

A PHOTOGRAMMETRY-BASED STRUCTURE FROM MOTION ALGORITHM USING ROBUST ITERATIVE BUNDLE ADJUSTMENT TECHNIQUES

Styliani Verykokou*, Charalabos Ioannidis

Laboratory of Photogrammetry, School of Rural and Surveying Engineering, National Technical University of Athens, Greece
st.verykokou@gmail.com, cioannid@survey.ntua.gr

KEY WORDS: Structure from Motion, Bundle Adjustment, Exterior Orientation, Homography, Coplanar Feature Points, Template Matching, Oblique Images

ABSTRACT:

The purpose of this paper is the presentation of a novel algorithm for automatic estimation of the exterior orientation parameters of image datasets, which can be applied in the case that the scene depicted in the images has a planar surface (e.g., roof of a building). The algorithm requires the measurement of four coplanar ground control points (GCPs) in only one image. It uses a template matching method combined with a homography-based technique for transfer of the GCPs in another image, along with an incremental photogrammetry-based Structure from Motion (SfM) workflow, coupled with robust iterative bundle adjustment methods that reject any remaining outliers, which have passed through the checks and geometric constraints imposed during the image matching procedure. Its main steps consist of (i) determination of overlapping images without the need for GPS/INS data; (ii) image matching and feature tracking; (iii) estimation of the exterior orientation parameters of a starting image pair; and (iv) photogrammetry-based SfM combined with iterative bundle adjustment methods. A developed software solution implementing the proposed algorithm was tested using a set of UAV oblique images. Several tests were performed for the assessment of the errors and comparisons with well-established commercial software were made, in terms of automation and correctness of the computed exterior orientation parameters. The results show that the estimated orientation parameters via the proposed solution have comparable accuracy with those ones computed through the commercial software using the highest possible accuracy settings; in addition, double manual work was required by the commercial software compared to the proposed solution.

1. INTRODUCTION

Advances in photogrammetry and computer vision have led to the development of the Structure from Motion (SfM) approach, which has seen tremendous evolution over the years. SfM refers to the process of estimating the camera poses that correspond to a 2D image sequence and reconstructing the sparse 3D scene geometry. The combination of SfM and Multi-View Stereo (MVS) methods offer an automated workflow for the generation of high-accuracy dense 3D point clouds.

A variety of SfM methods have been proposed so far and can be basically categorized as incremental (e.g., Snavely et al., 2006; Frahm et al., 2010; Agarwal et al., 2011; Wu, 2013; Shah et al., 2014), hierarchical (e.g., Farenzena et al., 2009; Gherardi et al., 2010; Ni and Dellaert, 2012) and global methods (e.g., Arie-Nachimson et al., 2012; Jiang et al., 2013; Moulon et al., 2013; Wilson and Snavely, 2014; Cui and Tan, 2015). Among them, incremental SfM methods are the most widely used ones. Such methods refer to the gradual incorporation of successive views in the sparse 3D reconstruction, as one image is added at a time. According to the general workflow followed by incremental SfM methods, two or three starting images are automatically selected and the corresponding camera poses are retrieved either from the fundamental matrix of the starting pair or the trifocal tensor of the starting triplet respectively, up to a projective transformation of the 3D space. 3D coordinates of the corresponding points in the starting images are retrieved through triangulation. Several methods may be used for the registration of successive views, such as epipolar constraints that relate each image to its predecessor, resection or merging of

partial reconstructions (Robertson and Cipolla, 2009). A metric reconstruction is obtained either via an auto-calibration procedure or using known calibration data. The final stage of the SfM pipeline is usually a bundle adjustment, that is, a nonlinear optimization in order to refine the camera poses and the 3D coordinates of points. The georeferencing of the derived SfM results is usually accomplished by estimating the 3D similarity transformation between the arbitrary SfM coordinate system and the world reference system using ground control points (GCPs) and/or GPS data (Verykokou et al., 2018).

The purpose of this paper is the presentation of a novel algorithm for automatic estimation of the exterior orientation of images, which requires the measurement of four coplanar GCPs in only one image. The proposed algorithm uses an incremental SfM workflow for the orientation of images. One of the challenges that it is intended to solve is the reduction of the required manual work with the scope of increasing even more the automation of the SfM process. Yet another challenge is the difficulty in matching correctly feature points among multiple views; the proposed algorithm solves this by eliminating all the erroneous tie points through the combination of multiple checks and geometric constraints imposed during the image matching procedure and robust iterative bundle adjustment methods.

The rest of this paper is organized as follows. In section 2, the steps of the proposed algorithm are described in detail. Section 3 presents the experiments performed for testing the algorithm through an in-house developed software solution. In section 4, the results of the proposed algorithm are outlined and assessed in comparison with those ones achieved through well-

* Corresponding author

established commercial software. In section 5, the originality of our work is discussed and some indicative application scenarios for the proposed algorithm are outlined. Finally, the conclusions of our work along with future research steps are presented in section 6.

2. PROPOSED ALGORITHM

In this section, the proposed algorithm for the determination of the exterior orientation parameters of images is presented in detail. It requires (i) a dataset of images without the need for GPS/INS information; (ii) the pixel coordinates of four coplanar GCPs, measured in one image; (iii) the ground coordinates of the GCPs; and (iv) the interior orientation of the imagery along with their pixel size, or at least the focal length and pixel size of the imagery which are almost always available in the image metadata. Although the algorithm requires the input of the image coordinates of GCPs in one image, it assumes that the GCPs are also visible in a second one. For the efficient implementation of the algorithm, its corresponding to numbers ranging from 1 (that corresponds to the image where the GCPs are measured) to the total number of images are assigned to each image. The proposed algorithm is illustrated in Figure 1.

2.1 Determination of overlapping images

The overlapping images are determined in a first stage, so that the subsequent search of correspondences takes place solely in corresponding images, in order to minimise the computational burden and the processing time of image matching. As a first step, the images are resampled to a sufficiently low resolution to speed up the process. Then, features are extracted in each image after it has been converted to greyscale, using the speeded-up robust features (SURF) algorithm by Bay et al. (2008).

At the step of finding correspondences between the images, the feature points extracted from an image are compared to the feature points extracted from all the other images, using the criterion of the minimum Euclidean distance between their descriptors. The correspondences are rejected if the distance between the descriptors of the matched feature points is above a maximum accepted value. Also, they are geometrically verified via the random sample consensus (RANSAC) algorithm (Fischler and Bolles, 1981), through the computation of the fundamental matrix using the eight-point algorithm (Hartley, 1997). The remaining correspondences may still contain some outliers which, however, do not affect the reliability of the resulting information, that is, whether the images need to be matched in the next step. The output of this stage is the number of correspondences between every image pair, which determines whether the images are overlapping.

2.2 Image matching

SURF feature points are extracted in all images, after the reduction of their dimensions, for the acceleration of the process. The resizing factor and the Hessian threshold used here are smaller than the ones used in the previous stage, resulting in a greater number of features in each image. Feature-based matching is applied only to the overlapping images. A cross-check test is implemented, according to which the similarity of the descriptors is verified through reverse comparison using the criterion of the minimum Euclidean distance, which is constrained to be below a predefined threshold. The RANSAC algorithm is applied for the removal of outliers through the estimation of the fundamental matrix.

Despite this geometric constraint, some incorrect matches still remain; those features that are erroneously considered to match a feature in another image and happen to lie on the epipolar line of the homologous feature under consideration are not detected via RANSAC. Thus, a point-to-point constraint is also imposed, as suggested by Verykokou and Ioannidis (2016). Specifically, a homography is fitted to the matches via RANSAC, using a distance threshold for determining the outliers, which is small enough to reject the erroneous matches yet sufficiently large to

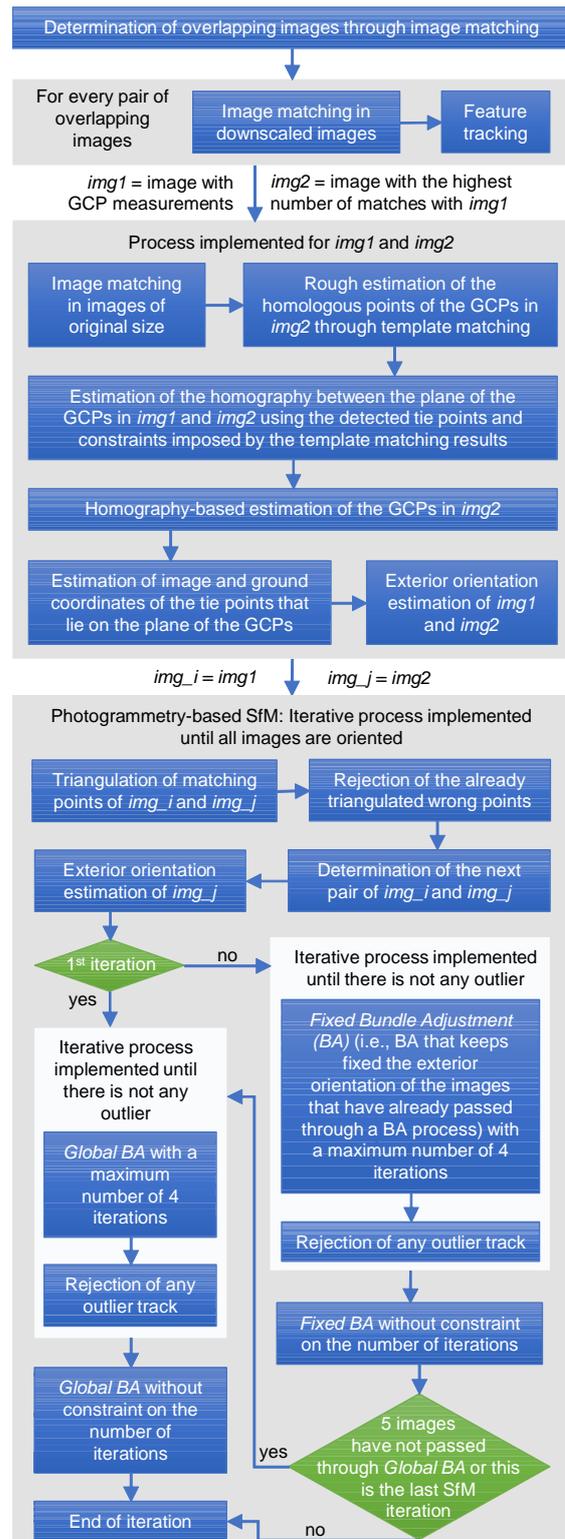


Figure 1. The proposed algorithm

cope with the homography being approximate and not representing the actual relation between two central projections in cases of non-planar scenes and images not acquired from the same point. The corresponding feature points for each image pair are stored in the same position of two vectors, for efficient implementation of the feature tracking process. Also, they are sorted by ascending order of the x pixel coordinates of the feature points of the first image of each pair, for the scope of faster search for points during the feature tracking process.

2.3 Feature tracking

The correspondences are organized into tracks, each containing the coordinates of the feature points in different images that correspond to the same 3D point. Feature tracking takes place for each pair of overlapping images after the image matching procedure for this pair. The proposed methodology for the implementation of feature tracking is based on the proper formation of one vector and two matrices, as explained in the following. An index of zero is assumed to correspond to the first row and column of these data structures.

- A vector (*processedImages*) that stores information about whether feature points have been stored for each image. Its size equals to the number of images. Its elements may take the values of 0 and 1; an element at position r that equals to 0 indicates that feature points have not been stored for the image with id $r+1$; a value of 1 indicates that feature points have been stored for the corresponding image.
- A matrix (*imgPoints*) that stores the image coordinates of feature points, with the number of rows being equal to the number of images and the number of columns being equal to the number of tracks, that is, the number of 3D points. Each row stores information for one image and each column stores information for each track. For instance, the element at row r and column c corresponds to the pixel coordinates of the feature point with id $c+1$ at image with id $r+1$. If a point is not visible in an image, the coordinates of (-1, -1) are stored in the corresponding element of this matrix.
- A matrix (*visibility*) that stores information about whether a point is visible in each image. Its dimensions are the same with the ones of the *imgPoints* matrix. Its elements may take the values of 0 and 1, with 0 indicating non-visibility and 1 indicating visibility. For instance, an element at row r and column c that equals to 0 indicates that the feature point with id $c+1$ is not visible in image with id $r+1$.

The *processedImages* vector and the *visibility* matrix are initialized with the value of 0, whereas the *imgPoints* matrix is initialized with the values of (-1, -1); their initialization takes place before the image matching process. Feature tracking is applied for each image pair by filling the aforementioned vector and matrices. The algorithm is presented in the form of pseudocode in Figure 2; *rowImg_k* and *colImg_k* represent the row and column number, respectively, of the vector or matrix, that corresponds to image k ; *maxColNum* represents the maximum column number of *imgPoints* that contains a feature point; *pointImg_{kp}* represents the feature point with id p in image k , respectively.

2.4 Exterior orientation estimation of the first image pair

The idea behind the computation of the exterior orientation of the first image pair based on coplanar GCPs manually measured in the first image is the automatic detection of points in both images that belong to the plane of the GCPs followed by computation of their ground coordinates and estimation of the

Algorithm: Feature tracking

```

1: if (processedImages(rowImgi) = 0 and
   processedImages(rowImgj) = 0) then
2:   for each pair  $p$  of corresponding feature points do
3:     find maxColNum
4:     for each  $img\ k \in [i, j]$  do
5:       colImgk  $\leftarrow$  maxColNum + 1
6:       imgPoints(rowImgk, colImgk)  $\leftarrow$  pointImgkp
7:       visibility(rowImgk, colImgk)  $\leftarrow$  1
8:     for each image  $k \in [i, j]$  do
9:       processedImages(rowImgk)  $\leftarrow$  1
10:  else if ((processedImages(rowImgi) = 1 and
   processedImages(rowImgj) = 0) or
   (processedImages(rowImgi) = 0 and
   processedImages(rowImgj) = 1)) then
11:   for each pair  $p$  of corresponding feature points do
12:      $m \leftarrow$  index for the processed img
13:      $k \leftarrow$  index for the not processed img
14:     if pointImgmp has been saved then
15:       find colImgm where pointImgmp has been saved
16:       colImgk  $\leftarrow$  colImgm
17:       do steps 6-7
18:     else
19:       do steps 3-7
20:   processedImages(rowImgk)  $\leftarrow$  1
21: else
22:   for each pair  $p$  of corresponding feature points do
23:     if pointImgip or pointImgjp has been saved then
24:       if its corresponding point has not been saved then
25:          $m \leftarrow$  index for img with the already saved point
26:          $k \leftarrow$  index for the other img
27:         do steps 15-17
28:       else
29:         do steps 3-7

```

Figure 2. Feature tracking algorithm applied for each image pair

orientation of each image via photogrammetric space resection. The steps of this algorithm are described in the following.

2.4.1 Image matching: The image (*img2*) that is going to be a pair with the image where the GCPs are measured (*img1*) is detected as the one with the maximum number of corresponding feature points with *img1*. The image matching procedure described in section 2.2 is applied in these images with small differences for the scope of extracting a greater number of corresponding feature points: (i) images of original dimensions are used; and (ii) a smaller Hessian threshold for SURF feature extraction is applied. Then, the fundamental matrix estimated via RANSAC is recomputed using the correct matches, i.e., the ones geometrically verified by RANSAC. Subsequently, the epipolar lines in *img2* for the GCPs measured in *img1* are computed using the estimated fundamental matrix.

2.4.2 Template matching: A template window, hereinafter symbolized as *template*, is defined around every GCP in *img1* and their homologous windows are estimated in *img2* through template matching. According to this method, *template* slides pixel to pixel in *img2*, left to right and up to down. At each location of this sliding window, a metric is calculated that represents the similarity between *template* and *img2* and is stored in a result matrix R . The metric used by the proposed algorithm is computed using equation (1), where i and j represent the row and column number of *img2*, while i' and j' represent the row and column number of *template*.

$$R(i, j) = \frac{\sum_{i', j'} [template(i', j') \cdot img2(i + i', j + j')]}{\sqrt{\sum_{i', j'} template(i', j')^2 \cdot \sum_{i', j'} img2(i + i', j + j')^2}} \quad (1)$$

After normalization of the matrix R , the best match is found as a global maximum. The top left corner of the homologous window of $template$ in $img2$ is the location of the global maximum and its width equals to the width of $template$. This template matching procedure applies for all the manually measured GCPs in $img1$. The center point of each rectangle detected in $img2$ as the homologous window of each $template$ is assumed to be the rough homologous point of each GCP in $img2$, found through template matching.

2.4.3 Transfer of the GCPs: An iterative procedure is implemented for the estimation of the correct homography transformation between the plane of the GCPs in $img1$ and the plane of their homologous points in $img2$. Specifically, a homography transformation from $img2$ to $img1$ is estimated via RANSAC using a small distance threshold between the actual image points and the projected ones via homography, so that the inliers that verify the computed transformation belong to a plane. The pixel coordinates of the GCPs are transformed from $img1$ to $img2$ using the inverse homography matrix. If the homography-based estimation of each GCP in $img2$ lies on the correct epipolar line, estimated as described in section 2.4.1, the distance in $img2$ between the homography-based estimation of each GCP and the rough homologous point found via template matching is calculated. If it is smaller than a threshold for every GCP, the homography-based estimations of the GCPs are considered to be the homologous points of GCPs in $img2$ and the computed homography is considered to be the correct one, that expresses the relation between the plane of GCPs in $img1$ and $img2$. Otherwise, the computed homography represents the relation between points that belong to another plane visible in these two images. In this case, the points that are returned as inliers by RANSAC are removed from the set of matches, as they do not belong to the plane where the GCPs lie, and the same process is repeated until the correct homography is found.

2.4.4 Detection of coplanar points: After computation of the homography that expresses the relation between the plane of the GCPs in $img1$ and the homologous plane in $img2$, the points that lie on this plane are identified in both images. Specifically, the corresponding feature points in $img1$ and $img2$ that are identified as inliers by RANSAC during the computation of the correct homography transformation are stored by the algorithm as points that belong to the plane on which the GCPs lie, in addition to the GCPs in $img1$ and their corresponding points in $img2$, estimated using the correct inverse homography matrix.

2.4.5 Estimation of ground coordinates: The homography transformation from the plane of GCPs in $img1$ to the real world plane is computed using the pixel coordinates of the GCPs and their ground coordinates in the reference system. The ground coordinates of the corresponding features of $img1$ and $img2$ that belong to the plane of the GCPs are calculated using their pixel coordinates in $img1$ in combination with the estimated homography transformation from $img1$ to the real world plane.

2.4.6 Space resection: The exterior orientation parameters of $img1$ and $img2$ are estimated through conventional photogrammetric space resection using the collinearity equations as the mathematical model. For this reason, the pixel

coordinates of the coplanar points in each image are converted to millimetres in the photogrammetric system and these ones along with the ground coordinates of the coplanar points in the reference coordinate system and the interior orientation parameters of the camera that acquired the images are combined in a least-squares solution for the iterative computation of the exterior orientation parameters of each image separately. The initial exterior orientation parameters required for the least-squares solution are estimated through a Levenberg-Marquardt (Moré, 1978) optimization method that solves the Perspective-n-Point problem (Lepetit et al., 2009), which minimizes the reprojection error, that is, the sum of squared distances between the observed image points and the projected ones in the image using the estimated camera pose parameters.

2.5 Photogrammetry-based SfM

An incremental SfM procedure that implements robust iterative bundle adjustment methods takes place as the last stage of the proposed algorithm. Its main difference compared to other state-of-the-art SfM techniques is its implementation using conventional well-established photogrammetric methods rather than computer vision techniques. The steps of the iterative procedure of the proposed SfM method, which is terminated when all images have been oriented, are described in the following. The starting images of the iterative procedure of the SfM algorithm are $img1$ and $img2$, i.e., the image where the GCPs have been measured and the image where the GCPs have been automatically found, respectively; the exterior orientation parameters of these images have already been estimated (section 2.4.6). The images that are used for the determination of the 3D coordinates of feature points in each iteration of the SfM algorithm are hereinafter symbolized as img_i and img_j .

2.5.1 Space intersection: The first stage is the triangulation of the feature points between img_i and img_j , that is, the computation of their 3D coordinates in the reference system. A least-squares solution that solves the conventional problem of photogrammetric space intersection is adopted for every pair of corresponding points, based on the collinearity equations. The initial data required for computing the 3D coordinates of a point depicted in img_i and img_j are (i) its coordinates in img_i (x_i, y_i) and img_j (x_j, y_j) expressed in the photogrammetric system; (ii) the interior orientation parameters of the camera that took the images; and (iii) the exterior orientation of the images. The initial values for the 3D coordinates of the point to be triangulated ($X_{approx}, Y_{approx}, Z_{approx}$) required for the iterative least-squares solution are estimated using equations (2)-(4).

$$Z_{approx} = \frac{(X_{0j} - X_{0i})c_i'c_j' - x_i'c_j'Z_{0i} + x_j'c_i'Z_{0j}}{-x_i'c_j' + x_j'c_i'} \quad (2)$$

$$X_{approx} = X_{0i} - \frac{x_i'(Z_{approx} - Z_{0i})}{c_i'} \quad (3)$$

$$Y_{approx} = Y_{0i} - \frac{y_i'(Z_{approx} - Z_{0i})}{c_i'} \quad (4)$$

In the above equations, X_{0i}, Y_{0i}, Z_{0i} and X_{0j}, Y_{0j}, Z_{0j} are the projection center coordinates of img_i and img_j respectively; x_i', y_i', c_i' and x_j', y_j', c_j' are computed via equations (5) and (6), where c is the principal distance and R_i is the rotation matrix of img_i , calculated using the angular exterior orientation elements (omega, phi and kappa angles) of img_i .

$$\begin{bmatrix} x_i' \\ y_i' \\ -c_i' \end{bmatrix} = R_i^T \cdot \begin{bmatrix} x_i \\ y_i \\ -c \end{bmatrix} \quad (5)$$

$$\begin{bmatrix} x_j' \\ y_j' \\ -c_j' \end{bmatrix} = R_j^T \cdot \begin{bmatrix} x_j \\ y_j \\ -c \end{bmatrix} \quad (6)$$

2.5.2 Removal of wrong points: If a point has already been triangulated through a different combination of overlapping images, the difference between its newly computed ground coordinates and its ground coordinates as computed in a previous iteration of the SfM process is calculated. If this difference is above a maximum accepted threshold, the entire track is considered to be erroneous and is removed from the initial data. If the computed difference is below this threshold, the already calculated 3D ground coordinates that correspond to the track are considered to be correct. Otherwise, if the point has not already been triangulated, its coordinates, calculated in the current SfM iteration, are stored by the algorithm.

2.5.3 Determination of the next pair of *img_i* and *img_j*: The non-oriented image with the maximum number of already triangulated feature points is considered to be *img_j*; *img_i* is the oriented image with the maximum number of triangulated points that match with those ones of *img_j*.

2.5.4 Space resection: The exterior orientation parameters of *img_j* are estimated through photogrammetric space resection as described in section 2.4.6.

2.5.5 Iterative fixed bundle adjustment: Some erroneous tie points may still exist; these are the points that have passed through the checks applied in the image matching procedure (section 2.2) and have not already been triangulated by a different combination of overlapping images, because in that case they would have been rejected, as described in section 2.5.2. An iterative process implementing a least-squares bundle adjustment method is applied in order to remove possible erroneous tie points and ensure that there are not any outliers in the solution. This stage is implemented in every iteration of the SfM procedure after the space resection step, with the exception of the first SfM iteration; in this case, the process described in section 2.5.6 is adopted. Specifically, this iterative procedure starts with a bundle adjustment solution which refines the exterior orientation of *img_j* and the 3D ground coordinates of the feature points that have been triangulated until the current SfM iteration; the exterior orientation parameters of all the other oriented images, as well as the interior orientation parameters are considered to be fixed. The mathematical model used by the bundle adjustment is expressed by the collinearity equations. The already calculated ground coordinates of the feature points either refined through a previous bundle adjustment or estimated only through space intersection (which applies to the newly triangulated feature points), as described in section 2.5.1, serve as initial values for the 3D ground coordinates of the points required for the bundle adjustment. The exterior orientation of *img_j* estimated through space resection, as described in section 2.5.4, is used as initial exterior orientation for this image in the bundle adjustment. A maximum number of four iterations for this bundle adjustment method is defined and the maximum value of the differences between the coordinates of the feature points as estimated in the third and the fourth iteration is detected. These differences are computed

in the vector *dx*, which contains the differences between the calculated values of the unknown elements as estimated in two consecutive iterations of the least-squares solution. If the maximum value of these differences is above a maximum accepted adaptive threshold, the value of which increases with the number of oriented images, the corresponding point is removed from the set of points and the bundle adjustment with the predefined number of iterations is repeated, until all outliers have been detected. At the end of this iterative procedure, the same bundle adjustment method is applied without constraint on the maximum number of iterations, using the correct feature points, for refining their 3D ground coordinates and the exterior orientation parameters of *img_j*. The iterations of this last bundle adjustment step are continued until all the values of the calculated vector *dx* are below predefined thresholds.

2.5.6 Iterative global bundle adjustment: This stage is implemented (i) after the space resection step, described in section 2.5.4, in the first iteration of the SfM method; (ii) after the iterative fixed bundle adjustment procedure, described in section 2.5.5, if a number of five images has been added in the incremental SfM procedure without a global bundle adjustment step; and (iii) after the iterative fixed bundle adjustment procedure, described in section 2.5.5, if *img_j* is the last image of the dataset being oriented. Similarly to the process described in section 2.5.5, an iterative procedure that implements a least-squares bundle adjustment is applied. However, contrary to the iterative fixed bundle adjustment solution, the iterative global bundle adjustment in this stage refines the exterior orientation parameters of all the oriented images, in addition to the 3D coordinates of the feature points that have been triangulated until the current iteration of the SfM process. Any possible outliers are detected and removed via estimating the differences of the 3D coordinates of the feature points as calculated in the third and fourth iteration of the bundle adjustment, as described in section 2.5.5. At the end of this iterative procedure, the same global bundle adjustment method is applied without any constraint regarding the number of iterations, for refining the 3D coordinates of the remaining correct feature points and the exterior orientation of all the already oriented images.

3. EXPERIMENTS

A software implementing the proposed methodology has been developed in the C++ programming language, making use of some functionalities offered by the OpenCV library (OpenCV Team, 2018), for image manipulations, and the Eigen Library, (Eigen Team, 2018) for matrix and vector operations.

The developed software that implements the proposed algorithm was tested using three datasets of UAV oblique aerial images (Figure 3), incorporating 15, 25 and 33 images respectively, taken by a Sony Nex-7 camera over the city hall in Dortmund, Germany, with a GSD varying from 1 to 3 cm. The images, with 6000×4000 pixels each and a focal length of 16 mm, are part of a larger dataset acquired for the scientific initiative “ISPRS benchmark for multi-platform photogrammetry”, run in collaboration with EuroSDR (Nex et al., 2015). This dataset also contains the ground coordinates of some targets on the façades of the city hall that exist only in the terrestrial images.

In order to obtain (i) the ground coordinates of four coplanar points for being measured in *img1* and (ii) reference exterior orientation parameters for comparison with the ones computed via the proposed solution, a block of 645 images (446 oblique and vertical aerial images and 199 terrestrial images) was

oriented using the Agisoft PhotoScan software (Agisoft LLC, 2018) through SfM with self-calibrating bundle adjustment using the “high alignment” setting, according to which the images are not downsampled for the tie point extraction process. Specifically, 34 targets of known coordinates on the façades of the city hall were measured in the corresponding terrestrial images (239 measurements in total) and 8 tie points were manually measured in multiple terrestrial and oblique UAV imagery (206 measurements in total). The average reprojection error of the manually measured GCPs and tie points is 0.14 pixels and the average absolute difference between the computed and the input coordinates of the GCPs is 1.7 cm.

The computed X , Y and Z ground coordinates of four coplanar points, lying on the roof of the city hall, which belong to the manually measured tie points of the reference dataset, triangulated via PhotoScan, served as the ground control of the starting image of the blocks that were used for testing the developed software.



Figure 3. Images of the first (D1), second (D2) and third (D3) dataset



Figure 4. Image matching results between a pair of images; top: detected correspondences; bottom: details illustrating the remaining outliers, marked with red and blue ellipses



Figure 5. Automatically detected coplanar points in *img1*

The checks and constraints imposed during the image matching procedure, as described in section 2.2, end up in removing the greatest number of outliers. However, a small number of erroneous points still remain; those feature points that (i) happen to lie on the epipolar line of the homologous feature point under consideration and (ii) are close to the correct homologous point, as the homography-based point-to-point constrained failed to detect it, are not rejected via the algorithm, as shown in Figure 4. Figure 4-top depicts the correspondences detected by the developed software; Figure 4-bottom presents a magnification of the matching results between the pair of images depicted in Figure 4-top; the tie point included in the blue ellipse and the two tie points included in the red ellipse are not correct, as they belong to the aforementioned cases. Thus, the proposed robust iterative bundle adjustment methods that reject these remaining erroneous matches are indispensable. The automatically detected coplanar feature points in *img1* that lie on the same plane as the four GCPs are depicted in Figure 5.

4. RESULTS ASSESSMENT

The exterior parameters of the three blocks of oblique UAV images that were estimated through the developed software were compared to the reference ones computed via PhotoScan using the whole block of UAV and terrestrial images. The maximum (Max), minimum (Min) and arithmetic mean (Avg) of absolute differences between the exterior orientation parameters were computed and the results are presented in the first three columns of Table 1. According to the developed methodology, four coplanar GCPs are measured in *img1*; the feature point extraction takes place in images that correspond to a resizing factor of 5; and the camera interior orientation is kept fixed, assuming a principal distance equal to the focal length, a principal point located at the image center and zero distortion.

For comparison reasons, tests were conducted in PhotoScan, using the same four GCP measurements in *img1* and also measuring these GCPs in *img2*. Tests took place using both the “medium alignment” and “high alignment” PhotoScan settings, corresponding to feature extraction in downsampled images with a resizing factor of 2 and in images of original size respectively. Also, tests were conducted in PhotoScan both using fixed interior orientation, as assumed by the developed software, and applying autocalibration. In all tests, conducted for each one of the three blocks of oblique imagery, the Avg, Max and Min absolute differences between the computed and the reference exterior orientation parameters were estimated. Quite similar results were derived using the “medium alignment” and “high alignment” PhotoScan settings; thus, the results using only the latter setting, are presented in the six last columns of Table 1.

The comparison of the exterior orientation results obtained through the aforementioned tests with the reference data shows that higher accuracy is achieved via the proposed algorithm than the accuracy of the corresponding results obtained through PhotoScan using fixed interior orientation, although the latter does not downsample the images for the tie point extraction process. The autocalibration process of PhotoScan improves the results, which are comparable with the ones obtained via the proposed solution without autocalibration, for the datasets of 15 and 25 images. The exterior orientation of the dataset of 33 images obtained via PhotoScan with autocalibration is of higher accuracy than the one computed via the developed solution. It is also observed that the accuracy of the exterior orientation computed via the developed solution is better when a smaller number of images is used, because of the ground control

	15 images								
	Developed solution			PhotoScan: fixed IO			PhotoScan: autocalibration		
	Avg	Max	Min	Avg	Max	Min	Avg	Max	Min
$ \Delta X_o $ (m)	0.14	0.30	0.01	0.43	0.73	0.24	0.34	0.45	0.27
$ \Delta Y_o $ (m)	0.25	0.33	0.16	0.40	0.59	0.22	0.06	0.14	0.00
$ \Delta Z_o $ (m)	0.24	0.35	0.13	0.26	0.54	0.04	0.04	0.09	0.00
$ \Delta \omega $ (deg.)	0.31	0.41	0.20	0.56	0.68	0.47	0.23	0.24	0.22
$ \Delta \phi $ (deg.)	0.33	0.61	0.01	0.91	1.26	0.67	0.57	0.62	0.54
$ \Delta \kappa $ (deg.)	0.08	0.25	0.02	0.21	0.23	0.19	0.09	0.13	0.02
	25 images								
$ \Delta X_o $ (m)	0.35	1.08	0.01	0.47	0.70	0.12	0.22	0.29	0.12
$ \Delta Y_o $ (m)	0.35	0.54	0.26	0.47	0.80	0.26	0.11	0.14	0.04
$ \Delta Z_o $ (m)	0.18	0.33	0.01	0.47	0.82	0.15	0.20	0.32	0.12
$ \Delta \omega $ (deg.)	0.23	0.45	0.02	0.28	0.41	0.18	0.18	0.21	0.15
$ \Delta \phi $ (deg.)	0.30	0.57	0.04	0.91	1.36	0.42	0.40	0.51	0.27
$ \Delta \kappa $ (deg.)	0.11	0.28	0.01	0.07	0.13	0.00	0.07	0.11	0.03
	33 images								
$ \Delta X_o $ (m)	0.40	0.97	0.00	0.27	0.46	0.11	0.23	0.30	0.12
$ \Delta Y_o $ (m)	0.47	0.78	0.26	0.48	0.75	0.26	0.06	0.10	0.00
$ \Delta Z_o $ (m)	0.24	0.34	0.11	0.49	0.75	0.15	0.23	0.34	0.14
$ \Delta \omega $ (deg.)	0.32	0.77	0.06	0.22	0.38	0.06	0.14	0.18	0.11
$ \Delta \phi $ (deg.)	0.47	0.73	0.18	0.76	0.92	0.42	0.27	0.34	0.16
$ \Delta \kappa $ (deg.)	0.18	0.38	0.02	0.09	0.17	0.00	0.04	0.07	0.01

Table 1. Average, maximum and minimum absolute differences between the computed and the reference exterior orientation parameters via various tests; columns 1-3: developed software solution – feature point tracking in images corresponding to a resizing factor of 5, fixed interior orientation (IO); columns 4-6: PhotoScan – feature point tracking in images of original dimensions, fixed IO; columns 7-9: PhotoScan – feature point tracking in images of original dimensions, autocalibration

existence in only one image of the block and the fixed approximate interior orientation values used. The differences between the exterior orientation parameters of the imagery computed via the developed software and the reference ones increase with an increase of the distance between the area depicted in the imagery and the area of the coplanar GCPs. Hence, better accuracy is observed in the orientation of images closer to the one with GCP measurements. Future work will focus on improving the achievable accuracy in large datasets. In each case, double manual work was required via PhotoScan.

5. DISCUSSION

The novelties/differences of the proposed algorithm compared to other SfM algorithms are (i) the estimation of the orientation of each new image based on a photogrammetric workflow, contrary to existing SfM algorithms that use computer vision methods; (ii) the robust iterative bundle adjustment methods for detection and removal of wrong tie points that remain after the checks and constraints implemented in the image matching process, which can also be applied by SfM algorithms for the removal of possibly wrong manually measured tie points; and (iii) the requirement for manual measurement of GCPs in only one image, thanks to the template matching method combined with the iterative homography estimation technique, contrary to existing SfM methods that require manual GCP measurements in at least two images for the scope of georeferencing of the SfM results (e.g., Moulon et al., 2017; Agisoft LLC, 2018).

The fact that the algorithm is fully automatic in combination with its requirement for a minimum number of measurements in solely a single image makes it easily adoptable by operators without expertise or even basic knowledge on photogrammetry. Some indicative use cases are the following: (i) applications for which a small number of coplanar GCPs (e.g., lying on the roof of a building) that are distributed in a small area compared with the area of the total image dataset is available; (ii) applications for which the measurement of GCPs in the area of interest is not possible; and (iii) applications for which the person in charge of the orientation process is a non-photogrammetrist. For instance, one application scenario for which the advantages of the proposed algorithm are particularly significant is the orientation of a block of images taken by an unmanned aerial vehicle (UAV) in the case of an emergency scenario for 2D mapping (e.g., Boccoardo et al., 2015) or 3D modelling (e.g., Ferworn et al., 2011; Verykokou et al., 2016a, 2016b; Verykokou et al., 2018) of a disaster area. In such emergency response situations (e.g., earthquake, explosion), the measurement of GCPs is usually not possible in the area of the disaster (e.g., debris of a building). Hence, in such situations four GCPs may be measured in a planar area close to the disaster and a UAV may fly towards the latter, starting acquiring images at the area of the GCPs. In addition, the operators in such a scenario (e.g., search and rescue crew members) are, almost certainly, people without knowledge of photogrammetry, which is another fact that highlights the advantages of using the proposed algorithm.

6. CONCLUSIONS AND FUTURE WORK

This paper proposes a methodology for the estimation of the exterior orientation parameters of images using an incremental photogrammetry-based SfM algorithm coupled with robust iterative bundle adjustment techniques. The initial data required are GCP measurements of four coplanar points in only one image, their ground coordinates and the camera interior orientation parameters. Neither GCP measurements in a second image nor GPS/INS data are required. The results achieved via the proposed algorithm, using datasets of UAV oblique aerial images, are comparable to the ones achieved through well-established commercial software and even better in the case that an autocalibration is not applied in the commercial software.

The proposed algorithm can be extended so that a final global autocalibrating bundle adjustment is implemented as a last step of the developed photogrammetry-based SfM approach. Along with the implementation of an autocalibration procedure, our future work will focus on the automatic recognition of the GCPs in every image where they are visible and not only on a second image of the block, for the scope of achieving better accuracy through the increase of automatic image measurements of GCPs. Furthermore, more tests will be carried out to investigate the results that can be achieved using different resizing factors for the process of feature tracking, as well as different number of images that are added in the SfM iterations without a global bundle adjustment. In addition, the proposed algorithm will be tested using different image datasets acquired through various camera configurations that include larger numbers of images and the achievable accuracy will be investigated in relation to the number of images. Finally, the proposed methodology for detecting coplanar points, as implemented in the first pair of images, may be applied for the detection of coplanar points in entire datasets of imagery, which may serve for the imposition of constraints in a bundle adjustment procedure.

ACKNOWLEDGEMENTS

Styliani Verykokou would like to acknowledge the Eugenides Foundation for the financial support through a PhD scholarship.

REFERENCES

- Agarwal, S., Furukawa, Y., Snavely, N., Simon, I., Curless, B., Seitz, S. M., Szeliski, R., 2011. Building Rome in a day. *Commun. ACM* 54 (10), pp. 105-112.
- Agisoft LLC, 2018. Agisoft PhotoScan Software. <http://www.agisoft.com/> (7 May 2018).
- Arie-Nachimson, M., Kovalsky, S. Z., Kemelmacher-Shlizerman, I., Singer, A., Basri, R., 2012. Global motion estimation from point matches. In: *3DIMPVT 2012*, pp. 81-88.
- Bay, H., Ess, A., Tuytelaars, T., Van Gool, L., 2008. Speeded-up robust features (SURF). *Comput. Vis. Image Underst.* 110 (3), pp. 346-359.
- Boccardo, P., Chiabrando, F., Dutto, F., Tonolo, F. G., Lingua, A., 2015. UAV deployment exercise for mapping purposes: evaluation of emergency response applications. *Sensors* 15 (7), pp. 15717–15737.
- Cui, Z., Tan, P., 2015. Global structure-from-motion by similarity averaging. In: *ICCV 2015*, pp. 864-872.
- Eigen Team, 2018. Eigen Library. <http://eigen.tuxfamily.org/> (7 May 2018).
- Farenzena, M., Fusiello, A., Gherardi, R., 2009. Structure-and-motion pipeline on a hierarchical cluster tree. In: *ICCV Workshops*, pp. 1489-1496.
- Ferworn, A., Tran, J., Ufkes, A., D'Souza, A., 2011. Initial experiments on 3D modeling of complex disaster environments using unmanned aerial vehicles. In: *IEEE SSRR*, pp. 167-171.
- Fischler, M., Bolles, R., 1981. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM* 24 (6), pp. 381-395.
- Frahm, J.-M., Fite-Georgel, P., Gallup, D., Johnson, T., Raguram, R., Wu, C., Jen, Y.-H., Dunn, E., Clipp, B., Lazebnik, S., Pollefeys, M., 2010. Building Rome on a cloudless day. In: *ECCV 2010. LNCS, vol 6314*. Springer, pp. 368–381.
- Gherardi, R., M. Farenzena, Fusiello, A., 2010. Improving the efficiency of hierarchical structure-and-motion. In: *CVPR 2010*, pp. 1594-1600, doi: 10.1109/CVPR.2010.5539782.
- Hartley, R. I., 1997. In defense of the eight-point algorithm. *IEEE Trans. Pattern Recogn. Mach. Intell.* 19 (6), pp. 580-593.
- Jiang, N., Cui, Z., Tan, P. A., 2013. A global linear method for camera pose registration. In: *ICCV 2013*, pp. 481-488.
- Lepetit, V., Moreno-Noguer, F., Fua, P., 2009. EPnP: An Accurate O(n) Solution to the PnP Problem. *Int. J. Computer Vision* 81 (2), pp. 155-166.
- Moré, J. J., 1978. The Levenberg-Marquardt algorithm: implementation and theory. In: Watson, G.A. (Ed.), *Numerical Analysis. LNM, vol 630*. Springer, pp. 105-116.
- Moulon, P., Monasse, P., Marlet, R., 2013. Global fusion of relative motions for robust, accurate and scalable structure from motion. In: *ICCV 2013*, pp. 3248-3255.
- Moulon, P., Monasse, P., Perrot, R., Marlet, R., 2017. OpenMVG: Open Multiple View Geometry. In: *Reproducible Research in Pattern Recognition. RRPR 2016. LNCS, vol 10214*. Springer, pp. 60-74.
- Nex, F., Gerke, M., Remondino, F., Przybilla, H.-J., Bäumker, M., Zurhorst, A., 2015. ISPRS benchmark for multi-platform photogrammetry. *The ISPRS Annals II-3/W4*, pp. 135-142.
- Ni, K., Dellaert, F., 2012. HyperSfM. In: *3DIMPVT 2012*, pp. 144-151.
- OpenCV Team, 2018. OpenCV Library. <https://opencv.org/> (7 May 2018).
- Robertson, D. P., Cipolla, R., 2009. Structure from motion. In: Varga, M. (Ed.), *Practical image processing and computer vision*. John Wiley and Sons Ltd., New York.
- Shah, R., Deshpande, A., Narayanan, P. J., 2014. Multistage SFM: Revisiting incremental structure from motion. In: *3DV 2014*, pp. 417-424, doi: 10.1109/3DV.2014.95.
- Snavely, N., Seitz, S. M., Szeliski, R., 2006. Photo tourism: exploring photo collections in 3D. *ACM Transactions on Graphics* 25 (3), pp. 835-846.
- Verykokou, S., Doulamis, A., Athanasiou, G., Ioannidis, C., Amditis, A., 2016a. Multi-scale 3D modelling of damaged cultural sites: use cases and image-based workflows. In: *Digital Heritage. Progress in Cultural Heritage: Documentation, Preservation, and Protection. EuroMed 2016. LNCS, vol 10058*. Springer, Cham, pp. 50-62.
- Verykokou, S., Doulamis, A., Athanasiou, G., Ioannidis, C., Amditis, A., 2016b. UAV-based 3D modelling of disaster scenes for urban search and rescue. In: *IEEE IST 2016*, pp. 106-111.
- Verykokou, S., Ioannidis, C., 2016. Automatic rough georeferencing of multiview oblique and vertical aerial image datasets of urban scenes. *The Photogrammetric Record* 31 (155), pp. 281-303.
- Verykokou, S., Ioannidis, C., Athanasiou, G., Doulamis, N., Amditis, A., 2018. 3D reconstruction of disaster scenes for urban search and rescue. *Multimedia Tools and Applications* 77, pp. 9691–9717.
- Wilson, K., Snavely, N., 2014. Robust Global Translation with 1DSfM. In: Fleet, D. et al. (Eds.), *Computer Vision - ECCV 2014. ECCV 2014. LNCS, vol 8691*. Springer, pp. 61–75.
- Wu, C., 2013. Towards linear-time incremental structure from motion. In: *3DV2013*, pp. 127-134, doi: 10.1109/3DV.2013.25.