IMAGE MEASUREMENT TECHNIQUES IN ROCK GLACIER MODELLING

J. De Matías [a, *], J.J. Guerrero [b], J. Moreno [a], J.J. De Sanjosé [c]

[a] Dpto. de Ing. de Sist. Inf. y Telemáticos, Universidad de Extremadura, Cáceres, Spain - (jmatias, josemore)@unex.es
[b] Dpto. Informática e Ingeniería de Sistemas, Instituto de Investigación en Ingeniería de Aragón, Zaragoza, Spain - jguerrer@unizar.es
[c] Dpto. de Exp. Gráfica, Universidad de Extremadura, Cáceres, Spain - jjblasco@unex.es

Commission VI, WG VI/4

KEY WORDS: Photogrammetry, Computer Vision, Geomorphology, Rock glacier.

ABSTRACT:

Rock glacier is a rare case of study for 3D modelling. We use a stereo feature-based method and multi-view stereo reconstruction algorithms to get different types of cartography. The proposed method has been tested in a real rock glacier example (Veleta in Sierra Nevada, Spain) for producing maps with good results. This rock glacier is of high scientific interest because it is the southernmost active rock glacier in Europe and it has been analyzed every year since 2001. The research on the Veleta rock glacier is devoted to the study of its displacement and cartography through geodetic and photogrammetric techniques.

1. INTRODUCTION

A rock glacier consists of rocks covering a body of water ice. The ice creeps and carries the rocks and sediment along (active rock glaciers). Rock Glaciers are an interesting geomorphologic phenomenon. Many studies try to do the monitoring of this type of structures because allow experts analyze climate change and his effects. Therefore cartography is very necessary in that purpose (Sanjosé 2007).

Using geomatic techniques (geodesy, global positioning system, photogrammetry) it is possible to study the dynamics of geomorphologic structures with high precision. In order to obtain 3D reconstructions in rock glaciers, conventional photogrammetry (metric camera and normal shots) and cheaper methods (semi-metric cameras and convergent photographs) can be used (Sanjosé 2007, Serrano 2006, Sanjosé2001, Corripio2004), but we aim for a more automatic method.

Obtaining detailed maps (cartography) in rock glaciers is difficult, because the use of flights to obtain photogrammetric images in these areas is expensive. Our aim is to develop a quick and cheap method to obtain enough 3D points to create a detailed DTM. Thus, two methodological goals are considered:

- Cartographic representation of the rock glacier, for example at a scale of 1/1000 with a margin of error in the points of less than ± 20 cm. This is the main objective of this study.
- Determination of the glacier dynamics through the time. Currently, we are working in this new methodology.

To fulfill these goals we use computer vision techniques to automatically detect as many points as possible of the natural environment which is specific to our example Veleta rock glacier. The application of computer vision techniques in this field has the following objectives:

- The creation of a Digital Terrain Model (DTM) with enough points in order to obtain a geomorphic map of the Veleta rock glacier.
- Replacing costly photogrammetric flights using planes with less expensive photographs taken from the top of the surrounding mountains, or using balloons or hang-gliders.
- Elimination of extensive office work by a human operator in obtaining 3D points or contour lines by replacing photogrammetric restitution with an automatic method, which also eliminates human errors.

Main idea is use known Computer Vision and Photogrammetry techniques and develope new methods more automatic which allow us produce 3D cartography products of static and dynamic structures. In that work we try to explain our methodology and experiences using these techniques for 3D model production in rock glacier areas.

We propose a methodology to work in this problem. They are some specific type of models with particular properties that make it a special application for multi-view stereo reconstruction algorithms. Rock glaciers areas are complex and shapeless structures with difficult conditions for take images.

Using common computer vision algorithms we can get the calibration and the exterior orientation of a set of cameras in an automatic way using powerful features, but once we get this information we need a technique to generate a multi-view dense reconstruction. There are a lot of techniques to obtain a dense reconstruction from images. Over the last years, a number of high-quality algorithms have been developed. The work by Seitz (Seitz 2005) categorizes existing methods according a set of properties. We have a specific field of application and consequently we search for adaptation of general solution for our problem. In (De Matías 2009) we propose a method for getting dense reconstruction in a rock glacier area. In this work we explain how to use this dense reconstruction to get specific cartography products and how to take advantage using it for geomorphologic studies.

* Corresponding author. This is useful to know for communication with the appropriate person in cases with more than one author.

2. METHOD DESCRIPTION

In this section we present the principal methodology steps:

2.1 Camera processing (pose estimation)

We work with a calibrated camera (focal length, principal point, radial distorion). Therefore if we want to get the complete camera system we need to calculate the exterior orientation (rotation and translation of cameras or camera pose). Using bundle adjustment method (Triggs 2000) with a set of control points, we calculate exterior orientation of cameras. Bundle adjustment is a non-linear method and his purpose is to get an optimization to minimize the reprojection error through a Gradient Descent solution. We calculate an inicialization using DLT (Direct Linear Transform) and use an iterative approximation to minimize reprojection error. In this bundle adjustment process we use a set of control points, that is to say points with known 3D coordinates. Although bundle adjustment could be used to obtain a self-calibration, we only use it for exterior orientation process (pose camera estimation) since we know camera calibration.

2.2 Extraction features

Extraction of point features of each image using the SIFT extractor (Lowe 2004). SIFT descriptors are very interesting because they are invariant to scale and rotation and have good properties with regard to repeatability, distinctiveness, and robustness, and then, SIFT has shown to be very useful for point matching purposes. In this way we obtain a set of point features for each image using SIFT descriptors, which we use for find pair of corresponding points.

SIFT method use DoG (Difference of Gaussian) detector to extract a set of interest point features. The regions extracted with DoG detector are described with a vector of dimension 128 and the descriptor vector is divided by the square root of the sum of the squared components to get illumination invariance. The descriptor is a 3D histogram of gradient location and orientation. SIFT descriptors are very interesting because they are invariant to scale and rotation and have good properties with regard to repeatability, distinctiveness, and robustness, and then, SIFT has shown to be very useful for point matching purposes. For example, we can distinguish the same point in several images even with big scale and orientation changes. In this way we obtain a set of point features for each image of stereo pair using SIFT descriptors, which we use for find pair of corresponding points.

Although there are other descriptors or methods it was demonstrated with different measures that the SIFT descriptors are superior to most others. Several works have compared existing descriptors and concluded that SIFT is a good feature point extractor method in a general use or in photogrammetric applications (Remondino 2006).

Therefore SIFT method provides a set of image locations and descriptors. We will use this information for matching points process. We obtain a list of image 1 descriptors and a list of image 2 descriptors that form a database of keypoints. The best candidate for a matching is his closer keypoint neighbor in database, that is to say, the keypoint with lowest euclidean distance respect the correspondent descriptor. Nevertheless, many of the image points do not have an equivalent matching point. Thus, it is necessary a way to remove points whithout a good matching in database.

A priori, the more successful solution is to establish a threshold for the distance to the closer descriptor and even better to compare the distance between the closer neighbour and second closer neighbour. Using this method we remove 90% of wrong matched points and we loose only 5% of correct ones (Lowe 2004).

2.3 Initial terrain surface

This step requires the camera projection matrices, SIFT matching points and a triangulation process to obtain a 3D model of the geomorphologic structure and surface of the area. In this step the 3D reconstruction of the scene is performed to produce the DTM from the set of projection camera matrices. We know the internal camera parameters and the exterior orientation (pose estimation) from the camera processing step in section 4.4., therefore we have now a calibrated system.

As we know the camera projection matrices, the 3D structure may be recovered by triangulation (Hartley 2004). A simple scheme is proposed using a linear triangulation method to recover the 3D structure. This scheme is based in the fact that the image points have noise and therefore the rays back-projected from the noisy matches in the images are skew in the space. Since the rays do not intersect in general in just one point, the measured points do not satisfy exactly the epipolar constraint. Then, triangulation by back-projecting rays from the matches will fail. This problem is solved if the matches points are corrected by minimizing a cost function which represents errors in the image. Once the points are corrected, a linear triangulation method can be carried out to obtain the 3D structure. This 3D structure can be improved using bundle adjustment process as we do for pose estimation.

At the end of the triangulation we get a cloud of 3D points. The last step is to convert this 3D point cloud in a surface using a Delaunay triangulation.

Using this previous terrain surface of the rock glacier we can estimate the distance between the camera and the rock glacier. The intersection of camera rays with this terrain surface gives us a good initial estimation of depth.

2.4 Dense reconstruction

We use the multi-view stereo reconstruction algorithm proposed in (Goesele 2006) for rock glacier areas. The algorithm consists of two principal steps:

- Reconstructing a depth map centered in each input view. For each view, we select a set of neighboring views against which we correlate it, using a robust window matching. For each pixel, we cover along its back-projected ray inside the depth interval established respect to the scene being reconstructed.
- Merging the different depth maps into a global mesh model.

2.4.1 Deph map generation

Principal algorithm input is a set of views $V = \{V_1, \ldots, V_n\}$. For each reference view $R \in V$, we select a set of k neighboring views $C = \{C_1, \ldots, C_k\} \in V - R$ against which we correlate *R* using a robust window matching.

For each pixel *p* in *R*, we cover along its back-projected ray inside the depth interval established respect to the scene being reconstructed. We calculate the back-projection of pixel *p* in several values of *d* between the minimum and maximum depth with and increment of Δ_{depth}. For each depth value we compute the normalized cross correlation $NCC(R, C_j, d)$

using a $m \times m$ window centered on *p* and the corresponding windows centered on the projections in each of the views C_j ($j \in [1..k]$) with subpixel accuracy. When two views show a similar area of a textured object, we will obtain a high *NCC* score for value of *d*. If, otherwise, there is for example an occlusion (few important for rock glaciers examples), specular highlight, or other componding factor, the *NCC* value obtained will be low for all depths. We will rely on a depth value only if the window in the reference view correlates well with the corresponding window in several views. We define that a depth value *d* is valid if $NCC(R, C_j, d)$ is larger than a threshold *thresh* for at least two views in *C*. The set of views with *NCC* larger than *thresh* for a given depth *d* is denoted as $C_v(d)$.

For a valid depth *d* we compute a correlation value *corr(d)* as the mean of the *NCC* values of all views in $C_v(d)$:

$$corr(d) = \frac{\sum_{C_j \in C_v(d)} NCC(R, C_j, d)}{\|C_v(d)\|}$$

$\|C_v(d)\|$ evaluates the number of elements in $C_v(d)$. For each pixel *p* in *R*, the depth is chosen to be the value of *d* that maximizes *corr(d)*, or none if no valid *d* is found. The algorithm also compute a confidence value *conf(d)* for each recovered depth value as follows:

$$conf(d) = \frac{\sum_{C_j \in C_v(d)} NCC(R, C_j, d) - thresh}{\|C\|(1 - thresh)}$$

The confidence function increases with the number of valid views and is used to inform the merging step.

Another important aspect in our solution is the depth definition. Usually depth is measured as the distance in the direction perpendicular to the image plane. In our solution we consider the depth as the distance along the line joining the camera center and the 3D point which we want to reconstruct. In this way it is easier to define the prior interval where our algorithm searchs the final depth.

One problem we found using this dense reconstruction method is the estimation of an interval of initial depth for searching the final depth. In this way we have a good initial depth value where start searching the correct depth. The reconstruction method gives us a way to complete the initial terrain surface and to get a dense 3D reconstruction.

2.4.2 Depth map fusion

The previous step produces a set of incomplete depth maps with confidence values. In the following step, we merge them into a single surface mesh representation. To do that merging, the freely available implementation of the volumetric method of Curless and Levoy (Curless, 1996; Vrippack) is used. This approach was originally developed for merging laser range scans. It converts each depth map into a weighted signed distance volume, takes a sum of these volumes, and extracts a surface at the zero level set.

This merging approach has a number of nice properties that make it particularly appropiate for our algorithm, in particular robustness in the presence of outliers and representation of directional uncertainty. The merging algorithm starts by reconstructing a triangle mesh for each view and downweighting points near depth discontinuities and points seen at grazing angles. These meshes are then scan-converted using per-vertex weights into a volume merging. Outliers consisting of one or two samples are filtered out automatically, because they cannot form triangles in the first phase of the algorithm. Larger handfuls of outliers will be reconstructed as small disconnected surfaces; these surfaces will have low weight, since all the points are near depth discontinuities and are probably not substantiated by other views. They can be eliminated in a postprocessing step by removing low confidence geometry or by extracting the largest connected component. In addition, the approach has been shown to be least squares optimal under certain conditions, particulary assuming uncertainty distributed along sensor lines of sight which by constructions applies to the depth maps from previous step.

3. RESULTS AND DISCUSSION

The geomorphologic structure observed in this study is the Veleta rock glacier (Sierra Nevada, Granada, Spain). The Veleta rock glacier is located in "Sierra Nevada" (37º N - 3º W) in a cirque at the bottom of the mount "Veleta" (3,398 m. a.s.l.). Sierra Nevada is a National Park at the south of Spain. This mountain is very vulnerable to gelifraction which turns out in much more rock material on the Veleta rock glacier. The Veleta rock glacier forms an open figure "╗". Its source is attached to the wall of the "Corral" and continues in a northern direction and later extends to the west direction. It has an average slope of 20º, with the front at 3,090 m. a.s.l. and the root at 3,175 m. a.s.l. It is a tongue-shaped rock glacier 35 m wide and 109 m long (Gómez 2004). This rock glacier is of high scientific interest because it is the southernmost active rock glacier in Europe and it has been analyzed every year since 2001. We test the proposed methodology in Veleta rock glacier using 2007 and 2009 campaign images. We use a digital camera (Canon EOS 5D) and a total station (Topcon GTS-502E) to locate control points that are being supervised in the rock glacier area. The baselines between cameras are in an interval of 10 m and 50 m and distance to rock glacier is about 350 m.

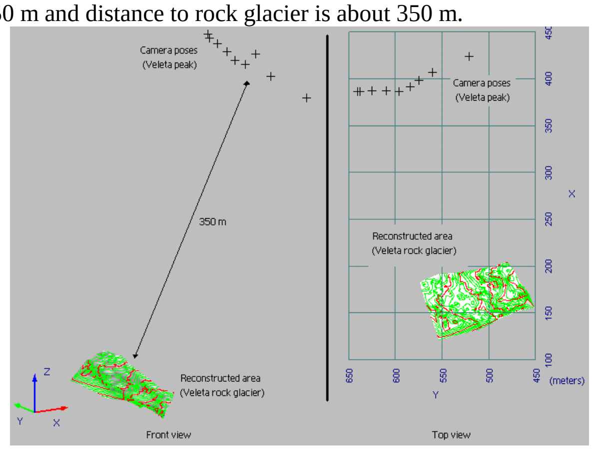

Figure 1. Camera situation respect the reconstructed area

We use photographs of 2007 and 2009 and 9 control points that are being supervised in the rock glacier area in 2007 and in 2009.

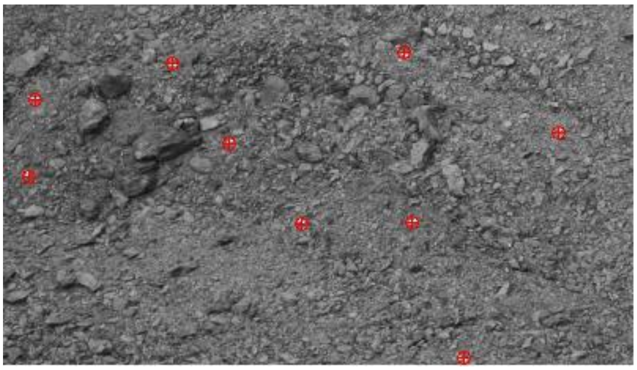

Figure 2. Control points spread around rock glacier.

3.1 Calibrated camera system

We work with a calibrated camera (focal length, principal point, radial distorion) and we only need to calculate the exterior orientation (rotation and translation of cameras). Using bundle adjustment method (Triggs 2000) and the 9 control points (Figure 2) we calculate exterior orientation. In Table 1 the reprojection error (in the control points E1, E2, ..., E9) obtained computing the cameras (P1, P2, ..., P9) in 2007 are shown.

	Average (in pixels) respect points {E1, …, E9}
P1	0.687
P2	0.434
P3	0.535
P4	0.382
P5	0.529
P6	0.403
P7	0.449
P8	0.425
P9	0.631
Average all cameras (in pixels)	0.497

Table 1. Camera adjustment reprojection error (pixels)

3.2 Feature-based method (Automatic initial terrain surface)

Althought the calculation of the initial terrain surface it is not the aim of this work it can be interesting know how we get an acceptable surface if we want use it to calculate initial depth estimation.

In this work we implement a feature-based method using SIFT descriptors (Lowe 2004). Using SIFT we obtain a set of features which will be the basis to get 3D points. SIFT is used in computer vision applications and even in photogrammetry (Remondino 2006). It allows us find intereting points in a image which we can match with points in other image (De Matías 2007) This method can work with changes in rotation, scale or point of view with very good performance.

Once we obtain these matching points we have to triangulate them (using our known camera system) (Hartley 2004) to get a set of 3D point (≈8000 points). The last step is to convert this 3D point cloud in a surface using a Delaunay triangulation. The result is a surface which we can use for initial depht estimation as we explain in Method description section.

3.3 Input parameters value

Once we know the orientation of the cameras (interior and exterior orientation) and before use the dense reconstruction algorithm we have to choose several extra input parameters:

- Δ_{depth} : depth increment. It is important because it determines the resolution in depth searching. We have worked with values between 0.1 and 1 meter.
- *thresh*: It is the threshold that we use to determine when a NCC value is valid or not. Values between 0.6 and 0.8 are used.
- *k*: number of neighbors. We use 3 or 4 neighbor images, as it is adviced in Goesele work (Goesele 2006).
- Initial depth estimation. A depth interval for searching the final depth is required by the method. As explained above, we use a first depth estimation using a previous terrain surface. In our case we set a range of 20 meters around the initial depth computed. Without the automatic initial terrain surface improvement we should use a depth interval between 220 meters and 360 meters to obtain comparable results but with ≈ 7 times more computational cost (Veleta rock glacier example).

3.4 Results with rock glaciers images

After the use of algorithm (with Veleta rock glacier images) we have obtained one depth map per photograph with ≈200000 points per map. Normally it is difficult calculate quantitatively the error of 3D model because we have not a comparable “real” model. In our example we test the algorithm with a set of additional terrain points (TP1, TP2, ..., TP10) measured with Total Station. Their 3D location is obtained with an accuracy about ± 3-4 cm. It is important to use points that were not used in any step of the method (points of Figure 2 were used in exterior orientation of cameras). In these terrain points we calculate the reprojection error obtained by the dense reconstruction algorithm. In Table 2 we can see the reprojection error obtained during the calculation of a concrete deph map. In this case we show the reprojection error obtained in the reference image (Img1) and in the 3 neighbors images (Img2, Img3, Img4). As Table 2 shows, we get little errors (with a maximum of about 1 pixel).

	Img1	Img2	Img3	Img4	Average (pixels)
TP1	0.05	0.06	0.28	0.58	
TP2	0.05	0.03	0.10	0.22	
TP3	0.06	0.14	0.63	1.20	
TP4	0.01	0.01	0.03	0.07	
TP5	0.05	0.14	0.59	1.14	
TP6	0.08	0.26	1.08	1.20	

TP7	0.07	0.18	0.77	1.17	
TP8	0.08	0.09	0.42	0.85	
TP9	0.10	0.05	0.24	0.50	
TP10	0.02	0.01	0.05	0.10	
Average	0.057	0.097	0.419	0.703	0.319

Table 2. Terrain points reprojection error (pixels)

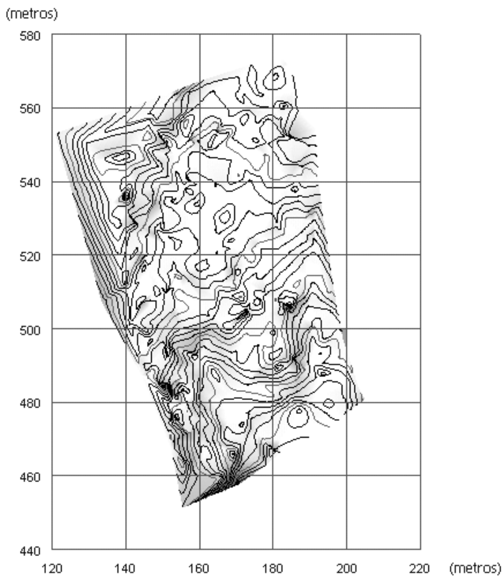

Figure 3. 2D map of rock glacier

Also it is important the results obtained comparing 2007 and 2009 reconstruction. In Figure. 4 we show a comparison of surfaces to obtain loss volume information. Blue line represents rock glacier contour, red areas represent areas with a big loss of volume and blue areas represent areas will a little loss of volume. After comparing 2007 surface and 2009 surface we obtain a loss of volume of 2178 m^3.

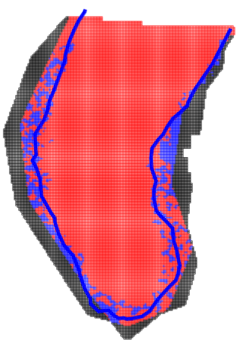

Figure 4. Loss of volume between 2007 and 2009

4. CONCLUSIONS

We propose an automatic technique for rock glacier reconstruction using digital images. A significant amount of work in feature matching techniques has been reported in computer vision in recent years improving old techniques. Using this kind of techniques the post processing cost and time needed to analyze image information of geomorphologic structures can be dramatically reduced avoiding the arduous task of manual photogrammetric restitution.

3D reconstruction of rock glacier surfaces is a specific case of 3D modelling with particular properties which motivated us to get a particular 3D reconstruction methodology. Using the implemented algorithms we can obtain a rock glacier reconstruction with a big amount of points with enough accuracy, merging the information of several images. Main objective is obtain cartography products interesting for geomorphologic studies. We present cartography products obtained in Veleta rock glacier and show how these products can be useful in rock glacier monitoring tasks.

In order to show the validity of the methodology experiments of the rock glacier have been performed and the most representative are shown. From the results given we can conclude that the set of matches obtained is very good. Future work is focussed toward on extending this method to images taken at different periods of time to study the evolution of the rock glacier. The objective is to have a cheap and fast tool for studying the evolution of the rock glacier along the time, which is our current research goal.

ACKNOWLEDGEMENTS

The authors are grateful for the collaboration on the geodetic data collection at the Veleta rock glacier to Antonio Gómez Ortiz, Ferran Salvador Franch (University of Barcelona) and Alan D.J. Atkinson Gordo (University of Extremadura). The financial support for this research is from a Nacional Parks Project (Ministry of Natural Environment of Spain, Ref. 018/2007).

REFERENCES

Corripio, J.G.; 2004. Snow surface albedo estimation using terrestrial photography, *International Journal of Remote Sensing*, Vol. 25 (24), pp. 5705-5729.

Curless, B., Levoy, M.,.1996. A volumetric method for building complex models from range images. *In SIGGRAPH*, pp 303-312.

De Matías, J., Moreno, J, Guerrero, J.J., 2007. . Automatización de trabajos fotogramétricos mediante técnicas de Visión Artificial. *Congreso Internacional de Matemáticas en la Arquitectura y la Ingeniería.*

De Matías, J., Moreno, J, Guerrero, J.J., 2009 A multi-view dense reconstruction for rock glacier modelling. *15th International Conterence on Virtual Systems and Multimedia*, Viena Austria

Goesele, M., Curless, B., Seitz, S., 2006. Multi-view stereo revisited. *In CVPR*

Gómez, A.; Schulte, L.; Salvador, F.; Palacios, D.; Sanjosé, J.J.; Atkinson, A.D.J., 2004. Deglaciación reciente de Sierra Nevada. Repercusiones morfogenéticas, nuevos datos y perspectivas de estudio futuro. *Cuadernos de investigación geográfica de la Universidad de la Rioja*, Logroño, Spain; Vol. 30, pp. 147-168.

Hartley, R.; Zisserman, A., 2004. *Multiple View Geometry in Computer Vision*. Cambridge University Press.

Lowe, D., 2004. Distinctive Image Features from Scale-Invariant Features. *International Journal of Computer Vision*, Vol. 60(2), pp. 91-110.

Remodino, F., 2006. Detectors and descriptors for photogrammetric applications. *Photogrammetry and Computer Vision*, Bonn, Germany.

Sanjosé, J.J.; Atkinson, A.D.J.; Salvador, F.; Gómez, A., 2007. Application of geomatic techniques to monitoring of the dynamics and to mapping of the Veleta rock glacier (Sierra Nevada, Spain). *Zeitschift für Geomorphologie*, Vol. 51 (2), pp. 79-89.

Sanjosé, J.J.; Lerma, J.L., 2001. La fotogrametría digital: Una herramienta idónea para el cartografiado y modelado de zonas de alta montaña. *Congreso de periglaciarismo en montaña*, Potes, Spain, pp. 185-203.

Seitz, S., 2006. Multi-view stereo evaluation web page. http://vision.middlebury.edu/mview/.

Serrano, E.; Sanjosé, J.J.; Agudo, C., 2006. Rock glacier dynamics in a marginal periglacial high mountain environment: Flow, movement (1991-2000) and structure of the Argualas rock glacier, the Pyrenees. *Geomorphology,* Vol. 74, pp. 285-296.

Triggs, B.; McLauchlan, P.; Hartley, R.; Fitzgibbon, A., 2000. Bundle adjustment – A modern synthesis. *In Proc. Int. Workshop on Vision Algorithms*, pp 298-372.

Vrippack: Volumetric range image processing package.Available at http://grail.cs.washington.edu/