" the individuals coming to visit the Ragnarock Museum; hence, conceptually, it tries to mimic the attention given by the public to the "rockstar". By creating a sculptural element that takes the average individual and transposes her or him to a "stage" that gratifies and offers a temporary

axis-mundi status to everyone, the conceptual red thread of the Ragnarock regains wholesomeness and strengthens the idea of democratic space and freedom of expression within the overall Musicon Masterplan identity.

1.2 Theoretical development

To develop a system that is capable of reproducing a real-time, three-dimensional animated model, we need to first understand the capabilities of today's scanning technologies. One of the most accessible products on the market with such capabilities is the Kinect, device, classified as a motion sensor input device. Originally developed by Microsoft in 2010, as a means of interaction with the XBOX 360 gaming platform, it has been quickly embraced by the MAKER community and open source software was developed to access the data generated by this device. [2]

The Kinect can generate a point cloud of the environment it "sees", which, through processing, can later be translated into a real-time tree-dimensional model of the environment. However, the generated model is an incomplete representation of the environment, having numerous artefacts as well as occlusion issues, where the device is incapable of detecting surfaces. The major issue with using one device lies in the fact that the generated model is not a full three-dimensional object, but it is more similar to a classical relief. Hence, even though the model can update real-time it is incapable of generating a fully three-dimensional output that can later be used to populate a three-dimensional display area. However, this paper will try to address this issue through theoretical means.

The second part of the project has to deal with the display area of the model. In order to understand what the output methodology will be we firstly have to grasp how an object manifests itself within physical space. An object found in the physical or digital realms presents itself as a compact collection of voxels. Each of these individual voxels has a series of properties, such as dimension and a coordinate system relative to its world coordinates. An accumulation of voxels results in the generation of a form. Using basic geometric principles, we can generate a form through the appropriate distribution of voxels which can be directly translated within a display area, which has identical voxel representation parameters, such as voxel size and spacing of the composing elements. [3]

2 Design

Recreating the form of an individual through light is a complex issue, mainly having to do with the voxel resolution of the form. When dealing with the visual perception of a form, experimentation regarding the appropriate spacing of the voxel arrays is critical. After this first step has been completed, considerations for the spatial qualities of the installations have to be investigated. Lastly, the scanning area has to be established and the software dealing with compilation of the individual cloud points and merger into one continuous three-dimensional model has to be developed.

2.1 Spatial properties

Having to deal with the human body as an interaction element, the installation area has to be able to accommodate the physical dimension and interaction space, which is within the physical grasp of an individual.

The Modulor is one of the most well-known studies in the field of design. Developed by a Swiss-French architect Corbusier, it proposes a unifying scale between the imperial and metric system by adopting a methodology which implies using the human body as a means

to impose scale upon objects and spaces. Even though in many ways it is a system that could seem obsolete or outdated, it has a significant advantage, having as a basis a highly humanistic approach to design, while also putting the individual at the centre of rationalisation as opposed to an aleatory mechanical like establishment of dimensions. [4]

Based on this investigation, it was concluded that the optimal spatial properties of the installation, which are able to accommodate a large range of the movement manifested through the anthropomorphic form generated, are as follows: a height of 2.26m, a width of 2.26m and a depth of 1.13m, as shown in Figure 1. [4]
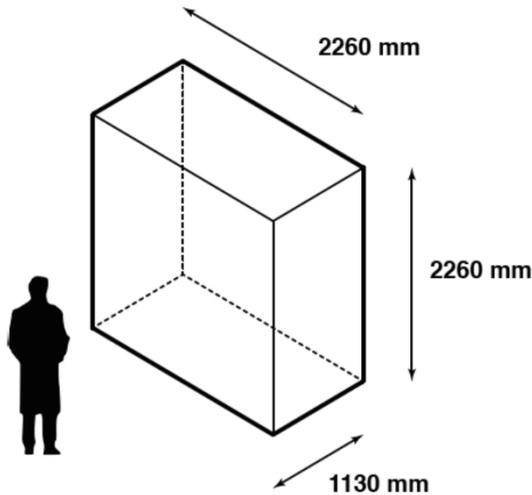

Fig. 1. Spatial properties of the display area and its relative size to an average size individual.

The interaction and scanning zone also needs to be able to facilitate the room required for individuals to interact with the installation as well as allowing for non-users to transit the area. Three scanning pillars are placed in around the area in order to form an equilateral triangle, measuring 4.75m between the posts. In the centre of this perceived triangle lies the actual interaction, as shown in Figure 2.
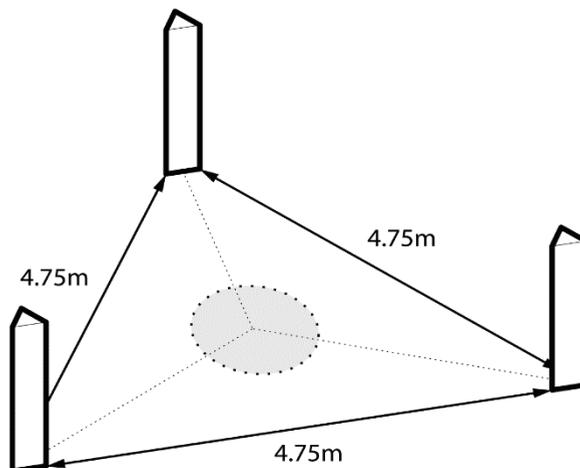

Fig. 2. Spatial properties of the scanning area and interaction zone.

2.2 Scanning area and technological principles

As mentioned before, a system capable of registering the 3D has to be put in place to develop a fully three-dimensional real-time model.

As such this system needs to be capable of not only registering this data but also compiling it into our desired output. In order to cover the 360 degrees around a model, a minimum of 3 scanning devices have to be used. Furthermore, each device needs to cover an area of minimum 130 degrees of the scanned model. This ensures an overlapping area between the three different surfaces, which helps in processing the date to a continuous 3D model. [5]

A software dealing with this intense computational process of patching together a 3D model from several sources must be populated with relevant information, such as the exact position of the scanning devices and the spatial relation between the three elements as well as the individual cloud point generated by each of them. Each individual cloud point brings a unique set of data to the system; however, as mentioned before, to create the fully three-dimensional model, the "overlapping" area between the different cloud points is critical. The reference position of each individual scanning device coupled with the "overlap" helps ensure a minimum of errors. The "overlap" establishes a bridging area between the cloud points; by comparing the different position between two points which are supposed to have an identical xyz coordinate, the software can compute the necessary adjustments. To identify which points from the different "clouds" correspond to each other, , the colour information of each voxel is compared to further ensure accuracy on top of the corresponding position approximation.

2.3 Display Area

The internal matrix structure will also be dimensioned according to the proportions coming from Le Corbusier's system. In order for the vision of the spectator to penetrate the matrix, so that he or she could easily understand the recreated shape, an appropriate spacing between the light pods needs to be take into consideration. As discussed above, the higher the density of the matrix is, the more accurate the physical representation of the form we obtain. However, as the elements become more and more compact, the problem of self-obstruction becomes a critical issue.

To solve this problem, a series of tests are necessary to establish the optimal spacing. Using a sequence or distances, relating to the Modulor, has led to a conclusion that a grid spacing of 48 mm could be the most optimal distance for our intended purposes. [4]

Because of this arrangement, the matrix will contain 48.668 LEDs, meaning that if all light sources were to be turned on, they would have a total power consumption of approx. 4.000 W/H and a total luminous intensity of 83.000 cd. However, such a scenario will never occur, most likely the maximum number of LEDs that will be active when a user is interacting with the installation will be around 4.500.

3 Prototype

The development of a working prototype has been a priority from the very first moment of the project idea, having in mind both proving that the technology required to achieve the desired outcome already exists and identifying any potential design flaws. It has also been a way of understanding the intricacies of the project, varying from the physical properties of the "medium" in the installation to the phenomenology of the anthropomorphic representation, which is unveiled using light.

3.1 Prototype development

Through intensive experimentation the prototype has allowed for establishing the principles that have further improved the final design in efficiency, through the development of possible construction technics, power supply methods, data transmission protocols and software engineering.

The prototype has been a key element in revealing the physicality of this work, both from an artistic perspective and as a logistic exercise. Despite a certain "mechanical" look, the installation is softened by the gentle light effects playfully moving within its domain.

The physical prototype was developed using the knowledge acquired at the initial digital experimentation stage, as well as the through the use of additional components, which could not be tested until this point. In order to achieve a fully working scale partial model of the sculpture, we needed to incorporate the additional components of the installation, such as a control system for the matrix, a scanning method and a power supply fitting the energy requirements of the prototype.

For the intended purpose of creating a proof-of-concept prototype, the control system was aligned to a standard PC, used in running the developed software, coupled with a FadeCandy micro-controller, which is able to output the computed data to the LED matrix. FadeCandy is the perfect tool for creating interactive light setups when using addressable LEDs. Being an open source hardware means that its Open Pixel Control protocol is compatible with many existing general high-level programming languages. The Fadecandy controller can drive up to 512LEDs, arranged in eight strips of up to 64 LEDs connected to a laptop through a USB port. [6]

To scan our environment, we will be using the Kinect standard. However, for the purpose of demonstration, the prototype only applies two of these devices, as opposed to the three used by the full-scale sculpture.

The matrix will be composed out of 512 PL9823 F5 RGB LEDs arranged in eight grids of 8x8 LEDs with a spacing of 48mm, which, when stacked on top of each other, result in a cubic lattice with a dimension of 35x35x35cm. This particular LED has a high data transmission stability and is able to receive and forward large information package output. It is able to produce luminous intensity up to 2000 mcd using 5v 20mA. This product has a very good performance and a long lifespan. [7]

The large number of LEDs require the appropriate power supply running at a low voltage. The lighting elements need supply that is able to provide the required current. As each individual LED needs 20mA per colour channel, a quick calculation reveals that in order to fully power the entire matrix at maximum brightness, we would need an electric source able to deliver 5V with a minimum of 40A.

3.2 Conceptual software prototype

One of the most challenging aspects in developing the prototype related to the development of the software driving the installation. It is an initial building block for a fully functional algorithm that can recreate a real time 3D representation. As discussed above, the prototype used only two Kinect, meaning that the computed data was incomplete. To transmit the data to the display area, the main idea behind the algorithm developed was to slice this representation to individual plates.

By using the OPC library within the Processing development environment, we were able to create an algorithm that uses the FadeCandy micro controller in order to address the appropriate Voxel (Pixel/LED) within the 3D Matrix of LED. For this purpose, , the model was sliced at eight different levels of depth; the outline was defined, and the shape was filled with a chosen colour.

The OPC library cannot work with voxels, only with pixels; therefore, the slices had to be converted into a two-dimensional plane to allow data collection, thus sending data to the variable "collector". The 512 collectors were spread out evenly on the surface coinciding with the screen space. [8]

3.2 Unifying Multiple Cloud Point Coordinates

The individual cloud points were generated using two Kinect sensors and the openKinect library for Processing. The library was developed as an open source tool for the exploratory applications of these sensors.

By using the RawDepth arrays generated by the library, we are able to translate the points to the Cartesian system using the following nested for loop.

```
for (int x = 0; x < Kinect.width; x ++)
for (int y = 0; y < Kinect.height; y ++)
int offset = x + y*640
int rawDepth = newDepth[offset]
Vector v = depthToWorld(x, y, rawDepth)
```

The depthToWorld functions performs the necessary computations for the points displayed in the appropriate 3D position. It deals with the internal calibration parameters of the sensors. [9]

Once the cloud points had been generated, attention had to paid to how the two relate to each other and what parameters are needed for their alignment and generation of the third cloud point containing the unified data.

In physics and engineering, Davenport chained rotations are three chained intrinsic rotations about body-fixed specific axes. Euler rotations and Tait–Bryan rotations are the cases of the Davenport general rotation decomposition. The angles of rotation are called Davenport angles because the general problem of decomposing a rotation in a sequence of three was studied first by Paul B. Davenport. [10]

This means that each vertex in the cloud point needs to be rotated in 3D space in order to create a unified coordinate system. In this convention Roll1 imposes the "heading", Pitch is the "inclination" (complementary of the elevation), and Roll2 imposes the "tilt". The matrixes that are needed for multiplication are as follows:

$$R_z(\phi) = \text{Roll}_1(\phi) = \begin{bmatrix} \cos\phi & -\sin\phi & 0 \\ \sin\phi & \cos\phi & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$R_y(\theta) = \text{Pitch}(\theta) = \begin{bmatrix} \cos\theta & 0 & -\sin\theta \\ 0 & 1 & 0 \\ \sin\theta & 0 & \cos\theta \end{bmatrix}$$

$$R_z(\psi) = \text{Roll}_2(\psi) = \begin{bmatrix} \cos\psi & -\sin\psi & 0 \\ \sin\psi & \cos\psi & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Following the calculation of the matrix multiplication, they had to be translated into the new coordinate system to create the unified cloud point. The origin of matrix multiplications is always a fixed point. Nevertheless, there is a common workaround using homogeneous coordinates to represent a translation of a vector space with matrix multiplication: Write the 3-dimensional vector w = (*wx, wy, wz*) using 4 homogeneous coordinates as w = (*wx, wy, wz*,1). [11]

To translate the individual cloud points, each homogeneous vector *p* contained within the matrix can be multiplied by the vector *v* (describing the position of each Kinect sensor) using the following translation matrix:

$$T_{\mathbf{v}} = \begin{bmatrix} 1 & 0 & 0 & v_x \\ 0 & 1 & 0 & v_y \\ 0 & 0 & 1 & v_z \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Thus, the multiplication will show the expected result as described below:

$$T_{\mathbf{v}}\mathbf{p} = \begin{bmatrix} 1 & 0 & 0 & v_x \\ 0 & 1 & 0 & v_y \\ 0 & 0 & 1 & v_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} p_x \\ p_y \\ p_z \\ 1 \end{bmatrix} = \begin{bmatrix} p_x + v_x \\ p_y + v_y \\ p_z + v_z \\ 1 \end{bmatrix} = \mathbf{p} + \mathbf{v}$$

Through the process described above, we can generate a continuous cloud point, which can be displayed with the domain of the physical installation. hence, creating a system that can generate real-time content as well as displaying it in three dimensions.

3 Conclusion

Following the development of the prototype, several conclusions can be made regarding the areas investigated. While developing the prototype, it has proven to be very challenging to assembly the display area. The process required to develop the individual plates of LEDs has proven to be extremely time consuming. Should this not be the case of the prototype, this aspect could have been avoided by involving manufacturers that can produce the LED sheets according to the design specifications.

The most challenging part of the experiment, as in a method of rationalising 3D space, was related to the development of the software prototype. Even though it was simplified in order to better understand the intricacies of generating a 360 degrees 3D model, the software development is a crucial part for the further advancement of the system. It appears that previous efforts in this field, such as the development of the NOVA system, were confronted with the same problems. However, previous systems dealt with pre-generated and not real-time content. [12]

Previous research in the field of real-time 3D data reconstruction has been done by different research groups. The most promising research results are presented in the following paper: "Real-Time, Full 3-D Reconstruction of Moving Foreground Objects From Multiple Consumer Depth Cameras" [13]. However, this study does not address the output aspect of data or any potential artistic implications of the technology.

The research in the representation of form and movement through light is far from being exhaustive; however, the initial experimentation has proven to be extremely promising. Further research and development of the software is needed to make this concept a reality.

References

1. MVRD and COBE, Rockmagneten Masterplan, (2016).
2. Zhang W., Microsoft Kinect Sensor and Its Effect. IEEE Multimedia. 19, 4–10 (2012).
3. Foley, James D.; Andries van Dam; John F. Hughes; Steven K. Feiner "Spatial-partitioning representations; Surface detail". Computer Graphics: Principles and Practice, (1990).
4. Cohen, J.-L., Le Corbusier's Modulor and the Debate on Proportion in France. *Architectural Histories (2014* .

5. Kwan Pang Tsui; Kin Hong Wong; Changling Wang; Ho Chuen Kam; Hing Tuen Yau, et al. " Calibration of multiple Kinect depth sensors for full surface model reconstruction ", *Proc. SPIE* 10011, First International Workshop on Pattern Recognition, 100111H (July 11, 2016).

6. Micha Elizabeth Scott, FadeCandy Data Sheet, (2016).

7. PL9823 F5 RGB, Data Sheet.

8. OLE (Object Linking and Embedding) for Process Control.

9. Daniel Shifman, "Getting Started with Kinect and Processing".

10. P. B. Davenport. "Rotations about nonorthogonal axes." AIAA Journal, Vol. 11, No. 6 (1973).

11. Richard Paul. "Robot manipulators: mathematics, programming, and control: the computer control of robot manipulators", MIT Press Cambridge (1981).

12. Schubiger-Banz S., Eberle M. The NOVA Display System. In: Adams R., Gibson S., Arisona S.M. (eds) Transdisciplinary Digital Art. Sound, Vision and the New Screen. Communications in Computer and Information Science, vol 7. Springer, Berlin, Heidelberg (2008) .

13. Real-Time, Full 3-D Reconstruction of Moving Foreground Objects From Multiple Consumer Depth Cameras, IEEE, Transactions on Multimedia Volume: 15, (2013).