EXTENDING THE LIFE OF VIRTUAL HERITAGE: REUSE OF TLS POINT CLOUDS IN SYNTHETIC STEREOSCOPIC SPHERICAL IMAGES

J. Garcia Fernandez [a,*], K. Tammi [b], A. Joutsiniemi [a]

[a] TUT, School of Architecture, Tampere University of Technology, Rakennustalo Building, 33720 Tampere, Finland - (jorge.garciafernandez, anssi.joutsiniemi)@tut.fi
[b] TAMK, School of Construction and Environmental Engineering, Tampere University of Applied Sciences, Kuntokatu 3, 33520 Tampere, Finland - kalle.tammi@tamk.fi

Commission II

KEY WORDS: Point Cloud, Terrestrial Laser Scanner, Reuse, VR, Stereoscopic, Spherical Images

ABSTRACT:

Recent advances in Terrestrial Laser Scanner (TLS), in terms of cost and flexibility, have consolidated this technology as an essential tool for the documentation and digitalization of Cultural Heritage. However, once the TLS data is used, it basically remains stored and left to waste.How can highly accurate and dense point clouds (of the built heritage) be processed for its reuse, especially to engage a broader audience? This paper aims to answer this question by a channel that minimizes the need for expert knowledge, while enhancing the interactivity with the as-built digital data: *Virtual Heritage Dissemination through the production of VR content.*

Driven by the ProDigiOUs project's guidelines on data dissemination (EU funded), this paper advances in a production path to transform the point cloud into virtual stereoscopic spherical images, taking into account the different visual features that produce depth perception, and especially those prompting visual fatigue while experiencing the VR content. Finally, we present the results of the Hiedanranta's scans transformed into stereoscopic spherical animations.

1. INTRODUCTION

During the last decade, laser scanner (LIDAR) technology has been consolidated as an essential tool for the documentation and digitalization of Cultural Heritage sites (Nothegger & Dorninger, 2009). Simultaneously, the cost of the required equipment has decreased significantly and more versatile software solutions, with advanced processing capabilities, have become available. The aforementioned factors have ensured the growth of point clouds' popularity (LIDAR native´s data), as a key data structure in digital Heritage management.

Point clouds are solving a wide range of needs of Cultural Heritage, especially in the areas of conservation and communication, and loads of digital treasures are therefore being captured and stored. However very few reuse alternatives have been implemented for generally unstructured point cloud data, while requiring high memory capacities for both management and storage purposes.

At the same time, from a more social perspective, the effective heritage communication has become a crucial part of the site's conservation, as well as an element for the education and cohesion of the society (Garcia-Fernandez, 2014). Particularly, immersive VR content opens a new door not only for widening the access to Heritage-related information, but for a better –interactive– understanding of complex historical sites (Jimenez et al., 2016).

How can highly accurate and dense point clouds (of the built heritage) be processed for its reuse, especially to engage a broader audience? This paper aims to answer this question by exploring a channel that minimizes the needs for expert knowledge, while enhancing the interaction with the as-built digital data: Virtual Heritage Dissemination through the production of VR content.

2. TERRESTRIAL LASER SCANNER

During the last few years the Laser Scanner Technology has gained a privileged position in the accurate documentation of the built environment. The substantial decrease in the hardware's price, the flexibility in terms of dimensions, weight, and scan speed, together with the increased processing capabilities of desktop computers, has turned terrestrial laser scanners (TLS) into a recurrent tool for the built environment's documentation (Garcia-Fernandez, 2016).

The use of this technology, and its typical data: the point cloud, has already transcended the boundaries of remote-sensing disciplines. Still, their eminently technical character restricts the use of TLS data to an expert use, especially for the representation, interpretation and expert analysis of cultural heritage. Multiple documentation projects based on TLS are performed every year worldwide. Large and rich volumes of data (point clouds of cultural heritage assets) are stored, at best, once they have met the specific demands. In a world where the permanence of digital data is so uncertain (Carraway, 2011), we are facing the risk to lose the chances of exploiting the added value of these documents, especially the value for a non-expert public, eager and in need to take advantage of its cultural heritage (Garcia-Fernandez, 2014).

2.1 Reuse of Point Cloud

The concept of point-cloud reuse has been properly addressed in the case of aerial LIDAR (Hsu et al., 2011), where national and international institutions have implemented online-available and free LiDAR Data Sources. However, for terrestrial laser scanner

of the keystone effect (objects are seen with a different obliquity from each camera). Even though from this disparity a bias in-depth perception is expected, there are no conclusive proofs that vertical disparity can induce visual discomfort (Speranza et al., 2002; and Allison, 2007). In our case (virtual toed-in configuration) the vertical disparity is not noticeable, since this effect is acuter in the corner of the photograms, and we propose to conduct the renderization in very narrow strips.

In order to define the zero parallax plane, it is necessary to accurately set the convergence angle (α) of the two cameras. For this, we suggest to follow the equation stated by Bourke (2010) (1).

$$\alpha = 2 \operatorname{asin} \frac{r}{D_z} \quad (1)$$

where α =the convergence angle
r =radius of the rotation circle
Dz = distance to the zero parallax.

3.3 Rendering

Having fully defined the camera configuration, the rendering process in our production path is conducted in Autodesk 3D Max, with the rendering application NVIDIA Mental Ray. Using the 'WrapAround' lens, Mental Ray allows us to get the entire 360-degree environment surrounding the camera, spherically projected onto a dome.

Blur effects in rendering

Aiming to minimize the visual discomfort and to better simulate the human depth perception, we have introduced the depth of field (DoF) blur effects, as suggested by Carnegie and Rhee (2015). The retinal image blur is one of the main monoscopic cues that helps our system to understand the object's size and the perception of depth. Besides, this effect softens the weights of disparities in the extreme DoF range, where maximum values and the visual discomfort are reached.

In the case of synthetic images (computerized process), the blur effect can be introduced by customizing the camera parameters (similarly to what happens in the 'physical' photography) or directly throughout the render motor (Figure 5).

Since we don't want to give the appearance of capturing a smaller world (easily recognized when capturing images with micro-photography), we suggest to use the human in-focus range as a reference, especially to represent the as-built heritage from terrestrial perspective. In this case, for a human being averagely having a DoF of 0.3D, at a focus distance of 3m, means that the in-focus range is from 1.6m-30m (Devernay and Beardsley, 2010).

Figure. 5 Testing different amplitude of Depth of Field in Mental Ray through 'Camera Effects Rollout'

For spherical images, blur introduction may have an unwanted effect in the near range, where cameras are set close to the floor (simulating the human being position). Panning the visualization closer to the nadir will bring a fully blurred screen representing the ground plane. This will not allow the eyes to find the optimal area to converge. One solution is to only set the blur effect to the far range, the other one is to limit the allowance of the camera tilt.

Textural Gradient in rendering

One of the main monoscopic depth cues contributing to the depth perception is the textural gradient. The amount of variation 'from point to point in the texture is directly related to the distance from the viewer' (Rosas et al., 2004). This cue will improve not only the depth perception but also the surface orientation in the case of surfaces slanted in depth.

Specifically, for the case of point clouds, we take advantage of the discretion of the points, in order to exploit the Textural gradient depth cue. The increases in density of the projected texture, is in our case related with the point cloud geometry itself: (i) proximity/distance from point to point in the cloud; and (ii) dimension of the point to be represented.

These parameters must be defined in close relation with the distance to the object and the resolution of the input data. In our test, with an average resolution of 1 point every 10mm at 10 meters, we decided neither to create new points, nor to increase the size point in the short range.

Automatic image seaming

Within the digital environment managed in 3D Max, we accurately know the camera's position and its orientation in the scene, in every rotation step. Besides, the virtual cameras do not contribute with errors due to optical aberration or lens misalignment. These make unnecessary the overlapping and further blending function (Huang and Hung, 1998). The rendered slits can be directly stitched.

In order to automate the stitching process, we have implemented the task in Image-J, a former neuro-science software for expert image analysis. Within Image-J, the plugin *Stitch Sequence of Grid of Images*, based on the stitching algorithms created by

Stephan Preibisch (Preibisch et al., 2009), was preferred to conduct the automatic image seaming.

The selection of this plugin was done based on its capacity to manage the stitching in multiple loops, and organize them by grid along the three axis. These become critical when aligning multiple render loops, as in the case described in the following section.

3.4 Animations in a video sequence

One of the main potentials of conducting this process in a virtual environment, is that the as-built heritage scene can be enriched with 'almost?' any other digital information. Enhancing the digital environments and the stereoscopic spherical documents with storytelling-based strategies, can bring back inhabitants to the historical site, as well as populate it with monsters and princesses, recreating legends, myths, or old ways of life. This turns simple (cultural heritage) diffusion of data into a rich interpretative and educational product (Garcia-Fernandez & Medeiros, 2014).

Empowering the scene with animations developed in time and space, first requires the implementation of the added information (in our study case 3D models); and second, the re-configuration of the cameras in terms of the loops (frames) and path.

In order to optimally visualize an animation with moving objects, we need to develop the video sequence with a ratio higher than 15 (spherical) frame per seconds. This multiple rendering is a significant time-consuming task, requiring also a high processing capacity. The final quantity of slips (S) can be easily estimated through (2).

$$S = 2nqt \quad (2)$$

Where n = the number of steps in each loop (frame)
q = number of frames per second
t= amount of second for the whole video sequence

In our test, in order to keep the time manageable while using average personal computers, we decided to set one capturing position, and to conduct the process in two steps. First we rendered the 3D environment without the animation, and second we rendered the animation without the environment, setting the background in pure green to easily extract it in post processing (Figure 6).

We implemented this procedure combining a spherical image of the mansion's attic with a 3D monster walking along the scene (free monster model downloaded from http://3drt.com/). During the tests, we have set 360 steps (α=1°), and 25 photograms per second along 20 second in left and right cameras. That means a total 360.000 slits rendered in batch during 12h. For this task we have used an eight-CPUs Intel Core i7 personal computer at 2.8GHz; 32 Gb of RAM; NVIDIA Quadro K21000M graphic card, and running on Windows 7.

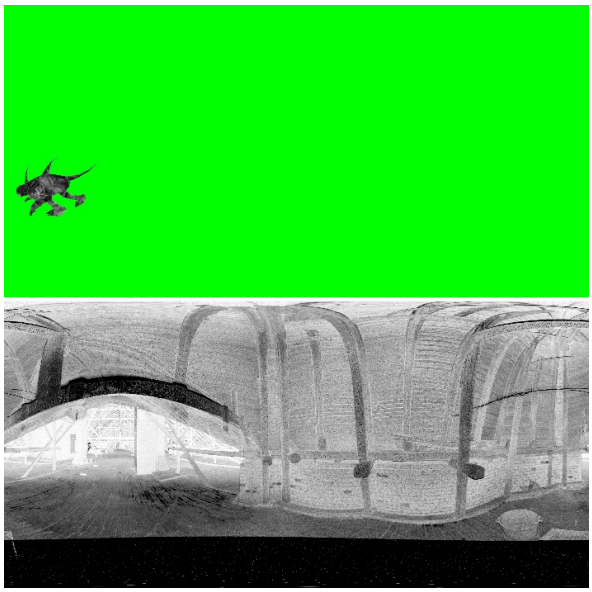

Figure 6. Top, single photogram of the animation (after stitching); Bottom, left spherical image representing the environment

Specifically, for managing this large number of frames, we have developed a script that allows to rename and organize the spherical frames, specifically for complying the input data demands in the stitching processing using *Stitch Sequence of Grid of Image* (Figure 7).

```
do while XFICH<=107999
if XPRE<=9; XPREFIJO="S00"+str(XPRE,1)
   else
    if XPRE>=9.and. XPRE <=99; XPREFIJO="S0"+str(XPRE,2)
    else; XPREFIJO="S"+str(XPRE,3)
    endif
endif
 do while CICLO<=360
if CICLO<=9; CONS="_00"+str(CICLO,1)
   else
    if CICLO>=9.and. CICLO <=99; CONS="_0"+str(CICLO,2)
    else; CONS="_"+str(CICLO,3)
    endif
endif
IF XFICH<=9;XENTRADA="R000"+STR(XFICH,1)+".JPG"; XSALIDA=XPREFIJO
+CONS+".JPG"; !REN &XENTRADA &XSALIDA
 ELSE
 IF XFICH>=10 .AND. XFICH<=99; XENTRADA="R00"+STR(XFICH,2)+".JPG";
XSALIDA=XPREFIJO+CONS+".JPG"; !REN &XENTRADA &XSALIDA
  ELSE
  IF XFICH>=100 .AND.XFICH<=999; XENTRADA="R0"+STR(XFICH,3)+".JPG";
XSALIDA=XPREFIJO+CONS+".JPG"; !REN &XENTRADA &XSALIDA
   ELSE
    IF XFICH>=1000.AND. XFICH<=9999; XENTRADA="R"+STR
(XFICH,4)+".JPG"; XSALIDA=XPREFIJO+CONS+".JPG";!REN &XENTRADA
&XSALIDA
     ELSE; IF XFICH>=10000.AND. XFICH<=99999; XENTRADA="R"+STR
(XFICH,5)+".JPG"; XSALIDA=XPREFIJO+CONS+".JPG"; !REN &XENTRADA
&XSALIDA
      ELSE; XENTRADA="R"+STR(XFICH,6)+".JPG"; XSALIDA=XPREFIJO
+CONS+".JPG"; !REN &XENTRADA &XSALIDA
ENDIF
 ENDIF
  ENDIF
    ENDIF
     ENDIF
CICLO=CICLO+1; XFICH=XFICH+1
 ENDDO
```

Figure. 7. Code for adapting multiple frames to the plugin *Stitch Sequence of Grid of Image* (J-Image software)

The final video creation includes the removal of green backgrounds, the overlapping of both environment and animation sequences, and the transformation of the multiple single frames into continuous video format. This task can be conducted in almost any video editing software; for the purpose of this research, we used Sony Vega Pro.

3.5 Visualization

The race for allowing VR content in large-audience platforms is already on its way. YouTube, Facebook, Hulu, Google Play, HBO, NETFLIX, Jaunt, among others have already implemented solutions for it.

Considering the free and easy access, and the large user community, we have decided to implement our study cases in YouTube. This platform currently allows top-bottom and side-by-side VR representation, suitable for anaglyph in desktop display and stereo images for head-mounted devices (HMD).

Metadata injection for the recognition of the stereo spherical images on YouTube can be solved with the crowdsourced developed app: Spatial Media (https://github.com/google/spatial-media). A user-friendly interface version is accessible through YouTube's official page. The two screenshots in Figure 8 shows the Lielahti's Attic animation, uploaded on YouTube.

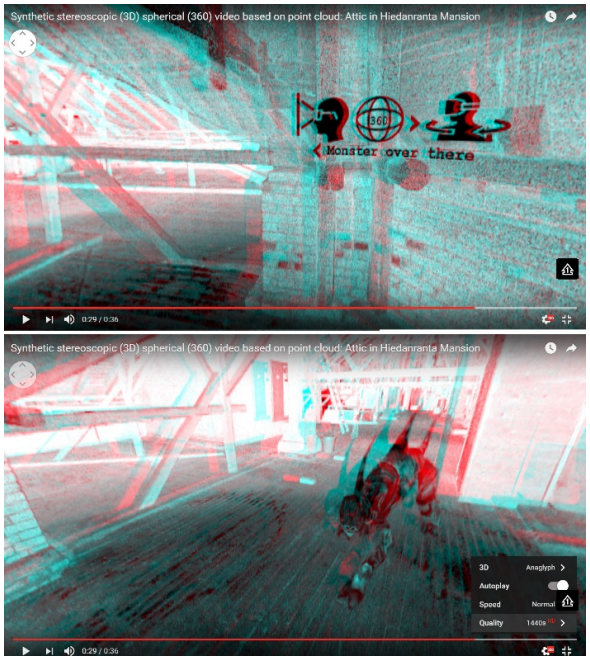

Figure 8. Animation based on sythethis stereoscopical spherical images, Lielahti 's attic, Tampere Finland

4. CONCLUSIONS AND FUTURE RESEARCH

This paper has first drafted the current state of terrestrial laser scanners in Cultural Heritage and the imperative need of giving a second life to the point cloud, especially for heritage communication purposes. Second, we present a semi-automatic production path for the creation of synthetic stereoscopically spherical images, addressing the main elements causing visual fatigue in its production.

This study attempts to reach an optimal -legible- solution for the semi-automatic stereoscopic VR content creation. The production path was demonstrated and tested by the study case of Lielahti mansion, as part of the EU funded regional ProDigiOUs project.

Given the complexity of the topic, some elements got out of the scope, particular an in-depth analysis about Accommodation-Convergence conflict in the creation of spherical stereoscopic images. Further research must analyze the aforementioned conflict, its relation to nearness of the visualization screen (especially in HDM), and the specific ranges of horizontal disparities from crossed to uncrossed disparities in toed-in camera configuration.

4.1 References

Allison, R., 2007. Analysis of the influence of vertical disparities arising in Toed-in Stereoscopic cameras, *Journal of Imaging and Technology*, 51:4, pp. 317-327

Atallah, P., Pelah, A., and Wilkins, A., 2012. Visual stress symptoms from stereoscopic television, IEEE/IET Electronic Library (IEL) Conference Proceedings, pp. 1-7

Bourke, P., 2006. *Synthetic Stereoscopic Panoramic Images*, in Lecture Notes in Computer Science (LNCS), Springer, V. 4270, pp. 147-155

Bourke, P., 2010. Capturing omni-directional stereoscopic spherical projections with a single camera, *16th International Conference on Virtual Systems and Multimedia, VSMM 2010*, pp. 179-183

Carnegie, K., and Rhee, T., 2015. Reducing Visual Discomfort with HMDs Using Dynamic Depth of Field, *IEEE Computer Graphics and Applications*, pp. 34-41

Carraway, L., 2011. On preserving knowledge, *The American midland naturalist*, 166:1, pp. 1-12

Chen, J., Hsu, P., and Lin, C., 2012. Model retrieval based on point cloud encoding of airborne LIDAR, *32nd IEEE International Geoscience and Remote Sensing Symposium*, IGARSS 2012; Munich; Germany; 22 July 2012 through 27 July 2012;

Devernay, F. and Beardsley, P., 2010. Stereoscopic Cinema, in *Image and Geometry Processing for 3-D Cinematography*, volume 5 of Geometry and Computing, Springer Berlin Heidelberg

Garcia-Fernandez, J., 2014. *Digital Culture for the enhancement of Heritage value, generation of heritage product with educational scope*, doctoral thesis, Published by the Editorial Service- University of Valladolid, Spain.

Garcia-Fernandez, J., 2016. An Assessment of Errors and Their Reduction in Terrestrial Laser Scanner Measurements in Marmorean Surfaces, *3D Research Journal*, March 2016, 7:2, pp. 1-11

Garcia-Fernandez, J., and Medeiros, L., 2014. *Chapter 3: Digital Storytelling*, in MOOC Course, New Strategies for the Safeguard of Cultural Heritage, Coord. Fernandez, J., Ed. University of Valladolid.

Huang, H., and Hung, Y., 1998. Panoramic Stereo Imaging System with Automatic Disparity Warping and Seaming, *Graphical models and image processing*, 60:3, pp. 196–208

Hsu, P., Chen, J., and Lin, C., 2011. 3D building model retrieval for point cloud modeling, *32nd Asian Conference on Remote Sensing 2011, ACRS*, V. 4, pp.2243-2248

Jimenez, B., Morabito, D., and Remondino, F., 2016. Access to complex reality-based 3D models using virtual reality solutions, *Journal of cultural heritage*, Ed. Elsevier, pp. 1-9

Nothegger, C., and Dorninger, P., 2009. 3D filtering of high-resolution terrestrial laser scanner point clouds for cultural heritage documentation, *Photogrammetrie, Fernerkundung, Geoinformation*, 2009:1, pp. 53-63

Pacheco, D., 2014. Spatializing experience: a framework for the geolocalization, visualization and exploration of historical data using VR/AR technologies, *Proceedings of the 2014 Virtual Reality International Conference 20*14, pp. 1 -4

Preibisch, S, Saalfeld, S., and Tomancak, P., 2009. Globally optimal stitching of tiled 3D microscopic image acquisitions, *Bioinformatics*, 25:11, pp. 1463 -1465

Quattrini, R et al., 2016. Virtual reconstruction of lost architectures: from the TLS survey to AR visualization, *International archives of photogrammetry and remote sensing*, v:41, pp. 383-390

Rosas, P., Wichmann, F., Wagemans, J., 2004. Some observations on the effects of slant and texture type on slant-from-texture, *Vision Research*, 44:13, p. 1511 -1535

Speranza F, Stelmach, L., Tam, W., and Glabb, R., 2002. Visual comfort and apparent depth in 3D systems: effects of camera convergence distance. In: Proc. SPIE Three-Dimensional TV, Video and Display, SPIE, vol. 4864, pp 146–156,

Yamanoue, H., 2006. The differences between toed-in camera configurations and parallel camera configurations in shooting stereoscopic images, ICME 2006,

Wei, M., 2003. Do visual cues contribute to the neural estimate of viewing distance used by the oculomotor system?, *The Journal of neuroscience*, 23:23, pp. 8340 -8350

Wann, J., and Mon-Williams, M., 1997. Health issues with virtual reality displays, Computer Graphics, pp.53-57

4.2 Acknowledgements

This work has been conducted in conjunction with the ‘ProDigiOUs’ regional EU project and supported by the School of Architecture, Tampere University of Technology (Finland); and the City of Tampere, Finland.