3D RECONSTRUCTION OF ARCHAEOLOGICAL WALLS USING KINECT

M. Quintana a, F. Zvietcovich a, B. Castaneda a*,

a Dept. of Electrical Engineering, Pontificia Universidad Católica del Perú, Lima 32, Perú – matias.quintana@pucp.pe, (fzvietcovich, castaneda.b)@pucp.edu.pe

KEY WORDS: Kinect, point cloud, 3D registration, 3D reconstruction, photogrammetry, ICP, SURF.

ABSTRACT:

3D acquisition and reconstruction is extensively used in documenting and preserving archaeological sites. Despite its benefits, price and portability of the devices used are still an issue in many cases. This work presents a characterization of the Microsoft Kinect sensor and a complete methodology of data acquisition, registration and 3D reconstruction of archaeological walls based on the application of image processing and computer vision algorithms. Consecutives range and color images are obtained by a Kinect for XBOX 360. Then, the range images are converted into point clouds and separated in groups where, based on their respective color images, are aligned using Speeded Up Robust Features (SURF) algorithm and refined with Iterative Closest Point (ICP) approach. Finally, an accumulative registration using ICP approach is applied to all point clouds. Subsequently, a 3D mesh is formed and compared with other 3D reconstructions such as one based on photogrammetry, and a commercial 3D laser scanner. Experiments applied to the Kinect provide errors of 6.3 millimeters in accuracy and 2.2 millimeters in precision. When 3D mesh models obtained using the proposed methodology are compared to a photogrammetry-based reconstruction, the accuracy error is 4.79 centimeters, and 1.49 centimeters in relation to a laser scanner model. This methodology provides relatively high accuracy metrics and proposes an inexpensive alternative for 3D scanning, surpassing other devices in price and time needed for data acquisition. This project deserves a more formal prospective study in using the high density point clouds to increase the resolution of the 3D reconstruction.

1. INTRODUCTION

3D scanning technology has evolved considerably in the last few decades. Among the available technology for the different applications, a variety of products can be found with ranges of price, accuracy, precision and resolution. Many techniques such as photogrammetry and time of flight scanners are used in archaeological sites [1,2,3] but their cost is not usually meet by the excavation budget, especially in developing countries with abundance of archaeological structures such as Peru. The importance of a low-priced system that meets a project's requirements is indispensable. Many available sensors in the market achieve accessible costs but portability and need of a technical user are still drawbacks. Nevertheless, the appearance of the Kinect sensor has managed to equilibrate the requirements. Despite the fact of its entertainment purpose, the technical characteristics of the Kinect (Microsoft, Washington USA) allow it to be used in many research areas [4,5]. The aim of this work is to develop a methodology of data acquisition, registration and 3D reconstruction based on a low-priced sensor, the Kinect, and image processing and computer vision algorithms. In order to compare the reconstruction results, the final 3D mesh will be compared with photogrammetric (Agisoft v1.0.4, 2014) and 3D laser scanner (NextEngine HD, NextEngine Inc., CA, USA) reconstructions to obtain an accuracy error.

This paper is organized as follows: - In section 2, we describe the Kinect characterization experiments. In section 3, we explain the entire methodology, the acquisition step and the algorithms utilized for registering a sequence of point clouds obtained with the Kinect for XBOX 360. In section 4, we compare the resultant 3D models with other reconstructions based on techniques that are normally used in archaeology. In section 5, the results of the reconstructions and of the devices used are explained and compared. The conclusions of the entire work are explained in this section.

2. KINECT CHARACTERIZATION

2.1 Elaboration of Testing Model

In order to determine the Kinect errors, a light wood testing model was built with dimensions of 33 centimeters wide and 45 centimeters long. The testing model consists on a plane platform with various geometric features of different sizes. As showed in Figure 1, the testing model has a solid paraboloid, four hemispheres and a staircase.

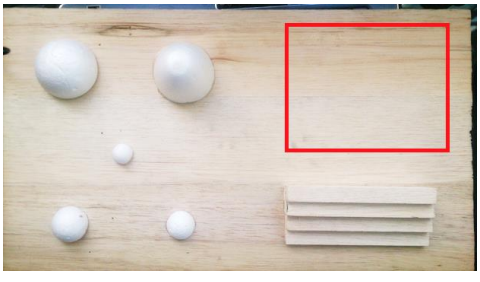


Figure 1: Testing model built for error measurement.

* Corresponding author. Tel.: +5116262000 ext. 4720. Benjamin Castaneda is at Laboratory of Medical Images Research and principal professor at Electrical and Electronic Department at the Pontificia Universidad Católica del Perú, Lima 32, Perú (e-mail: castaneda.b@pucp.edu.pe).

2.2 Precision Error

An amount of 300 consecutives samples were taken at different distances away from the Kinect with MatLab (MATLAB 8.1 and Microsoft Kinect for Windows Support from Image Acquisition Toolbox Release 2013a, The MathWorks, Inc., Natick, Massachusetts, United States). The first distance was 0.85 meters until 1.35 meters, with intervals of 0.5 meters. In addition, there were considered 3 light conditions, where Light 0 consisted in artificial lights turned off and closed black curtains on the windows, Light 1 had artificial lights turned on and closed black curtains on the windows and finally Light 2 had the artificial lights turned on and the curtains open. The area inside the red square in Figure 1 was the analyzed part because is a plane section and can be compared with a perfect plane equation.

The samples coming from the Kinect consist in color and range depth images. After acquired all the samples, the ones sharing the same distance were concatenated and the range depth images were transformed into point clouds. Later, a plane adjust was made to the point clouds and they were rotated to a plane Z = 0. In this way, distances could be measured and compared with the plane equation at which the data was taken.

Results of the standard deviation of the distance errors of the samples are showed in Figure 2. The deviation (precision error) increases as we move away from the device. We can conclude that the precision error of the Kinect for XBOX 360 ranges from 1.2 to 4.0 millimeters, in a range of 0.85 to 1.36 meters with indoor light conditions. These results can be compared with the ones obtained in [4,5].

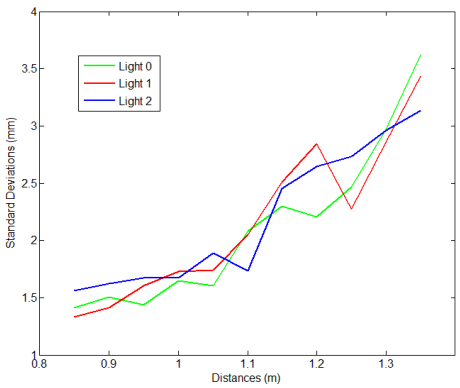


Figure 2: Precision error (standard deviation) versus distances at which the samples were taken.

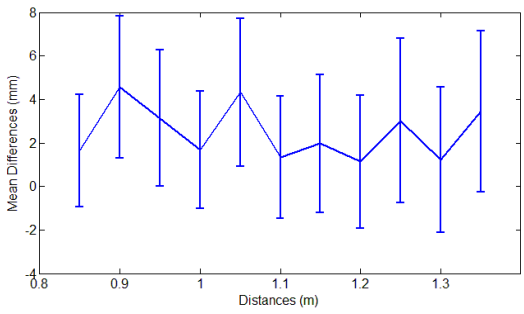


Figure 3: Accuracy error between the Kinect (data being evaluated) and NextEngine Laser Scanner HD (ground truth) for each distance.

Table 1 Precision and accuracy error comparison between the Kinect, NextEngine Laser Scanner HD [6] and photogrammetry [1].

Error	Kinect for XBOX 360	NextEngine Laser Scanner HD [6]	Photogrammetry [1]
Precision (mm)	2.2	0.107	16
Accuracy (mm)	6.3	0.127	4

2.3 Accuracy Error

In this case, the difference between obtained measures versus real measures was evaluated. The testing model was used as a whole and the samples were compared with a point cloud obtained with a 3D laser scanner NextEngine HD (NextEngine Inc, CA, USA) considered as ground truth. The light reference, Light 1, was chosen for this experience.

For each distance, the Kinect point cloud was compared with the 3D laser scanner point cloud. Figure 3 shows the mean differences between the point clouds as well as the maximum standard deviation. The error tends to be constant even when the distance from the device increases. From these results, the mean of the accuracy error is 2.5 millimeters within the range and light conditions already mentioned.

2.4 Results Analysis

Based on the parameters found in the work done in [6] and [1], Table 1 is constructed. Among the experiments in [6], the NextEngine HD is characterized, and in [1] the parameters obtained for photogrammetry are considered as acceptable for archaeological applications.

Based on these results, we can conclude that the Kinect for XBOX 360 is an appropriate device for archaeological applications of large dimensions in a distance range of 0.85 to 1.35 meters with very low solar light and standard artificial light.

3. METHODOLOGY

3.1 Data Acquisition

A cloudy day was chose for obtaining the data in order to have better quality range depth data. Some natural light was still needed because of the color images; otherwise they wouldn't be useful for the registration stage. The Inca Path, located within Pontificia Universidad Católica del Perú, Lima, Perú, was selected as target for the 3D reconstruction using the proposed methodology. It is formed by two well defined parallel walls made of mud mortar. This prehispanic path is associated to the Maranga Archaeological Complex that is formed by, at least, eight edification (*Huacas)* occupied by Ichma and Lima civilizations during 100-1470 A.D. period.

Figure 4 shows the positioning of the Kinect 1 meter away from an Inca Path, and the direction in which the Kinect sensor was moved during the acquisition process. Samples were captured in two zones of the Inca Path. The acquisition mode was set manually: one sample at a time, giving us time to move the

Kinect, 35 samples per zone; and automatically: 200 consecutives samples as we move rapidly the Kinect, only for the first zone. As a final step, the range depth images were transformed into point clouds, and the relation between each point to the pixel in the color image was preserved.

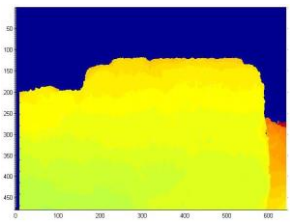
3.2 Keypoints localization and matching

As a first step, we used Speeded Up Robust Features algorithm (SURF) [7,8] to find the keypoints and their matching relation between samples. We select this algorithm among other such as Scale-Invariant Feature Transform (SIFT), because it gives us better descriptors in a lower computational time. If we choose the first sample as reference, the second one will have a good amount of matched keypoints but as we move forward, the tenth will only have a few (Figure 5).

(a) Kinect 1 meter away from the Inca Path

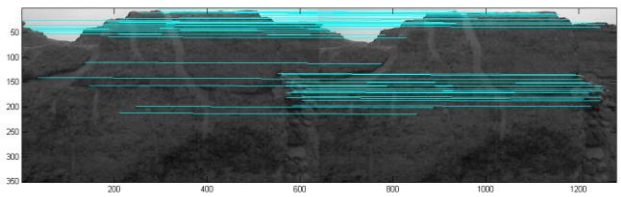
(b) The red arrow shows the direction in which the device was moved.

(c) Color image captured by the Kinect.

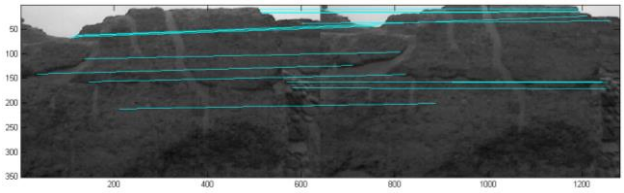


(d) Range depth image captured by the Kinect.

Figure 4: Kinect positioning next to the archaeological site. As the acquisition starts, the device will be manually moved (b) and the captured data will be a set of color images and a range depth images.



(a) Matching keypoints between consecutives samples



(b) Matching keypoints between first sample and tenth sample

Figure 5: Matching keypoints linked with a light blue bar. In (a) there are 63 keypoints and in (b) just 17.

In order to solve this very low quantity of matching keypoints, the samples were separated in small groups. In this way, the first sample of each group will become a reference of the rest of their respective group. The chosen group size was 4 and the number of matching keypoints between the first one and the fourth one are 55. Figure 6 shows how each sample in a group is analyzed with SURF to the first sample of the respective group.

3.3 First alignment determination

The second step consist in using the pixel to pixel relation between color images obtained previously for the calculation of the rotation matrix and translation vector to align two samples [9]. This is possible because we conserve the relation between each pixel in a color image with the 3D point obtained of the corresponding range depth image. Due to the fact that samples were separated in groups, samples within a group will be alignment but no necessarily with other samples from other groups. In Figure 7 (a) we see two consecutives samples of the first groups and in (b) the same samples aligned based on the matching of keypoints.

3.4 Alignment refinement

Until now, all the point clouds p belonging to a group are aligned within the group. In order to have a better alignment (increase the splotchy region in Figure 7 (b)), ICP algorithm is applied to the samples of each group in a combinatory fashion. The method is described in Figure 8.

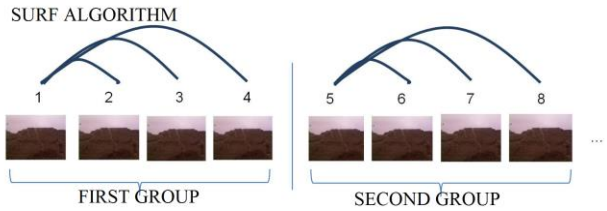
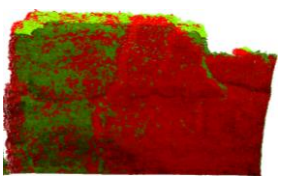


Figure 6: SURF algorithm applied to method group of consecutive color images. All the samples within a group are compared with the first sample of the group.

(a) The green point cloud is the reference and the red one will be rotated and moved.



(b) Aligned point clouds

Figure 7: The point clouds (a) will be aligned using R y T obtained after applying SURF. The result is shown in (b).

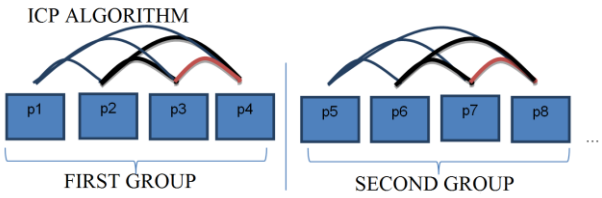


Figure 8: ICP algorithm is applied in a combinatory fashion within each group. For a group of 4 samples, the ICP algorithm is used 6 times.

3.5 Establishing alignment between groups

The next step will be to align each group with the first one. Nevertheless, because the point cloud has already being modified with SURF and ICP, the direct relation between the color image and the point cloud doesn't exist anymore. Then, it is not possible to apply SURF within those. On the other hand, if we apply directly ICP it will only bring the two point clouds as close as possible but will not take into consideration their common regions.

The link between two groups lies in the alignment of the last point cloud of each group (*n*) and the first point cloud of its consecutive group (*n+1*). If we use SURF and set n as the reference, it will align *n+1* with the original position of *n*. So far, n had a new position based on the *R* and *T* obtained by SURF and ICP within its group as described in Section 3.3 and 3.4. Therefore, in order to align *n+1* with the current position of *n*, it requires these to be transformed using *R* and *T*. Furthermore, when any point cloud is about to be align with the previous point cloud, it can be done by using SURF in addition to the values of *R* and *T* that the previous point cloud has used. Finally, its new position is calculated taking in account the actual and previous transformation of the clouds. We see that this is an accumulative process that requires all of the R and T in order to calculate the new value of the point clouds *p*. Equation 1, and 2 describe the general form for the final value of R and T and the visual representation is showed in Figure 9.

$$R_n = R_n(R_{n-1} + T_{n-1}) + T_n \quad (1)$$

$$T_n = T_n + T_{n-1} \quad (2)$$

4. 3D Model Comparison

After having applied the registering algorithm, the new point cloud is converted into a mesh using Poisson Surface Reconstruction [10]. In Figure 10 (a) and (b) we see the first zone of the Inca Path and in (c) and (d) the second part. The colors of Figure 10 (a) vary due to the natural light conditions the day the samples were taken.

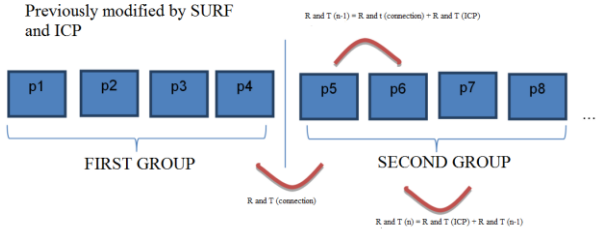


Figure 9: Accumulative process of operating with R and T for every point cloud.

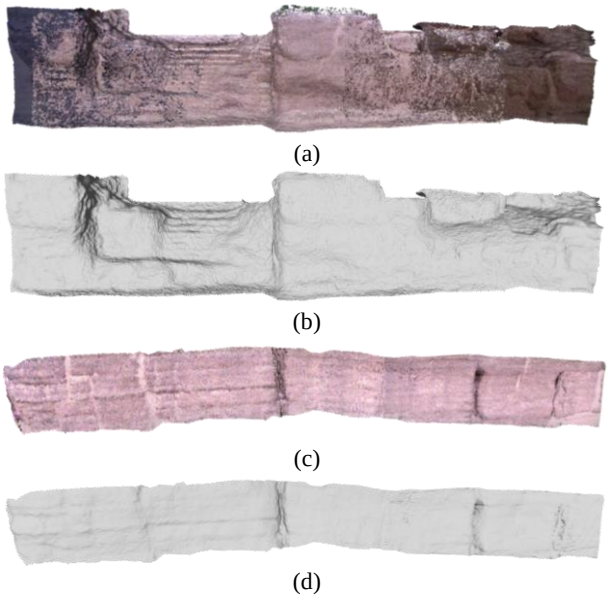


Figure 10: Registered point clouds (a) and (c), (b) and (d) represent the respectives 3D reconstructions.

Figure 11 shows the Kinect models in green and photogrammetry ones in full color. Due to the small area of data acquisition only a small part of the zone 1 was compared with the 3D laser Scanner. The differences between the 3D laser scanner and photogrammetry done in Agisoft is represented by a thermal scale image of the model compared and a histogram of the distance errors with a Gaussian adjust in Figure 12 and 13, respectively, extracted using the open source software CloudCompare (CloudCompare V2, v2.5, 2014).

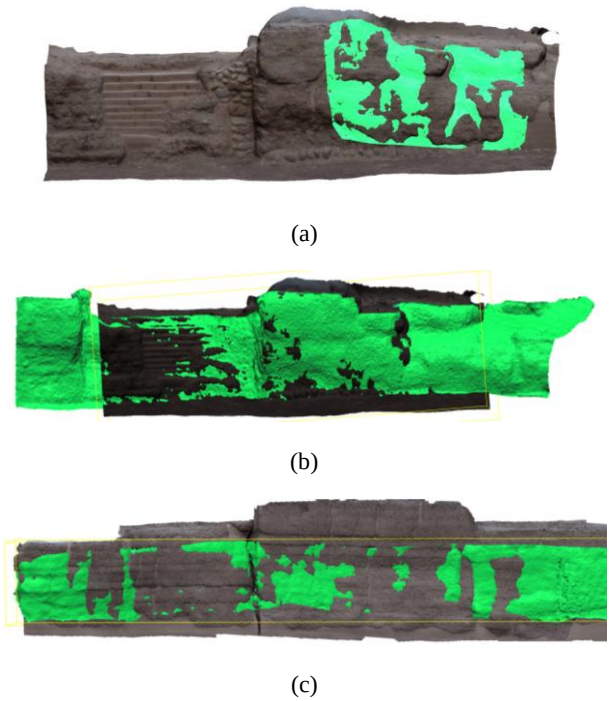


Figure 11: 3D reconstructions alignment of photogrammetry (color) and Kinect (green). Zone 1 of the Inca Path in (a) and (b) and zone 2 (c).

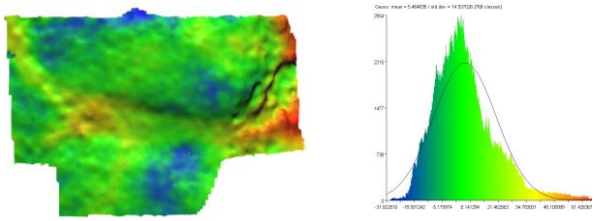


Figure 12: Comparison of the NextEngine Laser Scanner HD and Kinect from the first zone of the Inca Path.

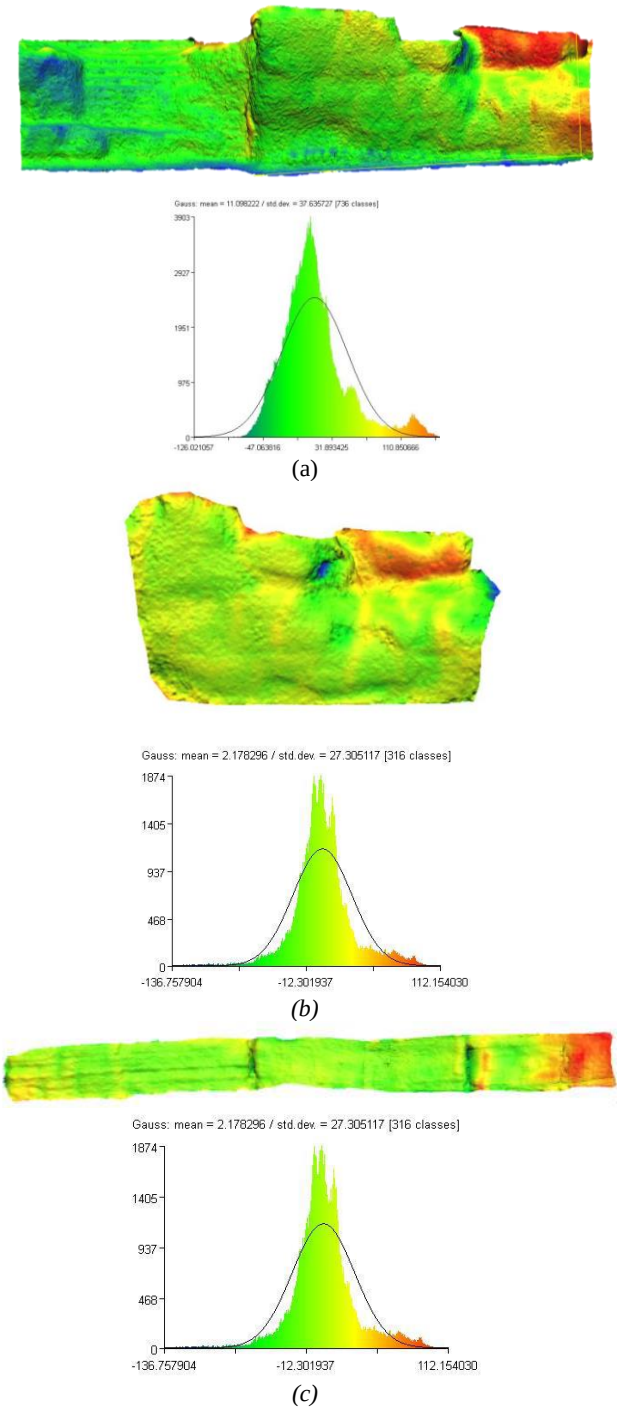


Figure 13: Comparison of photogrammetry and Kinect 3D reconstruction model. (a) and (b) from the first zone and (c) from the second one of the Inca Path.

The summery of the obtained error metrics can be found in Table 2. We found that the distance error between Kinect models and laser scanner is lower than those compared with photogrammetric models. Since laser scanner models are more accurate as described in Table 1, these results highlight the effectiveness of the reconstruction using Kinect and the proposed methodology. It is evident in Figure 13a and 13b that the error is higher (red areas) in regions more distant to the Kinect. This is due to the fact that, as demonstrated in the plot of Figure 2, the error increases as the distance of the target surface and the sensor increase.

I addition, we focused in the 3D reconstruction of cracks of the wall acquired previously. Accurate representations of cracks are fundamental in structural analysis for modelling stress and deformation which is useful for preservation and conservation of cultural heritage monuments [11]. Figure 14 shows the comparison of the crack geometric representation using both methods: Kinect and photogrammetry. The mean error is approximately 6 millimeters demonstrating that both techniques are highly similar in reconstructing complicated surfaces.

5. CONCLUSIONS

The Kinect for XBOX 360 is a potential alternative as a sensor for archaeological applications because of its low price, low acquisition time and an accuracy error of 1.49 centimeters when compared with a 3D laser scanner and a mean error of 4.74 centimeters with photogrammetry. This error may appear as significant; however, taking into account the dimension of the walls, the error represents the 5% of the total size of the model.

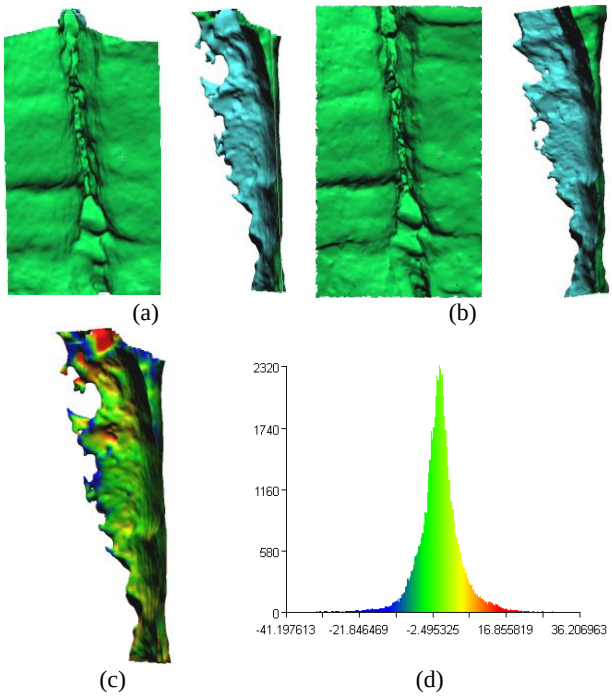


Figure 14: Comparison between 3D representations of cracks using Kinect and photogrammetry. (a) and (b) represents frontal an lateral view of the first crack in the model of Figure 10c, using Kinect and photogrammetry, respectively. In (c) and (d), the distance error (in millimeters) is calculated between models and represented in thermal scale image and histogram, respectively.

Table 2. Summary of the error values obtained when comparing Kinect models versus photogrammetric and laser scanner models.

Error	Comparison with photogrammetry			Comparison with NextEngine Laser Scanner HD
Accuracy (cm)	4.87	2.95	6.39	1.49

Table 3. Comparison between the techniques and devices used.

Parameters	Kinect for XBOX 360	Photogrammetry (SONY NEX-7)	NextEnginer Laser Scanner HD
Price (USD)	89.99	1,099.99 [12]	2,995 [13]
Samples quantity	200	30	20
Acquisition Time (min)	3	5	120

As we see in Table 3 it is by far the best option price-wise and provides very low acquisition time. Photogrammetry also has very low acquisition time but the processing time could reach up to 3 hours in a computer equipped with an Intel Core i7 processor with 4 cluster and 8 GB of RAM memory. The algorithm proposed takes a total processing time of 40 min in the same computer.

In addition, Kinect models reproduce the geometry of cracks barely the same resolution as models extracted with photogrammetric approach. A comparison between both models gives distance errors of 6 millimeters setting the photogrammetric model as reference. A more formal study using a time-of-flight laser scanner as a ground truth is needed in order to determine how accurate Kinect models are in representing cracks and structural failures.

This project deserves a more formal prospective study of the resolution of Kinect models. Since Kinect is able to acquire several samples per minute, the redundancy of points could be utilized for increasing the accuracy of models using a super resolution approach. For a country like Peru with great cultural and archaeological background, a device like the Kinect satisfies the fundamentals needs of price, portability and precision in order to digitalize and preserve the country's cultural heritage.

6. REFERENCES

[1] Koutsoudis, A., Vidmar, B., Ioannakis, G., Arnaoutoglou, F., Pavlidis, G., & Chamzas, C. (2014). Multi-image 3D reconstruction data evaluation. *Journal of Cultural Heritage*, *15*(1), 73-79.

[2] Guidi, G., Russo, M., & Angheleddu, D. (2014). 3D survey and virtual reconstruction of archeological sites. *Digital Applications in Archaeology and Cultural Heritage*, *1*(2), 55-69.

[3] Dellepiane, M., Dell'Unto, N., Callieri, M., Lindgren, S., & Scopigno, R. (2013). Archeological excavation monitoring using dense stereo matching techniques.*Journal of Cultural Heritage*, *14*(3), 201-210.

[4] Smisek, J., Jancosek, M., & Pajdla, T. (2013). 3D with Kinect. In *Consumer Depth Cameras for Computer Vision* (pp. 3-25). Springer London.

[5] Wan, Y., Wang, J., Hu, J., Song, T., Bai, Y., & Ji, Z. (2012, September). A Study in 3D-Reconstruction Using Kinect Sensor. In *Wireless Communications, Networking and Mobile Computing (WiCOM), 2012 8th International Conference on* (pp. 1-7). IEEE.

[6] Zvietcovich Zegarra, J. F. (2014). *Estimación del volumen de lesiones producidas por Leishmaniasis cutánea utilizando un escáner láser de triangulación 3D.* Undergraduate thesis. Pontificia Universidad Católica del Perú: Perú.

[7] Bay, H., Ess, A., Tuytelaars, T., & Van Gool, L. (2008). Speeded-up robust features (SURF). *Computer vision and image understanding*, *110*(3), 346-359.

[8] Juan, L., & Gwun, O. (2009). A comparison of sift, pca-sift and surf. *International Journal of Image Processing (IJIP)*, *3*(4), 143-152.

[9] Arun, K. S., Huang, T. S., & Blostein, S. D. (1987). Least-squares fitting of two 3-D point sets. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, (5), 698-700.

[10] Kazhdan, M., Bolitho, M., & Hoppe, H. (2006, June). Poisson surface reconstruction. In *Proceedings of the fourth Eurographics symposium on Geometry processing*.

[11] Brune, P., Perucchio, R. (2012). Roman Concrete Vaulting in the Great Hall of Trajan's Markets: *Structural Evaluation. Journal of Architectural Engineering* (18), 332-340.

[12] © Sony Electronics Inc. "Sony Online Store*",* California, United States. http://store.sony.com/alpha-nex-7-with-18-55mm-lens-zid27-NEX7K/B/cat-27-catid-Alpha-NEX-7-and-NEX-6?_t=pfm%3Dcategory (accessed 19 july 2014)

[13] NextEngine, Inc. "Everything you need to scan and build 3D models", California, United States. http://www.nextengine.com/products (accessed 19 july 2014).

7. ACKNOWLEDMENT

The authors would like to thank to Carlos Olivera, and San José de Moro Archaeological Program (Perú) for providing access to the site, technical information and professional assistance. This research was partially supported by the Interdisciplinary Research Project DGI 2013-0113 from the Pontificia Universidad Católica del Perú.