Computer Vision in Cultural Heritage:
a versatile and inexpensive tool

Claudio Giardino[1], Raffaele Rizzo[2]

[1] *University of Salento, Via D. Birago, 64, 73100 Lecce, Italy, claudio.giardino@unisalento.it*
[2] *University of Salento, Via D. Birago, 64, 73100 Lecce, Italy, raffaele.rizzo1@studenti.unisalento.it*

Abstract
The use of three-dimensional models is widespread in many sciences and has been in common use in archaeological studies, both in the phase of the study and in divulgation for a long time. When making decisions about the use of available technologies the following aspects should be taken into consideration: cost, accuracy, speed and easy of use. The Computer Vision is a relatively recent computer technology that is able to produce a three-dimensional reconstruction by a cluster of photos processed by automated software. The aim of this work is to evaluate the performance offered by the Computer Vision using low cost tools and software.

I. THE COMPUTER VISION, AN IMAGE-BASED TECHNOLOGY

Technologies are generally identified on the basis of the equipment used, followed naturally by the choice of different software.

The most common techniques used in the measuring of volume and surfaces for three-dimensional reconstruction use light radiation as a source of information and it is possible to make a macro division on available systems, based on the reception and interpretation of the light stimulus.

Thus, there are essentially two techniques: range-based (active sensors) and image-based (passive sensors). It is therefore important to emphasize that the various technologies should not be seen as competing with each other, but as various tools that can be used for one purpose [1, p. 191].

The principle of the range-based technique is the ability to emit light and to be able to observe, through a sensor, the response. The sensor device that is used is referred to as a range camera and one of the most widely used in archaeology is the 3D laser scanner.

Laser light, thanks to its peculiar physical characteristics, is the most commonly used light in the scanner, and it allows the creation of extremely focused spotlights even at long range intervals.

The use of these kinds of tools has reached their full maturity and today they are well structured and defined [2, p. 20]. Although active sensors are able to provide a 3D scale already metrically correct, the use of such technology requires high initial costs and qualified personnel.

Image-based techniques are referred to as "passive techniques", because the sensors used to capture images only records light in the environment and transforms it into digital information within image pixels. The structure from motion and image-based modeling software can then extract the information needed to create a 3D model from the images.

Computer Vision with software of image-based modeling and rendering (IBMR) is a relatively recent computer technology that is able to generate three-dimensional reconstructions from a series of photographic shots using automated software; it offers numerous possibilities for applications in archaeological research.

It originated with the aim of creating virtually starting from a selection of images, three dimensional models visible as to the human eye [3, p. 3].

The main advantage of this technology is the possibility of using low-cost a easily transportable equipment [2, p. 24], for example a common digital camera or even a smart phone. The computer programs are then almost completely automatic, and the operator's intervention is limited to the supervision of each processing phase and the correction of errors which occur during the developing process. They provide a more intuitive and simplified method than conventional geometric modeling software [4].

As imaged based technology, works through a combination of algorithms which, interact with digital images (pixel arrays that indicate brightness using a combination of Green-Red-Blue), additionally determining the shutter-release parameters of the machine (position and orientation) and features, common points between the various images. By combining this data with additional operating algorithms, a three-dimensional model is elaborated.

The theoretical basis of computer vision consists of studies made by the neuro-scientist David Marr [5, p. 25], who assimilated the human perception of reality to the mechanics of a computer for the first time.

According to Marr's theory software could elaborate the information on the basis of the integration of two images, in the same way as human sight.

To make this possible, it was necessary that the usable data were both unique (uniqueness), and also of a continuous nature as well as being of the same scene or object [6, p. 284]. It additionally required a correction to problems or limitations that might have emerged during the image association.

Assuming these requirements as indispensable, or "operational rules," Marr and Poggio elaborated an algorithm capable of reading and interpreting such evidence correctly [6, p. 285]. Lucas and Kanade's work [7] was the first step towards a revolution that would lay the foundations to elaborate computer vision; they enabled the creation of an image-technical-registration software, which is capable of converting a data set into a single coordinate system; the feature tracker (Kanade-Lucas-Tomas) was developed, a software that can determine the shooting position by using a gradient of a scalar function.

The SIFT (*Scale-invariant feature transform*) algorithm was developed by Lowe [8] in order to recognize a set of common features or characteristics in a series of images. This algorithm works by identifying some features and comparing these evidences in each new image: from the complete set of matching key points that are identified, all of which match position, scale and orientation of the object. The subsets are then filtered and reconditioned to eliminate false correlations.

II. CASE STUDIES

Three examples are presented here to analyze and they compare various operational steps based on the achievable results and the complexity of the work that was requested.

Archaeological Evidence:

A) Dolmen Stabile (or from Quattromacine) [9, pp. 347-365; 10, p. 110; 11, p. 75; 12, p. 140; 13, p. 75], a megalith from Salento located on the neighboring road of Quattromacine, Giuggianello (Lecce).
B) Burial V [14] of the Late Helladic necropolis of Kambi (Zakynthos, Greece).
C) Alabastron FS84 [14, p. 112; 15, p. 480; 16, p. 124] found in the Late Helladic necropolis of Kambi (Zakynthos, Greece).

Equipment:
A) Canon EOS 600D Camera;
B) Computer Toshiba Satellite C660D-1FV;
C) Smartphone with 13 Megapixel camera, Android 6;
D) Drone Phantom 3 standard.

Software:
Agisoft Photoscan, version 1.1.6.2038.

A. Photographic documentation

The photographic campaign provides data that can be analyzed by computer algorithms; its accurate execution decreases drastically (from a few millimeters to a few inches) the margin of error of the three-

dimensional model; it must take in consideration the needs of the software.

It must therefore consider some guidelines(Fig. 1):

a) a complete view of the object and where possible shot at 360°.
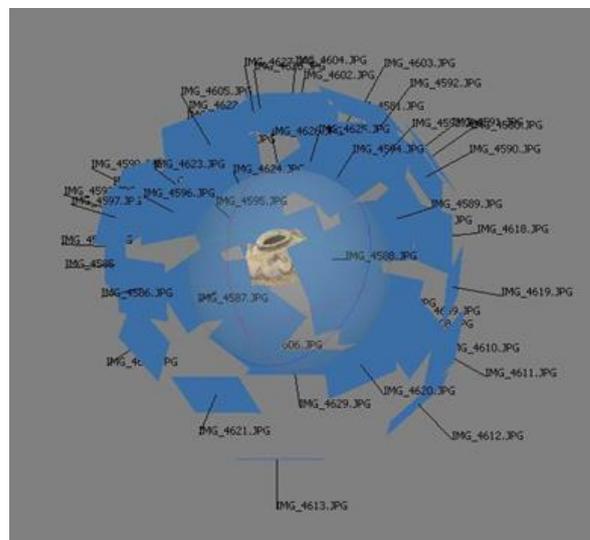b) it displays overlapping images at 40-60%.

Fig. 1. Features detection:the alabastron with parameters and positions of the camera.

Once the images have been captured a selection of those of the highest quality, optimum light and shadow conditions, among them should be made, in order to avoid subsequent interventions on the model. There is no exact limit on the images that can be submitted to the software.

A larger amount of data needs more processing time, but it does not guarantee more accurate results. However, a small number of images can affect reconstruction by producing black or empty dots in the model. In this case, it is possible to perform corrections in successive steps, adding new photos or data, or creating meshes by using the appropriate tools.

In the last few years, the use of unmanned aerial vehicles (UAV or drones) has made it possible to take shots from perspectives and heights otherwise difficult to obtain; in addition, the geographic coordinates are recorded by the drone driving system itself.

For a detailed 3D reconstruction, you must first perform a "calibration" procedure [17, p. 272] to improve camera precision and shutter quality. A 3D back-up with a calibrated camera can give up to 10 times more accuracy than the same calibration return. Some algorithms, such as bundle adjustments, can simplify the calibration procedure by identifying unknown system parameters: the orientation of the camera (internal and external), the 3D coordinates of the homologous points measured in the images and the additional parameters [1, pp. 185-186].

In the case of restitution 3D of the Mycenaean alabastron it was necessary to operate on a neutral background, using a white cloth mounted on supports; a lamp, placed at 90° from the alabastron, thus providing a uniform vision of light and shades. As a

whole, 47 photo shots were taken, including some showing just handles and decorations, made in ten minutes (Fig. 2).

Fig. 2. Photographic documentation: the alabastron

The Mycenaean burial site was taken at 360°, making a series of shots that could be overlapped by 30-40 %, and photographing of the tomb and the surrounding area with an inclination of 90°, 45° and 30°. Many photographs were taken in order to reconstruct some surfaces in a more detailed way - such as walls or internal burial sites - by proceeding from a longest distance to a closer one. So that the software, consequently had no difficulty in highlighting common points, and were then used to rebuild the model. The need for shots of 90° required the use of a pole to support the camera; while for 45° shots a tripod stand was preferred (Fig. 3). Shooting required optimum light conditions so that any shadows would not be shot. The set of shots was taken in 15-20 minutes.

Fig. 3. Photographic documentation: Kambi, burial V.

The photographic campaign of the dolmen structure was the subject of two different experiments:
a) use of the drone for the 90° and 45° shooting;
b) replacement of the Canon EOS 600d camera with a smartphone (13 MP camera).
This is to verify both the interaction with the aircraft and also the use of computer vision thanks to a tool which is not only cheap, but that is also daily use, and is therefore easily available even in emergency

conditions. In total, 179 shots were made, of which 11 thanks to the drone aid.

B. Model processing methodology

The first step is to realize the "feature detention"; organized into different processes, which lead to the creation of a cloud of low density points. Firstly the analysis of the features, small portions of images, allowing the location of the position, and the extension and the color gradient [18, p. 104]. The selected points must have characteristics that are clearly visible in most images, to make proper matches, so that they are "key point features". In this first operational phase, the Lowe SIFT algorithm is fundamental [8, pp. 25-26], because it guarantees excellent results even with scale or orientation variations between the various images. Subsequently, a special algorithm crosses the data emerging from each shot, reconstructing the symmetries between the photos.

The parameters and position of the camera are defined thanks to the simple principle that objects close to the observation point move in space faster than distant elements [18, p. 104]. In order to achieve this aim, modern operating programs use algorithms capable of synthesizing Structure from Motion, Image-Based Modeling and Image-Based Rendering softwares. So, once the parameters of the inner and the outer orientation of each frame (motion) are obtained, as well as of the geometry of the scene (structure), a cloud of dotted points is created (Fig. 4a-c).
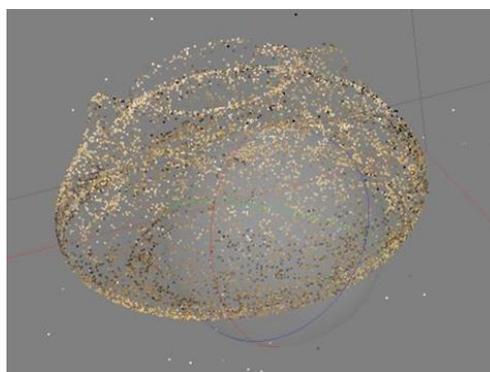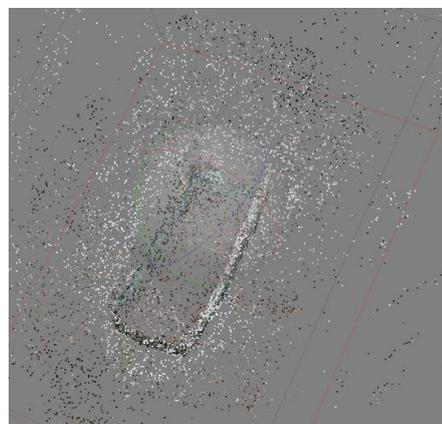

Fig. 4a. Feature detection: the alabastron.

Fig. 4b. Feature detection: Kambi, burial V.

Fig. 4c. Feature detection: Dolmen Stabile.

It is based on the alignment of the various common points highlighted in the individual photos and is already capable of expressing the main features of the model. From this first step one can notice the quality of what will be the final result, identifying errors in the processing, if there are any.

It is only possible to continue working after making the necessary corrections, which may include adding or deleting some photos, or even manually evidence markers to facilitate the reconstruction of some critical areas (key markers are fundamental if you want to use them to associate GPS coordinates). The model obtained, however, represents only a scattered reconstruction of the 3D scene; but also dense cloud is needed.

The use of algorithms of "dense image matching", though not necessary for model realization, improves the accuracy of the final product. Photo shots are re-analyzed and compared.

Two operations, expand and filters, are performed [18, p. 104] respectively to expand areas that are closer to the detected points and to correct any errors. Although these processes are largely automatic, the operator can discretely perform a "cleansing" of the dense cloud to eliminate non-essential elements.

The pattern skeleton arises from the correct succession of these operations (Fig. 5a-c), which will then be integrated with mesh and text modeling processes, and extrapolated from the images used.

Fig. 5a. Dense cloud: the alabastron.

Fig. 5b. Dense cloud: Kambi, burial V.

Fig. 5c. Dense cloud: Dolmen Stabile.

The dense cloud will then be converted into a polygon surface (mesh) (Fig. 6a-c) and then "coated" with digital images (texture) (Fig. 7a-c). During mesh processing, it is possible to intervene in any shadow zones or blind spots by initiating a correction made by the software that, thanks to the features detention, will integrate the 3D.

Fig. 6a. Mesh modeling: the alabastron.

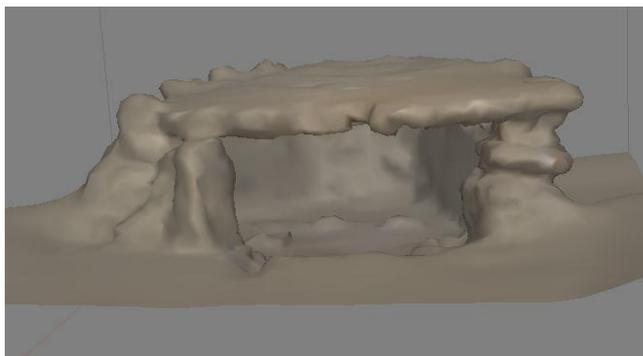*Fig. 6b. Mesh modeling: Kambi, burial V.*

Fig. 6c. Mesh modeling: Dolmen Stabile.

In the course of the investigation, it was observed that the presence of a neutral background during the photographic campaign results in being substantial, though not indispensable, for the realization of the models. This background made it easier to identify common points during the features detection process and significantly reduced the need for the corrective action of an operator.

The realization of the three-dimensional model of the structures did not cause any particular difficulty, but took longer if compared to the 3D reconstruction of the alabastron; this was due to a greater amount of software data and of common points among the different photos. It was therefore crucial to have an accurate photographic campaign that would ensure shots with similar light conditions.

The use of the drone and the use of the tripod for the camera proved to be equally valid; however, the use of the drone allowed an automatic recording of the GPS coordinates and increased control during shots, but it required optimum flight conditions and the help of a specialized pilot. The use of the smart phone did not result in any qualitative loss of the data which was collected or in complications of the various steps of the work.

Fig 7a. Texture modeling: the alabastron.

Fig 7b. Texture modeling: Kambi, burial V.

Fig 7c. Texture modeling: Dolmen Stabile.

III. CONCLUSION

The use of computing resources has provided a much better and greater understanding in archaeological studies, and digital archeology [19] is now a consolidated reality.

The possibility of using 3D programs has revolutionized the documentation and the study of cultural heritage. The use of three-dimensional models has rapidly spread in archaeological research, and computer vision is an innovation that offers new solutions and benefits, allowing a good combination of low cost, accuracy, rapidity, and manageability [20, p. 1; for a comparison between range-based and image based technologies, see 21, 22].

The data collection phase can be achieved without using any special equipment, it is easily transportable and cheap tools such as a camera or a smart phone, are able to achieve valid results. In addition, the process based on the elaboration of data, and the consequent realization of a three-dimensional model is implemented thanks to a few intuitive steps (features detection, dense cloud, mesh and texturing modeling) which the software can perform independently.

The intervention of an operator is only required to set the operating parameters of the various steps and to correct any imperfections. However, operations do not require specific computer skills, but only an overall and clear knowledge of the software used by the program. Even if it is possible to use online programs, such as web service or open source, that independently perform the processes (the work by Nguyen, Wünsche and Delmas [5] can be a useful starting point for discerning the potentialities and limits of each software, thus identifying the most appropriate field of action), it is

always preferable the continued supervision of the operator.

The speed of the data collection that C.V. offers, compared to other technologies, has been judged to be extremely useful during excavations, above all in situations where it is often required to work fast [23, p. 283]. New resources are currently being developed, programs which can be applied to archaeological research and are characterized by the use of computer vision, as in the case of the use of feature detection software to compare and catalog archaeological materials [24, pp. 477-480].

Using these technologies, it is possible to re-use old photographs in order to create three-dimensional models and also highlight damaged or missing evidence. This is a characteristic that opens up new horizons to investigating and enhancing the heritage of the past. A powerful tool for Cultural Heritage protection is the opportunity to create and to compare cheap 3D reconstructions of monuments. During our research, we compared old photos and drawings of the dolmen Stabile with its 3D model, realizing that some structural parts disappeared in the last 50 years.

In conclusion, Computer Vision does not have to be considered an enemy of older and more traditional methods, technological or not, but it can be quantified as a new and important resource, which is able to bring significant advantages to the discipline of archaeology by making it possible to employ an increased and improved use of three-dimensional models for the creation of reconstructions or virtual realities.

REFERENCES

[1] M. Russo, "Principali tecniche e strumenti per il rilievo tridimensionale in ambito archeologico", Archeologia e Calcolatori, 22, 2011, pp. 169-198.

[2] P. Cignoni, R. Scopigno,"Sampled 3D models for CH applications: A viable and ena-bling new medium or just a technological exercise?", Journal on Computing and Cultural Heritage (JOCCH), 1, 2008, pp. 1-23.

[3] R. Szeliski, "Computer Vision: Algorithms and Application", Berlin, 2011

[4] H. M. Nguyen, B. Wünsche, P. Delmas, "3D Models from the Black Box: Investigating the current state of Image-based modeling", International conference on computer graphics, visualization and computer vision, Praha, 2012, pp. 115-123.

[5] D. Marr, "Vision: A computational Investigation into the Human Representation and Processing of Visual Information", New York, 1982.

[6] D. Marr, T. Poggio,"Cooperative Computation of Stereo Disparity", Scienze, 194, 1976, pp. 283-287.

[7] B. D. Lucas, T. Kanade, "An iterative image registration to technique with an application to stereo vision", IJCAI'81 Proceedings of the 7th International Joint Conference on Artificial Intelligence, Vancouver, 1981, pp. 674-679.

[8] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Key points", International Journal of Computer Vision, 60, 2004, pp. 91-110.

[9] R. Whitehouse, "The megalithic monuments of south-eastern Italy", Man, 2, 1967, p. 347-365.

[10] F. Jesi, "Il linguaggio delle pietre", Milano, 1978.

[11] M. Cipolloni Sampò, "Manifestazioni funerarie e struttura sociale", Scienze dell'Antichità, 1, 1987, pp. 55-120.

[12] P. Malagrinò, "Monumenti megalitici in Puglia", Fasano, 1997.

[13] L.Coluccia, M .Merico, "Monumenti megalitici in Puglia", Le orme dei Giganti, Palermo, 2009, pp. 75-82.

[14] P. Agallopoulou, "Μυκηναικον νεκροταφειον παρα το Καμβι Ζακυντηου, Αρχαιολογικον Δελτιον", Athens Annals of Archeology, 28, 1973, pp. 103-116.

[15] P. A. Mountjoy, "Regional Mycenaean decorated pottery", Berlin, 1997.

[16] Ch. Souyoudzouglou-Haywood, "The Ionian islands, The Bronze Age and Early Iron Age 3000-800 BC", Liverpool, 1999.

[17] F. Remondino, El-Hakim,"Image-based 3D modelling: A review", The Photogrammetric Record, 21, 2006 ,pp 269-291.

[18] A. Bezzi, L. Bezzi, "Computer Vision e Structure From Motion, nuove metodologie per la documentazione archeologica tridimensionale", Open source, free software e open format nei processi di ricerca archeologica, Bari, 2011, pp. 103-111.

[19] T.L. Evans, P. Daly, "Archaeological theory and digital pasts", Digital Archaeology. Bridging method and theory, Oxford, 2006, pp. 2-7.

[20] M. Lo Brutto, P. Meli, "Computer vision tools for 3D modelling in archaeology", Palermo, 2012.

[21] J. Knibbe, K. P. O'Hara, A. Chrysanthi, M. T. Marshall, P. D. Bennett, G. Earl, S. Izadi, M. Fraser, "Quick and Dirty: Streamlined 3D Scanning in Archaeology", The 17th ACM Conference on Computer-Supported Cooperative Work & Social Computing, Baltimora-Febbraio, New York, 2014, pp. 1366-1376.

[22] T. P. Kersten, M. Lindstaedt, "Image-Based Low-Cost Systems for Automatic 3D Recording and Modelling of Archaeological Finds and Objects", Progress in Cultural Heritage Preservation, October 29 – November 3, Limassol (Cyprus), 2012, pp. 1-10.

[23] M. Sfacteria, "Foto modellazione 3D e rilievo speditivo di scavo: l'esperienza del Philosophiana Project", Archeologia e Calcolatori, 27, 2016, pp. 271-289.

[24] L. Van der Maaten, P. Boon, G. Lange, H. Paijmans, E. Postma, "Computer Vision and Machine Learning for Archaeology", Digital Discovery. Exploring New Frontiers in Human Heritage. CAA2006. Computer Applications and Quantitative Methods in Archaeology, Fargo, 2006, pp. 476-482.