Novel imaging of heritage objects and sites

Paul Bourke
iVEC @ UWA
The University of Western Australia
Perth, Western Australia 6009, Australia
Email: paul.bourke@uwa.edu.au

***Abstract*—Presented are three novel imaging techniques increasingly being used to record heritage indigenous sites, namely, 3D model reconstruction derived solely from photographs, high resolution panorama imaging and high resolution photographic mosaics. A highly desirable feature these techniques have in common is that they only require a good digital camera making them a viable option for recording at remote sites. In addition they involve largely automated processes and do not require specialist training. Examples will be presented of these technologies for heritage object and site recordings by researchers at The University of Western Australia. For each capture modality the principles will be outlined, the relative merits and limitations will be discussed along with their respective applications for research and as a means of creating valuable digital assets.**

***Keywords*—Photogrammetry, 3D reconstruction, mosaic, gigapixel, panorama, photography, feature points, image stitching.**

I. Introduction

The digital recording of heritage objects and sites can be challenging for a number of reasons. Objects are often delicate, they may be culturally sensitive with limited access, located at remote locations and may be unique or precious. These and other reasons can mean it may not be possible for them to be moved to a laboratory for scanning. They may also require non-intrusive scanning since they may not be able to be physically touched or marked. Challenging characteristics of heritage sites may be their remoteness, lack of electricity and they may require access on foot or at least involve a minimum payload due to access by small planes, boats or vehicles.

The researcher who gains access to these objects or locations in the face of these constraints would still like to make a recording that results in maximum value. That value may be measured in different ways including to the research process, as a digital record/archive, or for public dissemination and education through museums or art galleries. For this reason many involved in heritage recording have, over the years, been at the forefront of trialing new technologies for both the capture and presentation of digital heritage.

In what follows three technologies will be discussed in the context of their use by archaeologists at The University of Western Australia. A characteristic of archaeology in Western Australia is the often remoteness, cultural sensitivity and climatic harshness of the sites. As a result it may not be possible to have large groups operating in the field and there may be limited scope to which large equipment payloads can be transported, indeed some specialist equipment may not even survive or be suited to the conditions. The technologies presented satisfy the above mentioned constraints. They are based only upon access to a good camera, the imaging is non-invasive and can be readily transported to remote locations.

II. 3D Model Reconstruction

Photogrammetry is the general term given to the general process of deriving some 3D quality from photographs, typically two or more. While it is almost as old as photography itself, more recently, due to improvements in algorithmic developments in computer science, and in particular machine vision [1,2,3], it is increasingly possible to create fully textured and high quality 3D digital models solely from a collection of photographs. This presents an exciting new opportunity for site recording, creating a digital model of the three dimensional structure rather than just flat photographs. Given a 3D model, views can be derived from positions other than the original photographs and structural measurements and analysis performed, see fig 4.

In the context of remote sites it offers significant advantages over more traditional approaches such as laser scanning, in particular, it does not require any heavy equipment other than a camera, an important consideration for sites only reachable by foot. Compared to return of flight methods it more naturally deals with convoluted geometry, see fig 2, and can be performed in a wider range of environmental conditions. Depth camera or laser scanning generally require multiple scans and often a time consuming combining of the point clouds from each scanner position. Finally, the texture quality from photogrammetric methods is generally much higher than alternative approaches, not surprising since it is based upon a photographic process in the first place.

An active area of research involves 3D reconstruction from ad-hoc photographs, such as those found in on-line photography collections. However in order to achieve undistorted models a more robust process is required, in particular, uniform camera parameters. Applying some rigour to the photographic process can result in accurate and dimensionally correct models even when the scene is still free of any in-scene markers (often not possible for heritage objects).

There are three general topological categories which can require different photographic strategies. They are:

- Single objects but ones that are largely a surface, an extruded plane. These are typically able to be photographed from arbitrary positions around the object where the whole object is contained within each photograph. An example is shown in fig 1, an arrangement of marker stones by indigenous Australians.

- Extended objects where the photographs are generally taken along paths roughly parallel to the geometry and each photograph only contains a small region of the overall structure to be reconstructed. The key with these objects is to ensure sufficient overlap between photographs and multiple views of any one portion of the object.
- More challenging are self contained objects with convoluted geometric structure. Fig 2 is an example of a non-trivial toroidal structure, a headdress used as part of an indigenous dance.
- The most challenging is a combination of the above where there may be large scale concave structures as well as localised objects. Fig 3 is an example from a rock shelter with a central pile of rocks and a pit that is part of the archaeological excavation.

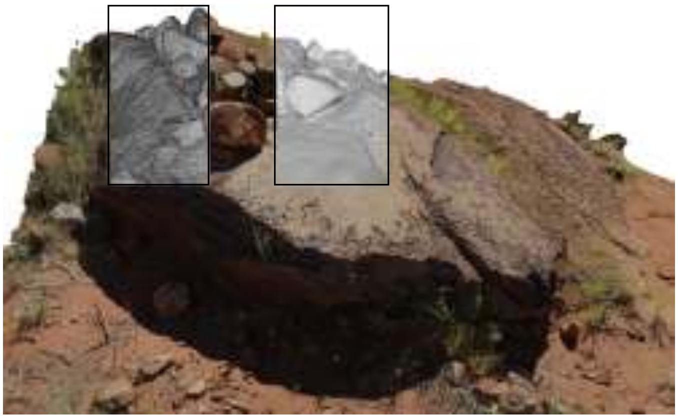

Fig. 1. 3D capture of "marker stones". Ngintaka story project, Museum of South Australia. Inset regions show mesh and geometric structure beneath the texture.

Fig. 2. Indigenous headdress reconstructed from 120 photographs. Textured mesh (left) and diffuse shaded raw mesh geometry (right). Ngintaka story project, Museum of South Australia.

The pipeline for 3D reconstructions of the sort presented here is as follows:

- Feature point detection [4,5,6], generally between all pairwise combinations of the photographic collection. The performance of this process can be improved if something is know about the order of the photographs.
- Numerical solving to estimate the camera positions and intrinsic properties of the cameras and at the same time deriving the 3D positions of the feature points. This is generally referred to as Structure from Motion (SfM) process, one implementation referred to as the Bundler algorithm.
- With a knowledge of the derived camera positions and the feature points one can now derive a denser point cloud.
- Form a, generally triangular, mesh over the dense point cloud.
- Re-project the photographs from each camera position and blend to form a texture set for the mesh.
- Perform various post processing stages such as removing unwanted parts of the model, closing holes, subsampling the mesh or texture resolution before exporting the model.

The process as described is largely automatic except for various algorithm parameters depending on the nature of the object being reconstructed, the type of images, and the desired object quality which can affect processing time. The only manual aspect is the last one depending on the degree of model cleaning required.

An important consideration in such reconstructions is the actual geometric resolution as opposed to the apparent resolution. Apparent resolution is referring to the appearance of geometric detail conveyed visually through the high quality textures. The relative importance of these two sources of detail depends on the intended application for the reconstructed object. For example as a digital asset of the object/site then one strives for the highest resolution for both the visual detail (texture) and geometry. For measurement or structural analysis one may not be interested in the texture resolution at all. Whereas for real time environments there can be constraints on the geometric resolution supported and providing apparent detail through good quality textures is acceptable. For educational and practical delivery of models online a compromise may be required. It should be noted that geometric detail is the most difficult to achieve and currently the area of active research and benefits from practical experience in the photographic acquisition. The relative importance of these two sources of detail is summarised in table 1.

Application	Geometric detail	Texture detail
Virtual environments	Low	High
Geometric analysis	High	Low
Education	Medium	High
Archive	High	High
Online	Low/average	Average

Table 1. Relative importance of sources of detail.

III. HIGH RESOLUTION PANORAMA IMAGING

Photographs are the most straightforward method of creating a digital record, albeit only a 2D projection, of heritage objects and places. Due to the upper limit of resolution of any digital camera system the tradition is to take close photographs of significant features but this fails to capture the context. Distant photographs can be taken to record the larger context but this fails to capture the fine details. The solution to capturing both the context and the detail in a single image is to take a large number of photographs, possibly but not necessarily arranged on a regular grid. If these photographs have sufficient overlap then computer algorithms can first find corresponding points between the photographs and then blend the photographs together to form a single large composite image.

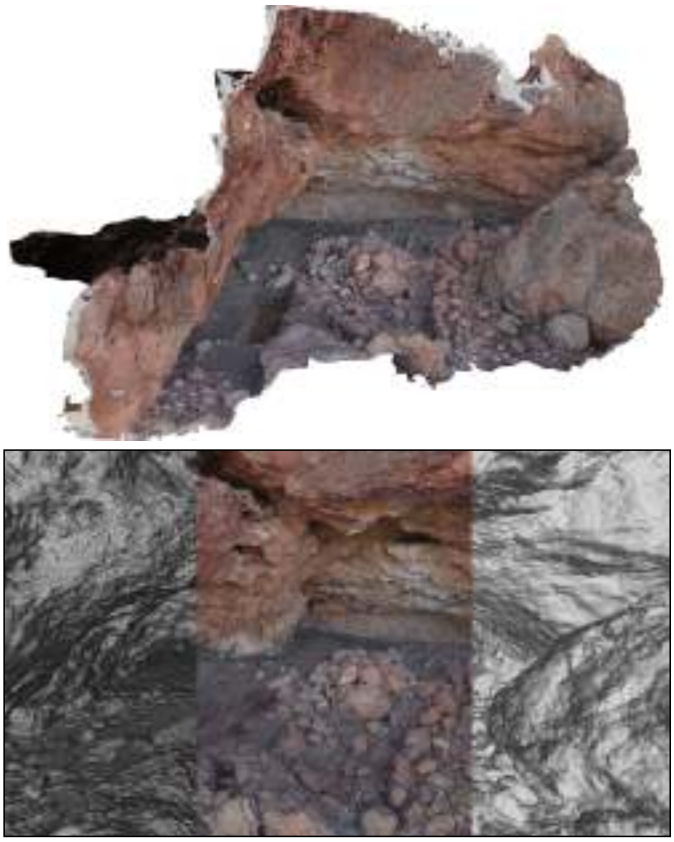

Fig. 3. Rock shelter reconstruction, Yallabilli Mindi. Constructed from 220 photographs. Lower image on left and right illustrates the underlying geometric data.

The final geometry of the composite photograph, usually referred to as a panorama, is commonly one of three types.

- Bubble photographs also known as called spherical panoramas or more formally as equirectangular projections [7]. The example in fig 5 is a complete 360x180 degree field of view panorama. In conjunction with immersive displays these images are often used to give one the sense of a place. Since everything visible from the camera position is recorded they also provide the ability to synthesize any other projection from that position.
- Partial (longitude) or full cylindrical projections [8]. An example is shown in fig 6 of a 220 degree horizontal panorama with a 30 degree vertical extent.

Perspective projections. These are essentially the same projection coverage as if one took a single lower resolution photograph, but at higher resolution. Fig 4 is an example of such a planar projection from a Western Australia rock art site.

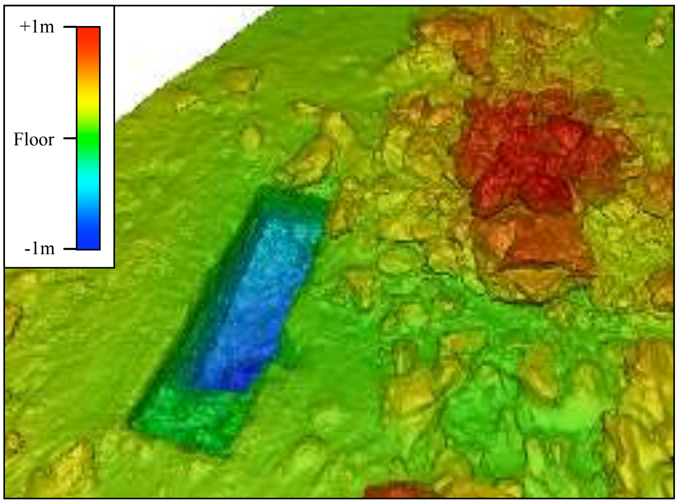

Fig. 4. Dimensions, contouring, and other structural analysis is possible from reconstructed representation, in this case the floor of the Yallabilli Mindi rock shelter shown in fig 3.

The ultimate resolution of these panoramas is a function of the number of photographs taken. This in turn is dictated by the field of view (FOV) of the lens, the smaller the FOV the higher the final resolution. Note that for dynamic scenes the time required may limit the number of photographs that can be taken before the scene changes too much. With care the collections of photographs making up these images can be taken by hand, with a simple tripod, using a precise camera rig designed for the task, or using an automatic robotic system. More dynamic scenes can be captured with camera arrays, this includes the possibility of capturing full video, these topics are considered out of the scope of the discussion here which is focusing on stationary objects and sites.

Fig. 5. 360 degree longitude x 180 degree latitude "bubble". Long Island, Houtman Abrolhos Island group, Batavia Dutch shipwreck project.

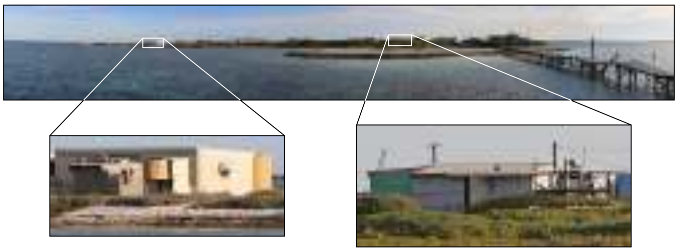

Fig. 6. Cylindrical panorama projection, formed from 50 images for a total of 750MPixels image. 220 degree longitude x 30 degree latitude panorama. Beacon Island, Houtman Abrolhos Island group, Batavia Dutch ship wreck project.

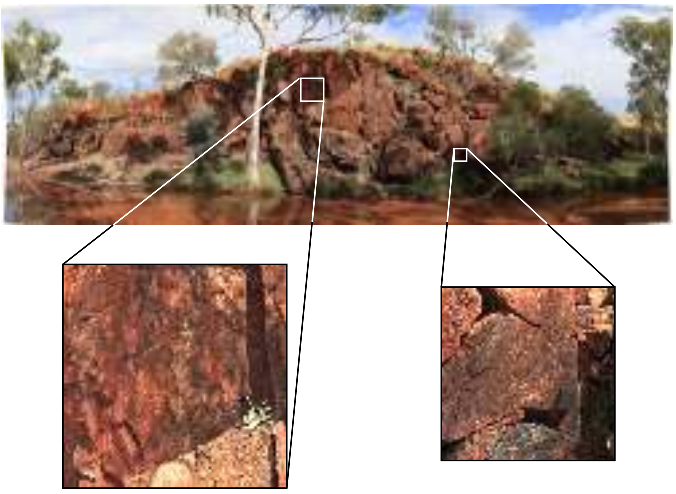

Fig. 7. 1.2GPixel [9] planar capture. Wanmanna rock art site, Newman, Western Australia.

The pipeline to form these high resolution images and for the image mosaics discussed next is similar. As with 3D reconstruction the process starts by looking for matching feature points between the images. If the images are laid out on a regular grid or if they are captured sequentially this information can be used to improve the quality and performance of the feature point search. The images are then linearly, or in the case of mosaics non-linearly warped to form the composite image. Finally regions across the overlap zones are blended together [10] to render final seamless image. Current algorithms, given a fixed nodal point camera arrangement and a calibrated lens can result in a high quality blend with minimal degradation of the input photographs across the blend zones. The most common cause of ghosting and image softness across the blend zones is due to movement within the scene.

IV. HIGH RESOLUTION PHOTOGRAPHIC MOSAICS

Presented here will be so called "photo-mosaics", although there is some confusion with the terminology. They are variously called photo-collages and image mosaics, and all three terms are also used to refer to an entirely different form of image generation. In the high resolution panorama photography of the previous section it is generally assumed the camera is rotated about its nodal point, indeed great care is usually taken to achieve this in order to minimise stitching and blending artifacts arising from parallax errors. In what follows the term photo-mosaic will be taken to refer to a composite image made up of a number of smaller photographs where the camera position varies, either by choice or necessity.

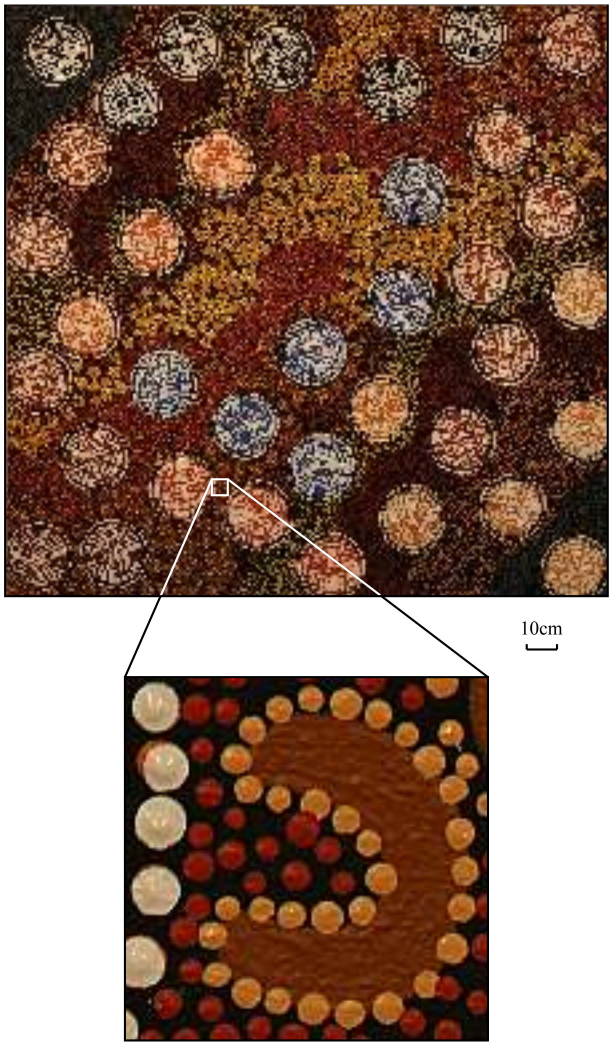

Fig. 8. 1GPixel image mosaic resulting from a 8x8 array of photographs. Indigenous dot painting by Margaret Whitehurst.

An example of capturing a photomosaic [11] of a flat painting is given in fig 8, the image presented is created from a 8x8 grid of photographs taken with a 20MPixels camera and 100mm lens. The resolution of the final image (for a fixed camera sensor resolution) is a function of the field of view of the lens, a narrower FOV means more photographs need to be taken to cover the object in question and therefore a higher final mosaic resolution. This scaling characteristic continues up to the limit or narrow field of view (FOV) quality zoom lens on the market. The same feature point detection, stitching and blending algorithms that are used for panoramas can be deployed here.

Strictly speaking, a photomosaic cannot be perfect except for perfectly planar objects [12,13]. One reason why a perfect stitching cannot be achieved for a scene where there are objects at different depths is illustrated in fig 9. This shows the side view of two camera positions with a blue sphere and red cube at different depths in the overlap zone between the two cameras. It is clear that it is not possible to seamlessly blend these two photographs together due to the relative locations in the two photographs of the red and blue objects. A second reason relates to parallax, elevated objects may have one side visible from one camera position but not the other. In the example shown in fig 9 the right hand side of the red box is visible to the right hand camera but not the left hand camera, as such a perfect stitch is not possible in the final mosaic. Note that it is possible to blend two images such as the ones illustrated in fig 9 for any particular depth, but not all depths at once. The situation improves for very narrow FOV lenses which increasingly approximates a parallel projection.

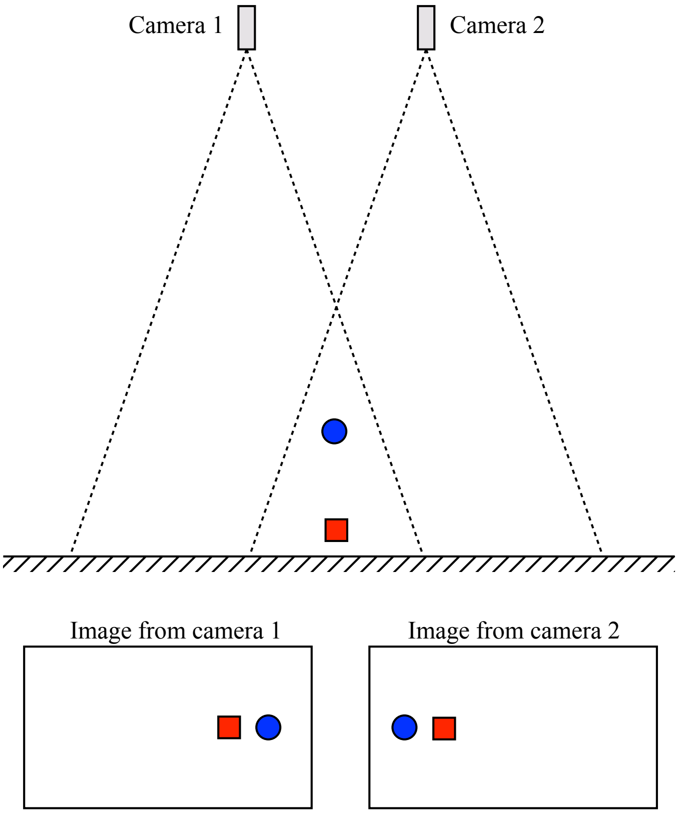

Fig. 9. Illustration of one reason why perfect stitching/blending with non-coincident camera is impossible.

The stitching limitations shown in fig 9 is a well known effect [14,15] for multiple camera video rigs such as the various GoPro rigs and the LadyBug series of cameras. While every attempt is made to locate the nodal point of the cameras as close together as possible, the parallax error still exists to some extent and fundamentally limits the stitching quality.

All is not lost, in many site recordings there are still benefits in being able to photographically map a large area despite potential errors in the stitching. The feature point detection and stitching/blending will warp the images in the face of these parallax errors. What one loses is the ability to reliably measure distances and angles since the image warping can be arbitrary in the algorithms attempt to collocate the detected feature points. Fig 10 illustrates a linear mosaic across a cliff wall covered by indigenous rock art. Fig 11 is an example from an unmanned aerial vehicle, namely an octocoptor, capturing a 2 dimensional grid of a significant rock mound. These are both cases for which perfect panoramas are not possible, in the first there is no single vantage point with a view of the whole cliff face, in the second the remote location and cost prohibit any scaffolding being build to gain an aerial view of the whole mound.

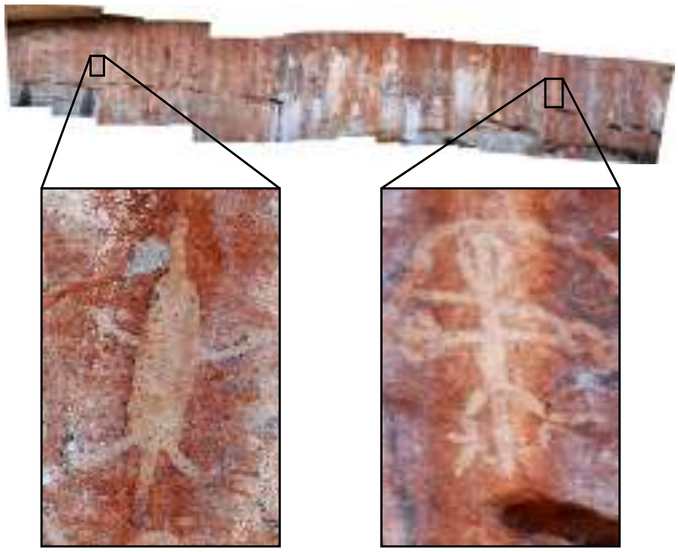

Fig. 10. Rock art cliff face. Only 16 photographs in the mosaic but no single vantage point exists for a single photograph or panorama.

Conclusion

Presented here are three data capture technologies which, while not necessarily new, have in recent times been undergoing steady improvements of the algorithms and as such are increasingly being applied as a means of capturing valuable digital assets of heritage objects and sites. A key characteristic of these technologies is that they don't rely on any hardware other than a good digital camera, a tool that is familiar, does not require extensive training, is readily transportable, and suffers from few environmental restrictions. All these characteristics are important in many heritage sites that may in challenging remote locations, and objects that may not be readily relocated or damaged.

Acknowledgments

The work was supported by iVEC through the use of advanced computing resources located at iVEC@UWA.

The Weld Range Web of Knowledge Project is funded by the Federal Government via the Indigenous Heritage Program. 3D visualisation was undertaken in collaboration with Ethical Engagement Consultancy and with assistance from Sinosteel Midwest Corporation.

Indigenous artwork "SKA Satellites on the Murchison" by Margaret Whitehurst.

Beacon Island recording managed by WA Museum and funded by Your Community Heritage: Batavia National Heritage Listing Grant from the Department of Environment.

Rock art access in conjunction with the Centre for Rock Art Research and Management, The University of Western Australia, with financial support from BHP Billiton.

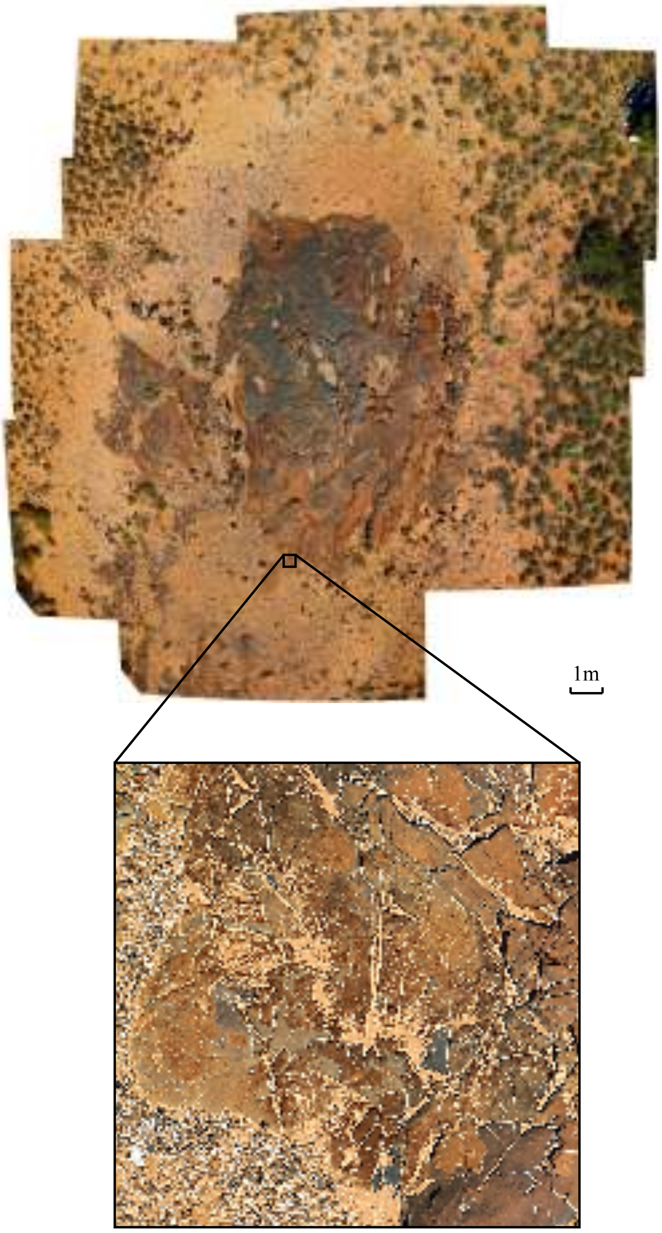

Fig. 11. 600MPixel image mosaic [16] from unmanned aerial vehicle (UAV) comprising of 50 individual photographs.

REFERENCES

[1] N. Snavely, S. M. Seitz, R. Szeliski. “Photo Tourism: Exploring image collections in 3D”. ACM Transactions on Graphics (Proceedings of SIGGRAPH 2006), 2006.

[2] O. Faugeras, S. Maybank. “Motion from point matches: multiplicity of solutions”. International Journal of Computer Vision, 4(3):225-246, June 1990.

[3] C. Tomasi, T. Kanade. “Shape and motion from image streams under orthography: A factorization method”. International Journal of Computer Vision, 9(2):137-154, 1992.

[4] J. Shi, C. Tomasi. "Good Features to Track,". 9th IEEE Conference on Computer Vision and Pattern Recognition. Springer. June 1994.

[5] D. G. Lowe. "Object recognition from local scale-invariant features". Proceedings of the International Conference on Computer Vision 2. pp. 1150–1157, 1999.

[6] D. G. Lowe, “Distinctive Image Features from Scale-Invariant Keypoints”, International Journal of Computer Vision, 60, 2, pp. 91-110, 2004.

[7] J. P. Snyder “Flattening the Earth: Two Thousand Years of Map Projections”. pp. 5–8, ISBN 0-226-76747-7, 1993.

[8] J. P. Snyder. “Album of Map Projections, United States Geological Survey Professional Paper”. United States Government Printing Office. 1453, 1989.

[9] M. Ben-Ezra. “High Resolution Large Format Tile-Scan - Camera Design, Calibration, and Extended Depth of Field”. In ICCP, Mar 2010.

[10] S. Wang, W. Heidrich. “The design of an inexpensive very high resolution scan camera system”. In Computer Graphics Forum, volume 23, pages 441-450. Citeseer, 2004.

[11] G. Ward. "Hiding seams in high dynamic range panoramas". Proceedings of the 3rd symposium on Applied perception in graphics and visualization. ACM International Conference Proceeding Series 153. ACM, 2006.

[12] S. Suen, E. Lam, K. Wong. "Photographic stitching with optimized object and color matching based on image derivatives". Optics Express 15 (12): 7689–7696, 2007.

[13] T. Shimizu, A. Yoneyama, Y. Takishima, "A fast video stitching method for motion-compensated frames in compressed video streams", International Conference on Consumer Electronics, 2006.

[14] M. Uyttendaele, A. Eden, R. Szeliski, "Eliminating ghosting and exposure artifacts in image mosaics", CVPR, 2001.

[15] W. Zeng, H. Zhang, "Depth Adaptive Video Stitching,", Eigth IEEE/ACIS International Conference on Computer and Information Science, ICIS, 2009.

[16] J. Kopf, M. Uyttendale, O. Deussen, M. Cohen, "Capturing and Viewing Gigapixel Images", SIGGRAPH 2007.