Aerial video geo-registration using terrain models from dense and coherent stereo matching

Susana Ruano, Guillermo Gallego, Carlos Cuevas and Narciso García

Grupo de Tratamiento de Imágenes, E.T.S.I. Telecomunicación,
Universidad Politécnica de Madrid, Madrid, Spain.

ABSTRACT

In the context of aerial imagery, one of the first steps toward a coherent processing of the information contained in multiple images is geo-registration, which consists in assigning geographic 3D coordinates to the pixels of the image. This enables accurate alignment and geo-positioning of multiple images, detection of moving objects and fusion of data acquired from multiple sensors. To solve this problem there are different approaches that require, in addition to a precise characterization of the camera sensor, high resolution referenced images or terrain elevation models, which are usually not publicly available or out of date. Building upon the idea of developing technology that does not need a reference terrain elevation model, we propose a geo-registration technique that applies variational methods to obtain a dense and coherent surface elevation model that is used to replace the reference model. The surface elevation model is built by interpolation of scattered 3D points, which are obtained in a two-step process following a classical stereo pipeline: first, coherent disparity maps between image pairs of a video sequence are estimated and then image point correspondences are back-projected. The proposed variational method enforces continuity of the disparity map not only along epipolar lines (as done by previous geo-registration techniques) but also across them, in the full 2D image domain. In the experiments, aerial images from synthetic video sequences have been used to validate the proposed technique.

Keywords: Geo-registration, terrain elevation model, variational method, dense 3D reconstruction, stereo matching, disparity map, surface interpolation, aerial imaging, multigrid method.

1. INTRODUCTION

In recent years, UAVs are increasingly being used in different domains, both for civil and military applications.[1] For these remotely piloted systems, electro-optical (EO) sensors are essential because they allow vehicle operability and enable the development of applications that build upon aerial imagery, such as surveillance, reconnaissance and remote sensing. One critical step towards processing aerial imagery is geo-registration,[2] which involves the assignment of 3D world coordinates to the pixels of an image (depending on the author, this may be called geo-positioning[3]). Geo-registration is a well known problem that requires precise measurements to achieve accurate results. Typically, these measurements will be given by on-board systems (e.g. Global Positioning Systems - GPS) or by previously geo-registered data. However, on-board positioning systems may not be reliable (due to the lack of accuracy or due to temporal inoperability), and so the availability of reference geo-registered data is important. The data commonly used as a reference are digital elevation models (DEM), which are raster representations of the real-world terrain surface.[4]

The use of geo-registered video data in a UAV can improve the operator's situational awareness by overlaying synthetic spatial information on the video received from the UAV, thus avoiding the burden of fusing related information in mission planning and mission execution displayed in separate screens.[5] A broad classification of geo-registration solutions can be done according to the dimension of the geo-referenced data: 2D if the system uses a collection of geo-referenced images,[3,6] or 3D if the system has an explicit terrain elevation map (e.g., DEM). In the latter case, some approaches use textured models,[7] while others do not.[2] In addition, in some solutions, geo-registration may be aided by inertial navigation systems (INS) and GPS.[8]

In,[9] a pipeline that allows geo-registration without a previous DEM, a low resolution DEM or an outdated one is presented. Following this idea, the main purpose of our work is the generation of a dense surface from

E-mail: {srs,ggb,ccr,narciso}@gti.ssr.upm.es

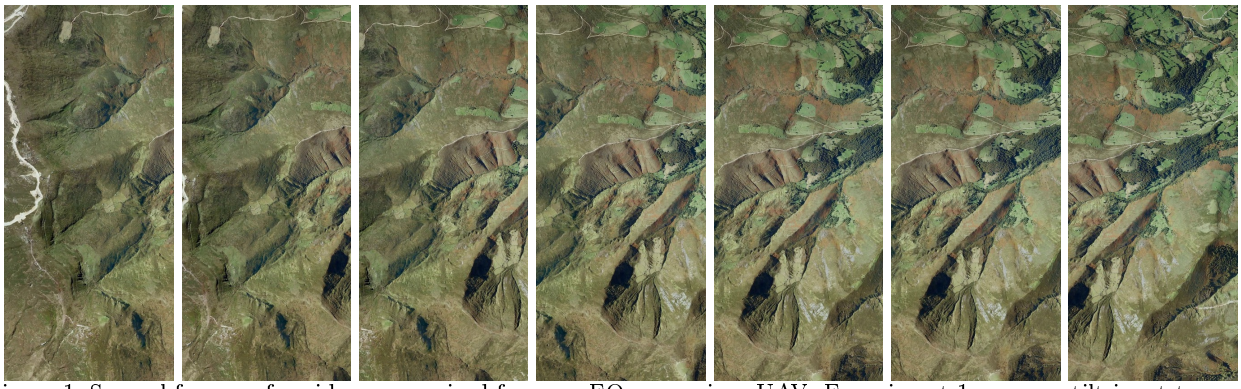
Figure 1. Several frames of a video as acquired from an EO sensor in a UAV. Experiment 1: camera tilt is set to zero.

the video sequence to substitute the DEM. The generation of a 3D surface from multiple images when the extrinsic camera parameters are unknown is a central problem in computer vision known as structure from motion (SFM).[10] Traditional strategies provide a sparse 3D reconstruction which is not enough to replace a DEM, thus dense reconstruction methods are considered. Our technique to obtain such dense reconstructions is based on the estimation of dense and coherent disparity maps between image pairs of the video sequence using multi-resolution variational methods that enforce continuity of the maps in the full 2D image domain. This produces a dense and coherent surface elevation model that may be used as a reference model in order to avoid the use of previous or outdated DEMs.

The rest of the paper is organized as follows: Section 2 describes how to generate the terrain model and mentions how to use this terrain surface to do geo-registration. The terrain generation methods are tested and their results are analyzed in Section 3. Finally, conclusions and future work are drawn in Section 4.

2. TERRAIN MODELS FROM IMAGES

In this section we present an automated solution to the problem of motion imagery geo-registration. The method is based on the scheme,[9] which aims at reducing the need for an input DEM to geo-register images. Thus it is intended to be used in case such a DEM is not available (it may not exist or is out of date) or it is of low resolution. Instead, the terrain elevation model is built from aerial images acquired from a UAV.

The method consists of two main steps: (1) building or refinement of the terrain elevation model, and (2) geo-registration of images using the computed terrain elevation model as reference. This work focuses on using structure from motion (SFM) techniques to implement the first step. The registration of new images is done according to the method in.[2]

2.1 Building the terrain model

2.1.1 Multi-view stereo processing overview

Let us show how to build accurate and robust terrain elevation models from aerial imagery, a process known as three-dimensional (3D) reconstruction in computer vision. To this end, we present a stereo processing pipeline in Algorithm 1. The principles behind stereo processing pipelines and their building blocks are presented in.[10–12]

Consider the scenario where a UAV is flying over a terrain of interest. The first step of the processing pipeline consists of acquiring images (see, for example, Fig. 1) using an electro-optical (EO) sensor, looking downward or in the direction closer to a forward looking camera (FLC). Next, point correspondences are established across images. To this end, repetitive and distinctive features are detected and matched in the images. There are multiple available detectors and descriptors[13] (SIFT,[14] SURF,[15] KAZE,[16] etc.). Such detectors have proven to be very effective for wide baseline matching like the case considered here.

The 2D point correspondences are the input of multi-view Structure from Motion[10] (SFM) algorithms, which reconstruct both the 3D location of the object points as well as the pose of the cameras in the scene based on

Algorithm 1 Multi-view stereo processing pipeline to build dense terrain elevation models.

1. Acquire images from a UAV.
2. Establish feature (e.g. point) correspondences across images.
3. Estimate the (sparse) 3D scene structure and the relative camera poses from 2D image point correspondences using Structure from Motion (SFM) techniques.
4. Generate dense depth maps of the scene using the variational disparity method between image pairs.
5. Combine multiple depth maps into a single surface by triangulation (i.e., back-projection) and averaging.
6. Geo-reference the terrain elevation model using either known position and orientation of the UAV's camera or known landmarks.

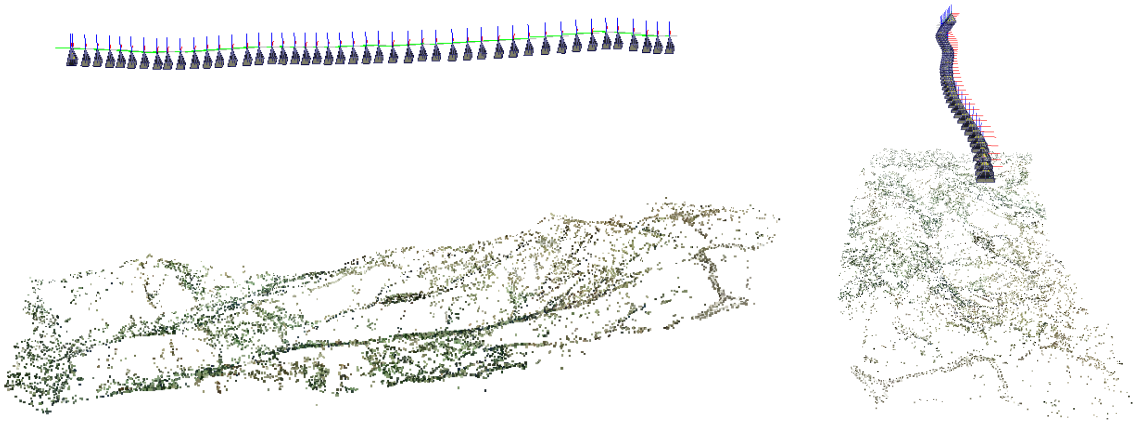

Figure 2. Sparse reconstruction of the viewed terrain in Experiment 1 (Sect. 3), including the trajectory (camera poses) of the UAV during the flight.

the observed point projections in the images. This is accomplished in an optimization framework with objective function given by the reprojection error between the predicted image points and the observed ones.

In most practical cases where there is access to the EO sensor, it is possible to assume that the internal camera parameters (focal length, principal point, etc.) are known, i.e., cameras are calibrated. This constitutes valuable information because it significantly constrains the 3D reconstruction problem by reducing the number of unknowns and avoiding camera self-calibration,[17,18] a very sensitive step. This is our case, and so we use SFM algorithms such as[19] or[20] that can exploit the calibrated camera constraint. First, two images with sufficient parallax are selected to initialize the reconstruction. The essential matrix[10] that encodes the relative Euclidean motion (rotation and translation) between both cameras is estimated as well as a set of 3D points that lie in front of the cameras. Then, the remaining images add new camera poses and object points to the existing reconstruction. Camera poses are estimated from object and image point correspondences (3D-to-2D matches) by solving the so called resection[10] or Perspective-n-Point (PnP) problem. New object points are added to the reconstruction by triangulation[10] of matched image points of a previously added camera pose. A global optimization of the camera poses and object points is usually performed to improve the fit of the reconstruction, a step known as bundle adjustment.[10] Robustness of the algorithm is achieved in multiple stages by outlier rejection using RANSAC.[21]

The output of the SFM step consists of a sparse 3D reconstruction of the scene (structure and camera poses), as shown in Fig. 2. Such a sparse reconstruction does not sample the objects in the scene (e.g., the terrain of interest) with enough density of points to provide accurate surface models. Nevertheless it is a useful step for several reasons: it yields accurate information by processing small amounts of data (sparse features) in reasonable

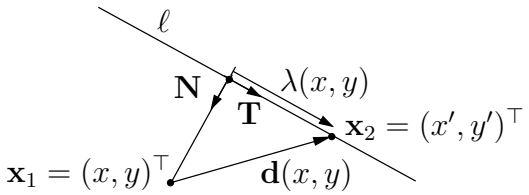

Figure 3. Disparity parameterization. At every point $\mathbf{x}_1$ in the reference image, the disparity $\mathbf{d}$ with respect to its corresponding point $\mathbf{x}_2$ in another image can be parameterized by the tangential disparity $\lambda(\mathbf{x}_1)$ along the epipolar line ℓ, i.e., $\mathbf{d} \equiv \mathbf{d}(\lambda(\mathbf{x}_1))$.

computational times, it allows the estimation of the camera poses in case they are unknown and it may be used as the starting point of a "densification" algorithm, as we show next.

2.1.2 Dense stereo matching via variational methods

Once the camera parameters are known, the stereo reconstruction problem is separable in two subproblems:[22] stereo matching (establishment of dense point correspondences between images) and depth recovery (back-projection of corresponding image points by triangulation, i.e., intersection of optical rays from the cameras). The first subproblem is significantly more difficult than the second one, and to solve it under the hypothesis of a Lambertian scene (where corresponding points in images have equal color levels) we use the variational disparity method between image pairs proposed by[23] and modified by[24] to reconstruct ocean waves. Ultimately, this variational method estimates a dense depth map of the scene with respect to both cameras indirectly via estimating the disparity between images. Such depth maps will be translated to terrain elevation displacements with respect to a reference plane.

The disparity map is a vector field that establishes correspondences between two images. A point $\mathbf{x}_1$ in image I_1 is mapped to point $\mathbf{x}_2 = \mathbf{x}_1 + \mathbf{d}$ in image I_2. The disparity $\mathbf{d}$ has two components: the displacements in both horizontal and vertical directions of the image. However, assuming there is no lens distortion (it has been corrected because we assumed calibrated cameras) and according to the epipolar constraint,[10,11] the disparity can be parameterized by a single displacement: the signed distance along the epipolar line, also called the tangential disparity λ, thus obtaining $\mathbf{d}(\lambda)$. Specifically, the disparity at $\mathbf{x}_1$ can be decomposed in the orthonormal frame $\{\mathbf{T}, \mathbf{N}\}$ adapted to the epipolar line ℓ corresponding to $\mathbf{x}_1$: $\mathbf{d} = \lambda\mathbf{T} + \gamma\mathbf{N}$, where γ is the distance of $\mathbf{x}_1$ to the closest point on its epipolar line ℓ, and so it solely depends on $\mathbf{x}_1$ and the camera parameters, but not on $\mathbf{x}_2$; only λ depends on the location of $\mathbf{x}_2$ along ℓ. This is illustrated in Fig. 3.

In this setting, the disparity map that solves the matching problem is obtained as the minimizer of the functional

$$E = E_{\text{data}} + \alpha E_{\text{smooth}}, \tag{1}$$

which consists of a weighted sum of a data fidelity term and a smoothness prior ($\alpha > 0$). The data fidelity term measures the photometric consistency between stereo images caused by a candidate disparity map, $E_{\text{data}} = \int_\Omega \frac{1}{2}\big(I_1(\mathbf{x}_1) - I_2(\mathbf{x}_2)\big)^2 d\mathbf{x}_1$, where $\mathbf{x}_1 \in \Omega$ lies in the reference image (origin for the disparity map) and $\mathbf{x}_1 \leftrightarrow \mathbf{x}_2(\lambda)$ are corresponding points in images 1 and 2, respectively, with observed intensities I_1 and I_2. This data fidelity cost is not symmetric with respect to the role of each image, but it can be easily symmetrized. The regularizer $E_{\text{smooth}}(\lambda) = \int_\Omega \frac{1}{2}\|\nabla\lambda\|^2 d\mathbf{x}_1$, $\nabla\lambda$ being the gradient of $\lambda(\mathbf{x}_1)$, enforces coherence (continuity) of the disparity map. The epipolar constraint allows to express (1) so that it depends on a single 2-D function, $E(\lambda)$.

In other stereo approaches, e.g. those that match points in corresponding epipolar lines on rectified images, continuity of the solution is only enforced at most along (1D) epipolar lines, but not across them. In our approach, however, continuity is enforced in the full 2D domain of λ within the image and it does not require rectified images. In addition, the weight $\alpha > 0$ allows to control the amount of smoothness of the solution disparity map, which is directly transferred to the smoothness of the terrain model because the disparity map contains depth information of the scene with respect to the cameras.

Cost (1) is minimized by gradient descent according to the necessary optimality conditions (Euler-Lagrange (EL) equations) of (1). These yield a non-linear elliptic partial differential equation (PDE) in the tangential

Figure 4. Variational disparity method. Predicted images by transferring intensities according to the correspondence given by the disparity map. Left: original image I_1 (outside the centered rectangle Ω, with some intensity scaling for visualization purposes) and predicted image $\hat{I}_1(\mathbf{x}_1) = I_2(\mathbf{x}_1 + \mathbf{d}(\lambda))$ (inside Ω). Center: Tangential disparity $\lambda(\mathbf{x}_1)$, pseudo-colored in grayscale expanding the range of λ, in this example $\lambda \in [-48.54, -41.11]$ pixels. Grid size (Ω): 513×257 pixels. Right: predicted image $\hat{I}_2(\mathbf{x}_2) = I_1(\mathbf{x}_2 - \mathbf{d}(\lambda))$ (matched region) and original image I_2 (outside).

disparity λ, $\alpha\Delta\lambda + (I_1(\mathbf{x}_1) - I_2(\mathbf{x}_2))\frac{\partial I_2(\mathbf{x}_2)}{\partial \lambda} = 0$, with homogeneous Neumann boundary conditions, where the symbol $\Delta\cdot$ stands for the Laplacian operator. The PDE is discretized on a rectangular 2D grid using finite differences and numerically solved using iterative multigrid methods,[25] which are among the most efficient numerical tools for boundary value problems.

In the same spirit of the hierarchical motion estimation framework,[9] multigrid methods are well founded multi-resolution tools. This increases robustness and significantly reduces the computational cost of establishing dense matches. Specifically, the full multigrid (FMG) method combines multi-resolution with a coarse-to-fine initialization, which is a fast and sensible approach to improve convergence of the iterative method and avoid local minima of the cost functional.

The numerical solver can be initialized in several ways. One option consists of using available software such as PMVS[26] to provide a dense point cloud representing the terrain model and project it to compute an initial estimate of λ. This is, however, overkilling: λ does not need to be initialized by such a detailed map since it is a slowly varying signal (in our use cases, see Fig. 4, center). Instead, an interpolation of the disparities given by the sparse reconstruction often suffices, specially if it is used to initialize the solver at the coarsest level of FMG. A third initialization option (used in[23]) consists of using a block based correlation matching algorithm. Finally, if the terrain of interest is flat, then λ can be reasonably initialized by the disparity corresponding to the mean plane through the scene.

Figure 4 shows the output of the variational disparity method overlaid on the original images (with some intensity scaling for visualization purposes). It can be observed, e.g., at the boundary of the highlighted matched region, that the predicted images via the estimated disparity map are a good fit to the original images.

2.1.3 Back-projection of dense disparity maps and surface generation

Once the variational disparity method has densely matched points between two images (Sect. 2.1.2), we recover the depth of the corresponding terrain points with respect to the cameras by intersecting the optical rays through the image points (i.e., triangulation[10]), according to the camera parameters obtained in the sparse reconstruction. This operation gives a dense cloud of 3D points for every image pair densely matched. Figure 5 shows the combined point cloud from a subset of the image pairs matched in a video sequence. Clearly, overlapping between point clouds is prone to exist and they must be combined to produce a consistent point cloud and/or meshed terrain model.

Next, 3D points are expressed with respect to a plane through the scene.[27] If this plane coincides with a horizontal plane, whose normal direction may be obtained by means of an inertial measurement unit (IMU) in the UAV, then this operation produces elevation information so that 3D points are expressed as $z_i = (x_i, y_i)^\top$. Interpolation methods are used to fit a surface $z^k(x, y)$ through the points corresponding to each (k-th) disparity map. Then, surfaces are merged by using statistical estimators (e.g. mean, median) on the functions z^k over a common grid domain. Figure 8 shows the merged surface from all disparity maps in a video sequence.

An alternative approach consists of using Poisson surface reconstruction[28] or similar algorithms to fit a surface model (e.g. polygonal mesh) to the combined point cloud from all disparity maps, without referring the points

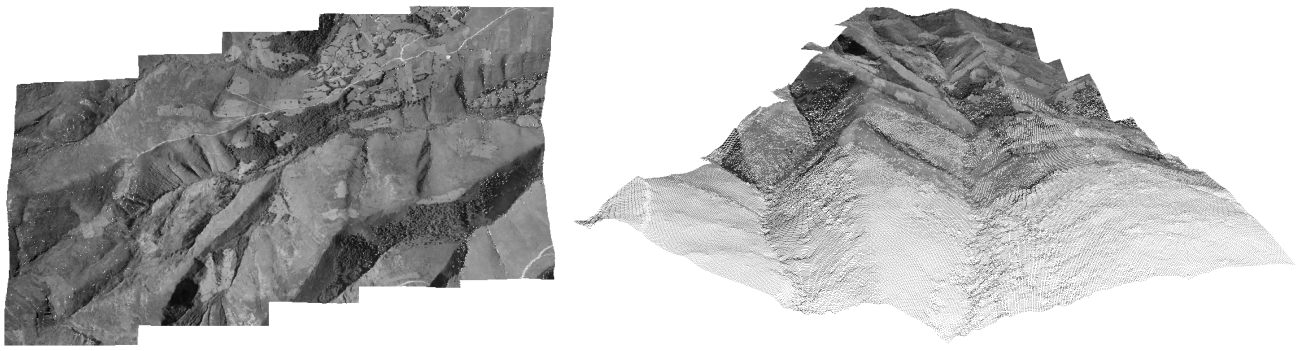

Figure 5. Point cloud obtained by triangulation of densely matched points with the variational disparity method. Example with points from five disparity maps of size 513×257, adding up to ≈ 0.66 Mpoints. Left: zenithal view. Right: close-up view. Individual points are more distinguishable as they are closer to the selected viewpoint.

to a plane. The fitted surface model needs to be constrained to be in the form of a graph $z(x, y)$ with respect to the horizontal plane so that it represents an elevation terrain model (e.g., DEM).

Finally, the resulting terrain model is geo-referenced (step 6 in Algorithm 1). This can be accomplished using additional information from the UAV (IMU, GPS, etc.) or from the scene (e.g. known geographic coordinates of landmarks). This is just a Euclidean change of coordinates, possibly including a scaling, between reference frames: the geographic world and the world in which the 3D reconstruction (cameras, terrain points) is given.

2.2 Geo-registration with the terrain model

Once the terrain elevation model has been formed, it can be used as a reference to find the camera pose (location and orientation) of new images of the scene acquired, for example, by another UAV flying over the same terrain. According to the method in,[2] the surface model, illumination conditions and candidate camera parameters of the new images are combined in a computer graphics pipeline to generate predicted images. Both, predicted and observed images are compared in an optimization framework that updates the camera parameters to achieve an optimal fit, i.e., to geo-register the new image with respect to the reference surface. A detailed discussion of this step can be found in.[2,5]

3. EXPERIMENTS

To validate the proposed method, we test it on simulated aerial video sequences obtained with Google Earth.[29] Two experiments are carried out: in the first one, the UAV camera is looking downward (with zero tilt and heading), whereas in the second one the camera is closer to the FLC setting, with a tilt of 45 degrees. The terrain under study corresponds to a mountainous area in the north of Spain. The videos consists of images of size 840×377 pixels, acquired at a frame rate of 20 Hz. The UAV flies at an altitude of 4.66 km, simulating a medium-altitude long-endurance (MALE) UAV (from 3 km to 10 km approximately) covering an area of 10×5 km in approximately half a minute. Figure 1 shows several images of the input videos to the terrain modeling pipeline.

Next, the images are fed to Algorithm 1 described in Sect. 2.1.1. Figure 2 shows the result of the sparse reconstruction stage (step 3). Although the density of terrain points is not sufficient to accurately model the surface, this step provides precise locations of the cameras with respect to the terrain, which is used in the next step of Algorithm 1.

Figures 8 to 7 show dense terrain 3D reconstruction results obtained with the variational disparity method, after step 5 of Algorithm 1. The domain Ω matched in the reference image(s) was discretized on a grid of 513×257 pixels, and a 6-level multigrid solver[25] with 200 iterations, 2 V-cycles per iteration and one pre- and post-relaxation sweeps per level was used. The weight $\alpha = 4000$ was empirically chosen to provide a sufficiently smooth surface. Figure 6 shows on the left, the terrain elevation model (geometric information only) and on the right, the same model with its corresponding texture (geometry + photometry). The figure also displays

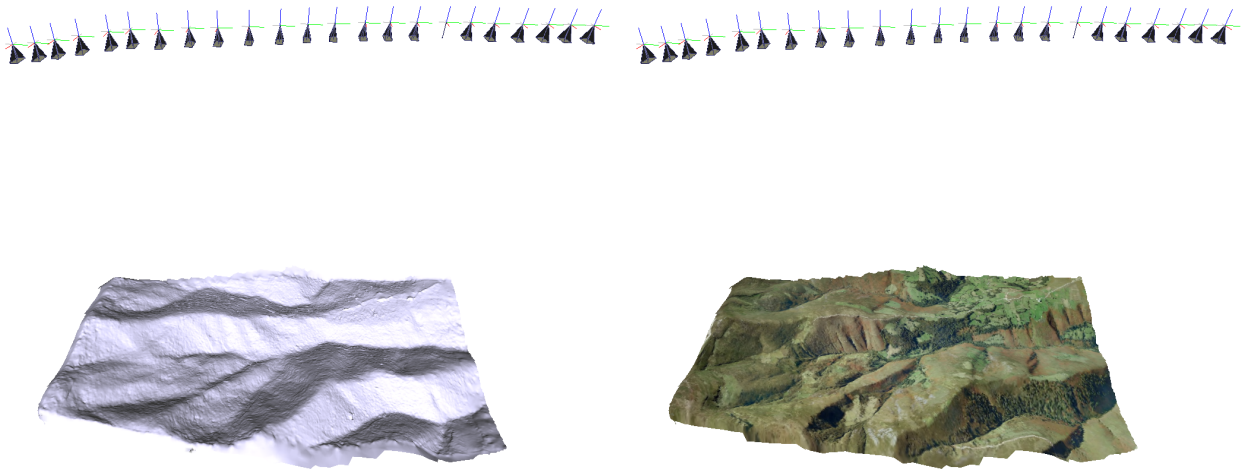

Figure 6. Experiment 1. A portion of the dense terrain elevation model and camera trajectory obtained from variational disparity method. Left: shaded model (geometry). Right: textured model (geometry and photometry).

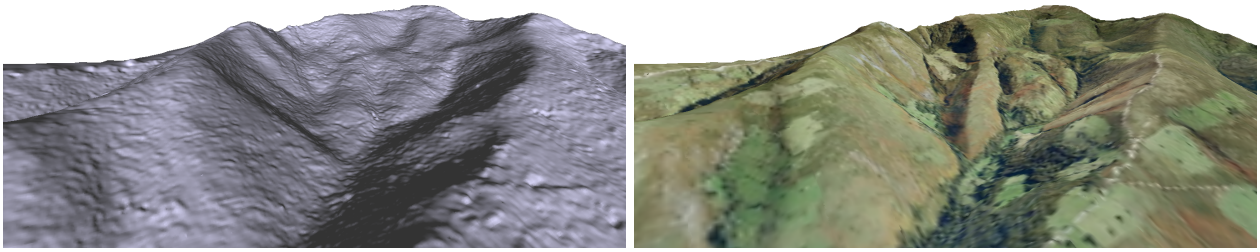

Figure 7. Experiment 1. Dense terrain elevation model obtained from variational disparity method. Close-up point of view.

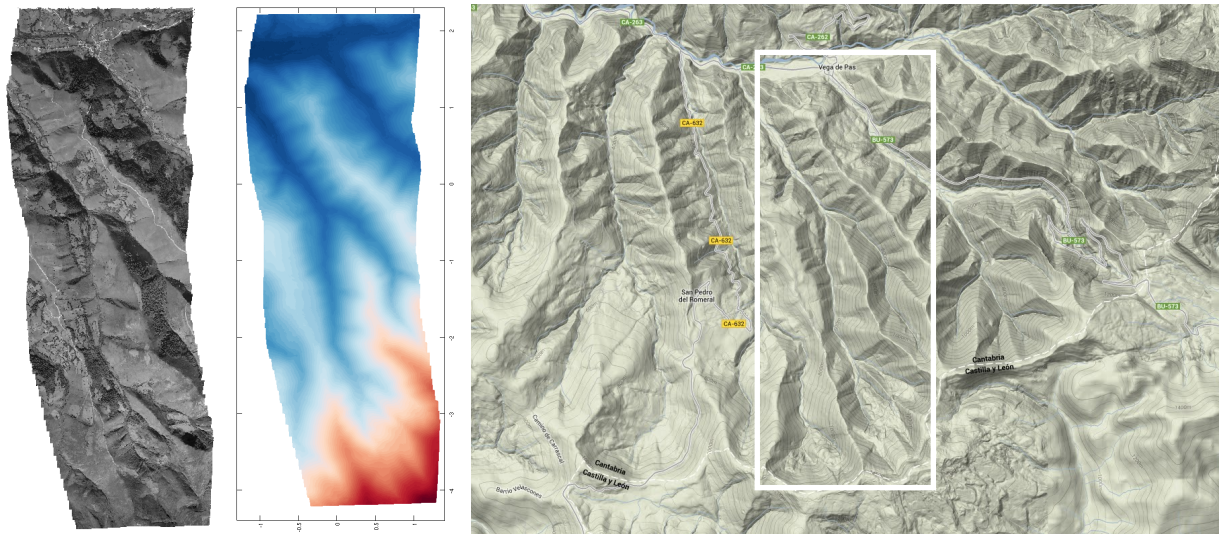

Figure 8. Experiment 1. Terrain elevation model in $43.079° \leq$ latitude $\leq 43.161°$ N, $3.758° \leq$ longitude $\leq 3.802°$ W, obtained by Algorithm 1: textured (left) and pseudo-colored (center), from blue (low) to red (high). Right: terrain elevation model (with shaded-relief details, obtained from Google Maps) of the area enclosing the region of interest (highlighted by a rectangle).

Figure 9. Experiment 2. Predicted images by transferring intensities according to the correspondence given by the disparity map (cf. Fig. 4). Left: original image I_1 (outside the rectangle Ω, of size 513×257 pixels) and predicted image $\hat{I}_1(\mathbf{x}_1) = I_2(\mathbf{x}_1 + \mathbf{d}(\lambda))$ (inside Ω). Center: Tangential disparity $\lambda(\mathbf{x}_1)$, pseudo-colored in grayscale expanding the range of λ, in this example $\lambda \in [-18.22, -8.21]$ pixels. Right: predicted image $\hat{I}_2(\mathbf{x}_2) = I_1(\mathbf{x}_2 - \mathbf{d}(\lambda))$ (matched region) and original image I_2 (outside).

the camera position in which the images were collected during the UAV flight. As it is expected, the variational method shows a remarkable performance in terms of the high density of sampling points in the surface model compared to the sparse reconstruction obtained by SFM methods, an effect that is even more perceptible in Figure 7, which shows a close-up of the reconstruction from a low altitude point of view. Hence, these high resolution models are appropriate for image geo-registration. Figure 8 shows a comparison of the obtained terrain elevation model with a high resolution terrain model of the surrounding area. Each input image pixel corresponds to a real world spacing of approximately 4.6 meters. The generated terrain model (Fig. **8**, center) is defined on a grid of 1000×2480 points, with a ground sampling distance (GSD) of 3.7 meters, thus covering an area of 9.17×3.7 km. Indeed, for fixed focal length acquisition conditions, distance between terrain samples depends on the UAV flight altitude. The generated terrain model will be most useful as a reference in the geo-registration phase for UAVs flying at altitudes around the acquisition altitude (4.66 km) and above. This is so because the accuracy of the model is limited by its resolution (GSD) and UAVs flying at lower altitudes may required higher resolution models.

Experiment with tilted camera. Figures 9 and 10 show the results of the second experiment (tilted camera). The UAV flies at the same altitude as Experiment 1 and covers approximately the same area in half a minute. The video acquisition settings and the processing steps are also the same. The reconstruction is more challenging than the previous one because the tilt angle implies that terrain points that are closer to the camera will be more accurately estimated than those further away. However, some of the latter will be better estimated as the UAV moves since they will be closer with respect to a later camera position. For this reason, as shown in Fig. 9 (left), we choose the location of the disparity grid (Ω) in the region of the image where the closer points to the camera project, since such points will be triangulated with less uncertainty.

In addition, observe that in this setup a rectangle in the image plane maps to a trapezoidal shape in the scene (in a quasi flat world), and therefore a uniform sampling in the image will give a non-uniform sampling of the terrain according to the perspective transformation. The front-to-parallel setup in Experiment 1 is a better strategy for sampling the terrain in a more uniform manner.

In Fig. 10 (center), the GSD of the obtained terrain is approximately 4.5 m, larger than in Experiment 1 due to the deviation from the front-to-parallel configuration: an image pixel covers more terrain area than before. Fine details can be better resolved than in the 90 m DEM of Fig. 10 (right), specifically this is more evident in the textured model (Fig. 10, left), e.g., in the road at the southernmost region.

4. CONCLUSION

We have presented a stereo processing pipeline for building a dense terrain model from images of the UAV video feed in case that a reference DEM is needed for geo-registration but it is unavailable or not present. The proposed method uses variational methods and multi-resolution to generate a dense disparity map, providing surface continuity not only along epipolar lines but also across them, and yielding a dense and coherent surface model.

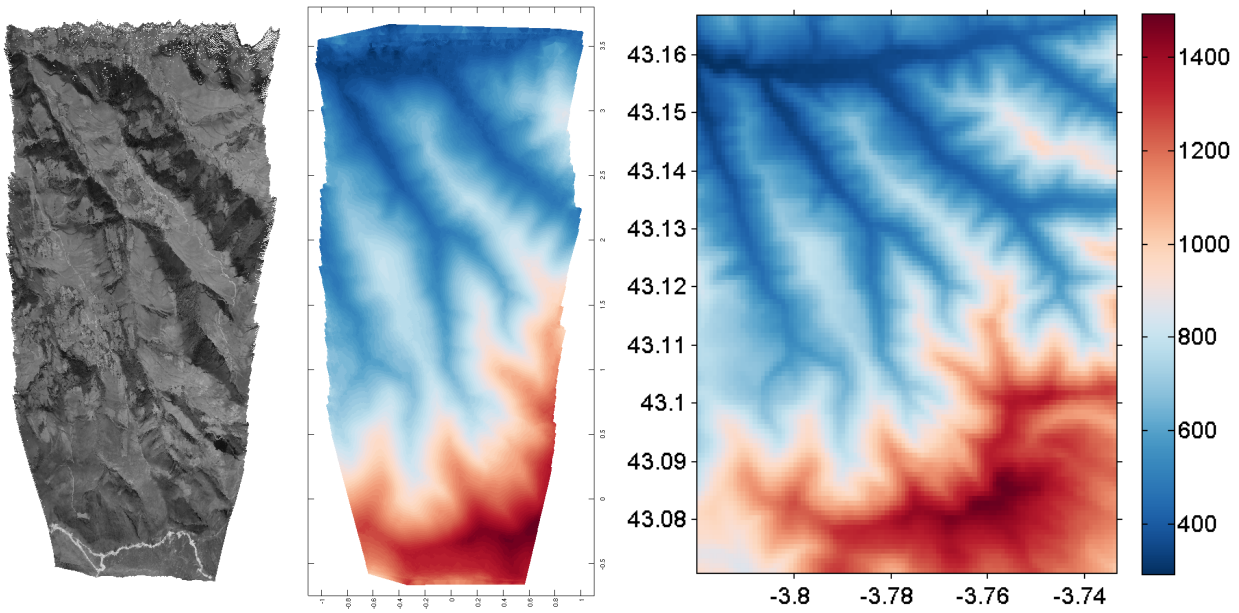

Figure 10. Experiment 2. Terrain elevation model obtained by Algorithm 1: textured (left) and pseudo-colored (center), from blue (low) to red (high). Right: (low resolution) DEM of the surrounding area enclosing the region of interest (NASA Shuttle Radar Topographic Mission (SRTM) 90m DEM obtained from[30]), also pseudo-colored; cf. Fig. 8. Axes are latitude and longitude (in degrees); color legend (elevation), in meters.

Experiments have been carried out simulating a MALE UAV with two different camera orientations, showing that terrain recovery is possible with both a downward looking camera and a tilted one, the first configuration being better than the second one from the point of view of a uniform sampling of the terrain. The surface models obtained with the proposed method are better than the sparse ones achieved with SFM techniques, providing the capability to create a high resolution DEM.

In the proposed disparity method, depth of the scene is not taken into account in the variational step because the cameras need only be weakly calibrated, nor is the normal of the surface with respect to the cameras. In future work, we plan to incorporate other techniques that take into account both the depth of the scene and the surface normals, in an object-centered reconstruction approach.

ACKNOWLEDGMENTS

This work has been partially supported by the Ministerio de Economía y Competitividad of the Spanish Government under project TEC2010-20412 (Enhanced 3DTV) and by Airbus Defence and Space under project SAVIER, Open Innovation Program.

REFERENCES

1. Zhang, J., Liu, W., and Wu, Y., "Novel technique for vision-based uav navigation," *Aerospace and Electronic Systems, IEEE Transactions on* **47**(4), 2731–2741 (2011).
2. Pritt, M. and LaTourette, K., "Automated georegistration of motion imagery," in [*IEEE Applied Imagery Pattern Recognition Workshop (AIPR)*], 1–6 (2011).
3. Doucette, P., Antonisse, J., Braun, A., Lenihan, M., and Brennan, M., "Image georegistration methods: A framework for application guidelines," in [*IEEE Applied Imagery Pattern Recognition Workshop: Sensing for Control and Augmentation (AIPR)*], 1–14, IEEE (2013).
4. Yue, T.-X., Du, Z.-P., Song, D.-J., and Gong, Y., "A new method of surface modeling and its application to DEM construction," *Geomorphology* **91**(1-2), 161 – 172 (2007).
5. Pritt, M. D. and LaTourette, K. J., "Georegistration of motion imagery with error propagation," in [*SPIE Defense, Security, and Sensing*], 838606–838606, Int. Soc. for Optics and Photonics (2012).

6. Mckay, T. and Hirsch, H., "A fast, accurate, cross-modality image geo-registration and target/object detection algorithm," *Proc. SPIE* **8747**, 874708–874708–8 (2013).
7. Shah, M. and Kumar, R., [*Video Registration, Chapter 8*], The International Series in Video Computing, Springer US (2003).
8. Eugster, H. and Nebiker, S., "Geo-registration of video sequences captured from Mini UAVs: Approaches and accuracy assessment," in [*5th Int. Symposium on Mobile Mapping Technology*], **12** (2007).
9. LaTourette, K. and Pritt, M., "Dense 3D reconstruction for video stabilization and georegistration," in [*IEEE Int. Geoscience and Remote Sensing Symposium (IGARSS)*], 6737–6740 (2012).
10. Hartley, R. I. and Zisserman, A., [*Multiple View Geometry in Computer Vision*], Cambridge University Press (2004).
11. Ma, Y., Soatto, S., Kosecka, J., and Sastry, S., [*An Invitation to 3D Vision*], Springer Verlag (2003).
12. Faugeras, O., Luong, Q.-T., and Papadopoulo, T., [*The Geometry of Multiple Images*], The MIT Press (2001).
13. Tuytelaars, T. and Mikolajczyk, K., "Local Invariant Feature Detectors: A Survey," *Found. Trends. Comput. Graph. Vis.* **3**(3), 177–280 (2008).
14. Lowe, D. G., "Distinctive Image Features from Scale-Invariant Keypoints," *Int. J. Computer Vision* **60**(2), 91–110 (2004).
15. Bay, H., Ess, A., Tuytelaars, T., and Gool, L. V., "Speeded-Up Robust Features (SURF)," *Computer Vision and Image Understanding* **110**(3), 346–359 (2008). Similarity Matching in Computer Vision and Multimedia.
16. Alcantarilla, P. F., Bartoli, A., and Davison, A. J., "KAZE Features," in [*Computer Vision – ECCV 2012*], 214–227, Springer Berlin Heidelberg (2012).
17. Ronda, J. I., Valdés, A., and Jaureguizar, F., "Camera Autocalibration and Horopter Curves," *Int. J. Computer Vision* **57**(3), 219–232 (2004).
18. Ronda, J. I., Valdés, A., and Gallego, G., "Line Geometry and Camera Autocalibration," *J. Math. Imaging Vis.* **32**(2), 193–214 (2008).
19. Snavely, N., Seitz, S. M., and Szeliski, R., "Modeling the World from Internet Photo Collections," *Int. J. Computer Vision* **80**(2), 189–210 (2008).
20. Herrero, N., Landabaso, J.-L., Gallego, G., and Pujol-Alcolado, J.-C., "In-loop feature tracking for structure and motion with out-of-core optimization," in [*IEEE Int. Conf. Image Processing (ICIP)*], 2937–2940 (2010).
21. Fischler, M. A. and Bolles, R. C., "Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography," *Commun. ACM* **24**(6), 381–395 (1981).
22. Trucco, E. and Verri, A., [*Introductory Techniques for 3-D Computer Vision*], Prentice Hall PTR, Upper Saddle River, NJ, USA (1998).
23. Alvarez, L., Deriche, R., Sánchez, J., and Weickert, J., "Dense Disparity Map Estimation Respecting Image Discontinuities : A PDE and Scale-Space Based Approach," *J. Visual Comm. and Image Rep.* **13**, 3–21 (2002).
24. Gallego, G., Yezzi, A., Fedele, F., and Benetazzo, A., "Two variational stereo methods for space-time measurements of ocean waves," in [*ASME 2013 32nd Int. Conf. on Ocean, Offshore and Arctic Engineering (OMAE2013)*], **5**, V005T06A041– (2012).
25. Briggs, W. L., Henson, V. E., and McCormick, S. F., [*A Multigrid Tutorial, Second Edition*], SIAM (2000).
26. Furukawa, Y. and Ponce, J., "Accurate, Dense, and Robust Multiview Stereopsis," *IEEE Trans. Pattern Anal. Mach. Intell.* **32**(8), 1362–1376 (2010).
27. Gallego, G., *Variational Image Processing Algorithms for the Stereoscopic Space-Time Reconstruction of Water Waves*, PhD thesis, Georgia Institute of Technology, Atlanta, GA, USA (2011). Directors: Yezzi, A. and Fedele, F.
28. Kazhdan, M., Bolitho, M., and Hoppe, H., "Poisson surface reconstruction," in [*Proceedings of the Fourth Eurographics Symposium on Geometry Processing*], *SGP '06*, 61–70, Eurographics Association, Aire-la-Ville, Switzerland, Switzerland (2006).
29. Google Earth. http://www.google.com/earth/.
30. The CGIAR Consortium for Spatial Information. http://srtm.csi.cgiar.org/.