Ellen Meijer

Structure from Motion
as documentation technique for Rock Art

Structure from Motion (SfM) is an user-friendly, low budget way of creating three dimensional image-Based Models from two dimensional photographs. The method has found its way into the documentation of rock art in recent years. In Sweden, the application on rock art has been developed by the Swedish Rock Art Research Archives (SHFA), primarily in Tanums World Heritage Area. The pilot project, commissioned by Länsstyrelse Väst, at Aspeberget was very succesful. It is a non-invasive, objective documentation method that is relatively easy to apply in the field that has the potential to become the standard in rock art documentation in the future.

The name of the technique – Structure from Motion – basically explains its principle: to photograph a fixed object (structure) through movement (motion). The tools needed are a digital SLR camera, a computer and a program to process the photographs. A Structure from Motion (SfM) approach allows the simultaneous computation of both the relative camera projection geometry and a sparse set of 3D points that represents the geometry/structure of the scene in a local coordinate frame, using only corresponding image features occurring in a series of overlapping photographs captured by a camera moving around the scene. In short, the method employs overlapping images acquired from multiple positions, similar to photogrammetry. However it differs from conventional photogrammetry in that the geometry of the scene, the camera positions and the orientation is determined automatically without the need to specify a network of targets (Westoby et al. 2012:301). Moreover, the program allows for images to be unordered, which means that photographs taken with multiple cameras and / or on various occasions can be used. This can be helpful when results from the initial processing of images reveal areas of the model in which further detail is needed (Severa & Goldhahn, 2011:261).

Structure from Motion versus Laser scanning

Prior to the implementation of Structure from Motion, laser scanners have also been used in the documentation of rock art. Although the results were promising (Johansson & Magnusson 2004), the technique was never implemented, due to the high costs and the heavy equipment that is difficult to handle in the field. An alternative to these laser scanners are handheld laser scanners, which are relatively easy to use in the field. As laser scanners uses invisible electromagnetic radiation (most of the time near-infrared) to determine the distance to an object, they are sensitive to direct sunlight. It is therefore advised to cover the surface that needs to be scanned with a party tent with a dark fabric in order to create shadow (Johansson & Magnusson 2002:130). Because the model is created during the scanning process, a powerful computer with external electrical supply is needed in the field as well. The additional equipment makes the technique less suitable for panels that are situated further away from an access point or for one person to operate. Furthermore, the high price of the equipment still continues to place laser scanning and its benefits out of reach of the majority of archaeological projects (Goldhahn & Severa 2011:262).

Figure 1: Documentation of a rock surface with a handheld laser scanner. Photo: Catarina Bertilsson, SHFA

An advantage of handheld laser scanners is the guaranteed depth accuracy and the flexibility in the choice of depth accuracy, during both acquisition and processing of the model. Moreover, the model is built up whilst scanning, making it possible to view the result in the field. A downside is the large amount of small targets that are needed to be able to scan the surface. However, more recent models are able to work without targets.

The advantage of Structure from Motion is that the image collection can be unordered and taken by multiple cameras, even with variations in focal lengths and image resolutions. This makes the technique highly accessible and usable to everyone with a camera and a computer (Goldhahn & Severa 2011:261). Its cost-effectiveness makes it possible to document a rock surface at regular intervals to, for example, monitor the progress of the weathering. Furthermore Structure from Motion does not need targets to create a 3D model. In practice it is, however, advised to make use of 4 - 6 targets, because it will increase the accuracy of the model and allows the extraction of correct metrical information (Plets et al, 2012:147). The limited amount of targets gives a large amount of flexibility, as whole surfaces can be photographed, inclusive ridges and other topographical aspects. In fact, during the pilot project new images were found merely because the adjacent parts of the rock was photographed to get a complete coverage of the surface.

In August 2015 some rocks were documented with both a handheld laser scanner and Structure from Motion. The preliminary results show that Structure from Motion can compete with a handheld laser scanner, provided that all criteria are met (see below). However, in practice, it may not be a question of either / or but rather of both / and. The documentation of a rock surface through Structure from Motion is quick and easy and results in a good image-based model revealing the motifs, the structure of the surface and the topography of the rock. The set of photos taken in the field can be used to create a model of both the whole surface as well as individual motifs. The results of these models can be used to discuss whether it is necessary to document a specific motif or rock surface with a handheld laser scanner (with a guaranteed depth accuracy). This could be desirable for severely weathered and/or very smooth surfaces, either to retrieve as much detail as possible or, in the case of the latter, to try and get more information regarding the pecking techniques used.

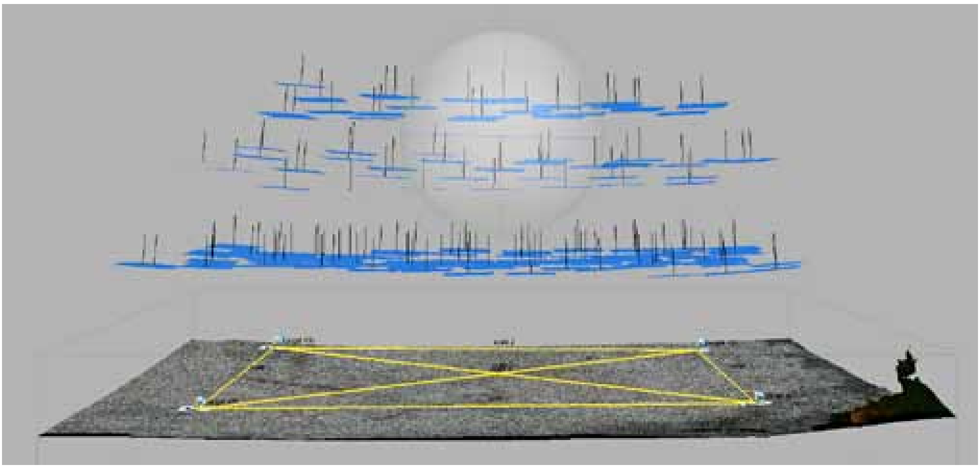
Structure from Motion from photo to a three dimensional Image-Based model

The first step – taking the photographs – is the most important step, as the quality of the photos and the amount of overlap determines the quality of the model. The photographs have to be sharp, well-exposed, evenly distributed over the object and taken at right angles to the object.

The process starts with the alignment of the photographs. During this step the program searches for common points in the photographs and matches them, calculates the camera positions for each picture (Agisoft 2014:V) and the internal camera parameters (focal length, principal point location and distortion) are computed. As this step largely determines the final accuracy of the model, it is useful to visually check the image alignment and the computed projection error directly after the alignment (J. De Reu 1110 et al. 2013: 1111). An advantage of SfM is that known targets and 'artificial' reference points are not necessary in order for the software to evaluate and merge the individual photographs. The information contained within the images themselves (Exif information) provides the location, orientation and geometry of objects through the application of computer vision technologies (Snavely et al 2006, p. 855). The result is a sparse point cloud and the camera positions (Goldhahn & Severa 2011:255).

Although it is not necessary to use artificial reference points, it is advised to use the targets supplied by the program. By defining the exact distance between two reference points in the field, the model is rescaled to an absolute model from which correct metrical information can be extracted. (Plets et al, 2012:147). However the use of targets can be harmful to the rock surface, especially if the surface is weathered, but it is not as tactile as the many target points necessary for laser scanning.

Figure 2: Result of the alignment of the photographs with the various camera positions (blue). Exact dimensions are created with the targets and scale bars (yellow). Photo: SHFA, Ellen Meijer



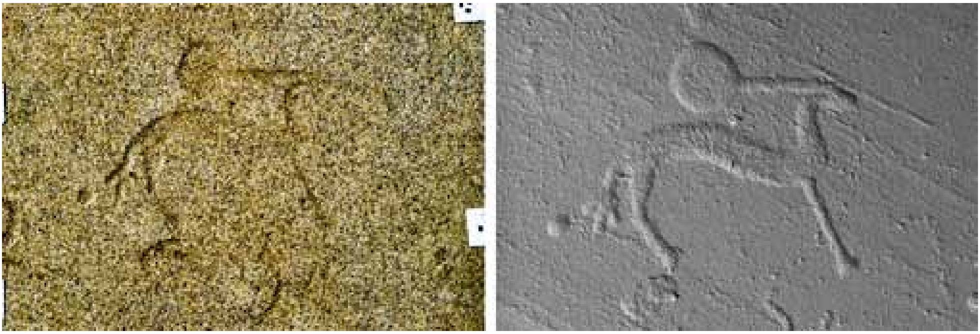


Figure 3: Left the dense cloud. Right: the meshed 3D model. Photo SHFA, Ellen Meijer

The second step in the process is the construction of the dense point cloud. The software now uses the camera positions, orientation and calibration, together with the sparse point cloud and the actual images, to reconstruct depth (distance) maps: for every pixel in the image, the distance between camera and object is computed and a dense point cloud is created (Plets et al 2012:147).

Finally, the dense point cloud is used to calculate a meshed 3D model. If so desired, this model can be texturized based on a selected photograph or a blend of various (selected) photographs. The texturing is not a necessary step in the creation of the model. The textured version, for example, cannot be illuminated in the same way as the meshed model and as such resembles a detailed 2D photograph. However, the textured version shows the coloration of the rock surface inclusive the (micro)vegetation, which can be particularly useful for research of the weathering.

During each step, adjustments can be made to increase the quality of the final model. The meshed model and 3D scene can be exported to different exchangeable formats (Wavefront OBJ, 3DS, Stanford PLY, Acrobat PDF etc.) that can be accessed and visualized in various software packages (e.g. freeware packages like Blender and MeshLab). In addition, orthophotos and 2.5D digital surface models (DSM) can be calculated (Plets et al, 2012).

The application of Structure from Motion within Rock Art

During the pilot project by SHFA in 2014, it became clear that a good 3D model of a rock carving with all its details requires a different approach than, for example, an excavation or landmark, where the difference in depth is considerably larger. The quality of the model is determined by the quality of the photos, the amount of overlap, the number of levels and the camera settings. However, the rock surface itself plays an important part as well. As Structure from Motion extracts the information from the photographs to produce the model, the photos need to be sharp and well exposed. As the point cloud is built up of matching pixels in the photographs, an overlap of 70-80% is advised. This means that an object should be visible in at least two, but preferably three photos. If the aim is to document a complete rock surface, it is advised to also take a set of oblique photographs around the rock.

The light conditions are equally important. For example, hard shadows should be avoided, because 3D reconstruction in the shadowed areas may be poor (Hesse 2014). The results, thus far, show that more detail can be obtained if the photos are taken at the "optimal time of the day", when the sun has the correct angle towards the rock and the images become visible, even to the untrained eye. As with other documentation methods, the best result is achieved on a dry surface, as water will produce a mirror

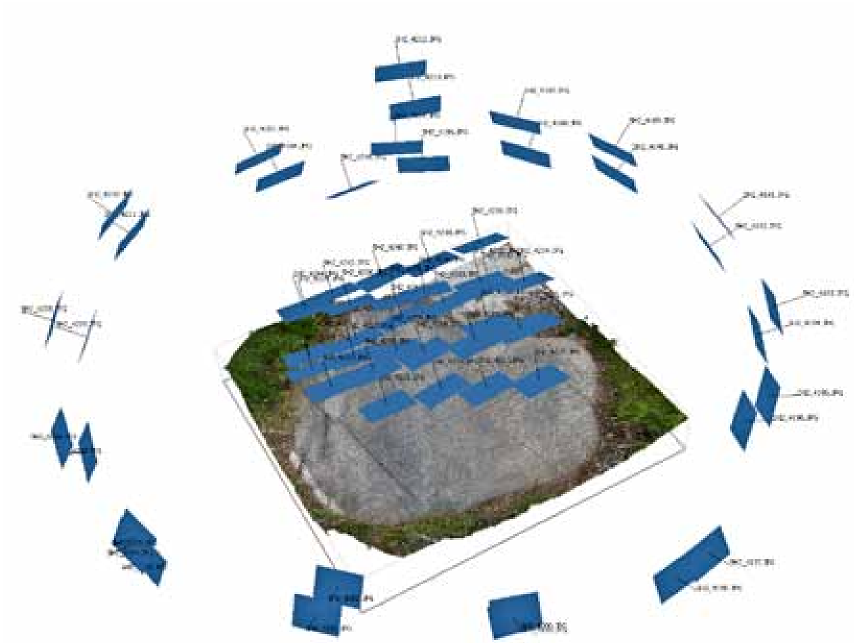


Figure 4: 3D model with camera positions indicated. The panel has been photographed from overhead in 3 levels. A set of oblique photos has also been taken to reproduce both the images and the topography of the panel. Photo SHFA, Ellen Meijer.

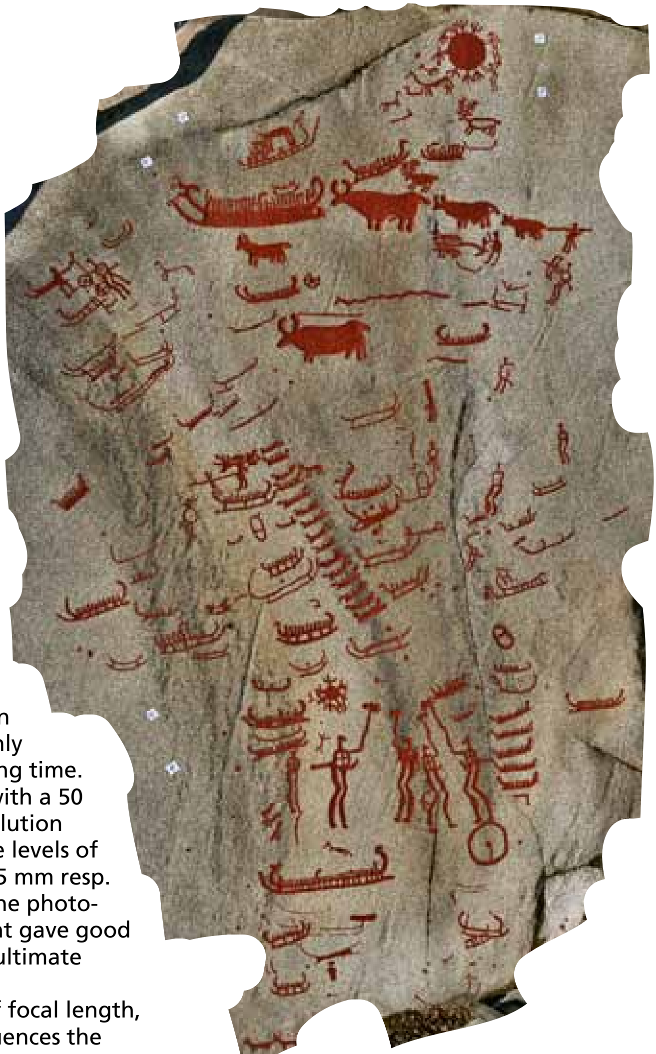
effect, consequently compromising depth reconstruction.

No rock is the same and the exact number of photographs required is dictated on a case-by-case basis. It is tempting to think that a large number of photos will lead to a better model, however, in reality, the difference between the required set of photographs and an excess of such is marginal. More photographs than necessary will only lead to an increase in processing time and a consequent higher demand on the computer's internal memory. It is better to focus on the sharpness of the photographs and the 70-80% overlap. The quality of the model is also influenced by other aspects, such as:

1. The resolution of the photos. As the number of pixels increases, the surface area covered by each pixel decreases, thus increasing the resolution and the sharpness of the object in each image.
2. The focal length of the lens. The longer the focal length of the lens, the higher the magnification and the narrower the angle of view. With other words, the longer the focal length, the closer the object and consequently the more detail.
3. The number of levels. It is advisable to take a set of photos from a higher level to ensure coverage of the whole surface as well as a set of photos from one, or, preferably, two lower levels to obtain as much detail as possible. As with the two points mentioned above, the closer the photo is taken towards the object, the more detail it will give.

4. The condition of the rock surface. A smooth surface will reveal more detail as the contrast between the surface and the structure of the motifs is clearer, even to the naked eye. A weathered surface with shallow carved motifs will require a different approach than a smooth surface with deeply carved motifs.

Figure 5: Aspeberget T12 photographed from three levels. The approx. 1200 photographs generates a 3D model of over 650 Mb. Photo: SHFA 2015



As the application of this technique is still quite new within rock art documentation, a variety of tests were undertaken by SHFA in 2015, mostly to fully comprehend the effect of camera settings, focal length and the number of levels on the quality of the Image-Based model. The four points mentioned above were more or less confirmed by the tests, which were taken on a relatively well preserved part of the Balken rock art panel (Tanum 262). The test results showed that an abundance of photographs, for example an extra level of height, did not result in a more accurate model, and only served to increase the processing time. The best result was obtained with a 50 mm focal length and high resolution photographs, taken from three levels of height. Shorter focal length (35 mm resp. 24 mm), lower resolutions of the photographs or fewer levels of height gave good results as well, but lacked the ultimate sharpness and details.

The optimal combination of focal length, resolution and levels, also influences the amount of photographs needed, which in turn influences the size of the final document and the ability to dissemate the Image-Based model within publications, on websites, or via other media. During the pilot project conducted by SHFA in 2014, processing of the images from some of the large panels at Aspeberget generated Image-Based models of 650 Mb to 1 Gb in size. This, however, can be adjusted by using smart mesh decimation techniques to reduce the size of the models and make them more manageable.

As long as there is a demand for illustrations in books and other publications, alternative documentation methods may still remain in use in the nearby future. A printed version of the Image-Based model is reduced to 2 dimensions, thus losing one third of the information originally gathered (Plets et al. 2012:140). From an aesthetical point of view, 2D representations of 3D models look less attractive than the alternatives, such as rubbing and night photogra-

Figure 6: The acrobat from Aspeberget (Tanum 14) reproduced with the three objective documentation methods: night photography (left), rubbing (center) and SfM (right). Each show a similar amount of information. (Photo: Ellen Meijer, rubbing: Tanums Hällristningsmuseum Underslös, SfM: SHFA, Ellen Meijer)

phy (see fig. 5), especially when they reveal a similar amount of information.

It is often argued that the conventional 2D reproductions, such as rubbings and night photography, do not show the same amount of information regarding the relative depth of the carvings or the relationship between the carvings and natural geological features. However, the details of a reproduction are to a large extend determined by the skills and experiences of the person responsible for the documentation. To summarize, night photography and rubbing can reveal a similar amount of information, but both lack the third dimension. These techniques, however, are more tactile than SfM and Laser scanning. The advantage of an Image-Based model is that it can be rotated and illuminated from various positions, which is necessary in order to study all details. As such, these models of rock art may be regarded as the most objective documentation, because the illumination is not predetermined, in contrast to, for example, a night photograph, where the photographer decided which angle of illumination is the best.

Whilst it appears that the implementation of 3D documentation within rock art research is still in its infancy, a more elaborate discussion is needed (Plets et al. 2012:150). Nevertheless, there is little doubt that the documentation of rock art through Structure from Motion, or similar methods to create 3D models, has the potential to become the standard for the future. It is a non-tactile, objective and simple method of documentation, which is least harmful to the rock, and will allow researchers to simulate a visit to the site from the comfort of their chair at home or at the office. At the same time, we should not forget that the obtained result, as with all other methods of documentation, remains only a representation of the original surface, and that the models created can be more detailed and objective than others. Even the ultimate documentation of a rock surface is a reproduction that will reveal inconclusive and / or interpretative details that require more research. No documentation, regardless of its accuracy, will be able to replace the research carried out on the rock itself, particularly finger-tip inspection and the resultant dialogue with the rock surface and its imagery that this skill generates.

Ellen Meijer
Tanums Hällristningsmuseum Underslös
SvenskHällristningsForskningsArkiv

I wish to extend a special thanks to the Swedish Rock Art Research Archive (www.SHFA.se) who have initiated and developed the Structure from Motion technique within the field of the documentation of rock art in 2013. Without SHFA the investigations mentioned in this article would not have been possible. A special thanks also to James Dodd, for helping me with the language. All models were processed with Agisoft PhotoScan Professional. There are, however, various freeware programs available on the internet, such as Autodesk 123D, Microsoft Photosynth or VisualSFM.

Bibliography

Bertilsson, U. 2015. Examples of application of modern digital techniques and methods Structure from Motion (SfM) and Multi-View Stereo (MVS) for three-dimensional documentation of rock carvings in Tanum creating new opportunities for interpretation and dating. In *XXVI Valcamonica Symposium 2015.*

Bertilsson U., Bertilsson C., Meijer E. 2014 Dokumentation av Världsarvets Hällristningar med 3D teknik – Structure from Motion. Etapp 1. Pilotstudien på Aspeberget 2014, Rapport Svenskt Hällristnings Forsknings Arkiv, 106s, Göteborg.

De Reu, J., Plets, G., Verhoeven, G., De Smedt, P., Bats, M., Cherretté, B., De Maeyer, W., Deconynck, J., Herremans, D., Laloo, P., Van Meirvenne, M., De Clercq, W. 2013. Towards a three-dimensional cost-effective registration of the archaeological heritage. In *Journal of Archaeological Science 40. 1108-1121.*

Dodd, J.A. 2015. New discoveries, new directions. Rock art research on Bornholm 2013 – 2014. With a contribution by Manuel Dueñas. *Adoranten* 2014. 112-122.

Goldhahn, J., Sevara, C. 2011 Image-based modelling of the present past: Building 3D models of archaeological environments from digital photographs. In *Conference Proceedings from Digital Media and its application in Cultural Heritage*. Amman, Jordan, 13-15 March 2011. Amman, CSAAR Press. 251-266.

Hesse, R. Structure from Motion: tips and tricks. https://blog23d.wordpress.com/2014/11/07/structure-from-motion-tips-and-tricks/ 2014.

Johansson, S.-O., Magnusson, J. 2004. Developing new techniques for 3D documentation of rock art. Ulf Bertilsson &Louise Mcdermott (ed.). In *The Valcamonica symposiums 2001 & 2002* – Rapport från Riksantikvarieämbetet 2004:6. Riksantikvarieämbetet/National Heritage board of-Sweden. 125-133.

Micheletti, N., Chandler, J.H., Lane, S.N. 2015. Structure from Motion (SfM) Photogrammetry. In *Geomorphological Techniques*, Chap. 2, Sec. 2.2, British Society for Geomorphology.

Plets, G., Verhoeven, G., Cheremisin, D., Plets, R., Bourgeois, J., Stichelbaut, B., Gheyle, W., De Reu, J. 2012. The deteriorating preservation of the Altai rock art: assessing three-dimensional image-based modelling in rock art research and management. In *Rock Art Research 2012* - Volume 29, Number 2, p. 139-156.

Plets, G., Gheyle, W., Verhoeven, G., De Reu, J., Bourgeois, J., Verhegge, J., Stichelbaut, B. 2012. Three-dimensional recording of archaeological remains in the Altai Mountains. In *Antiquity* 86, p. 884–897.

Rabitz, M. 2013. *Dokumentation af helleristninger – En kritisk analyse og diskussion af dokumentationsmetoderne*. Unpublished Masters thesis, Prehistoric archaeology, Copenhagen University.

Westoby, J., Brasington, J., Glasser, N.F., Hambrey, M.J., Reynolds, J.M. 'Structure-from-Motion' photogrammetry: A low-cost, effective tool for geoscience applications. In *Geomorphology* Vol. 179 (2012) p. 300–314.