REAL-TIME 3D RECONSTRUCTION FROM IMAGES TAKEN FROM AN UAV

A. Zingoni [a], M. Diani [a], G. Corsini [a], A. Masini [b]

[a] Dept. of Information Engineering, University of Pisa, Via G. Caruso 16, 56122 Pisa, Italy - andrea.zingoni@for.unipi.it, (m.diani, g.corsini)@iet.unipi.it
[b] Remote Sensing Department, Flyby s.r.l, Via Puini 97, 57128 Livorno, Italy - andrea.masini@flyby.it

KEY WORDS: 3D model, Dense Reconstruction, Real-time, Remote sensing, Surveillance systems, Unmanned aerial vehicle.

ABSTRACT:

We designed a method for creating 3D models of objects and areas from two aerial images acquired from an UAV. The models are generated automatically and in real-time, and consist in dense and true-colour reconstructions of the considered areas, which give the impression to the operator to be physically present within the scene. The proposed method only needs a cheap compact camera, mounted on a small UAV. No additional instrumentation is necessary, so that the costs are very limited. The method consists of two main parts: the design of the acquisition system and the 3D reconstruction algorithm. In the first part, the choices for the acquisition geometry and for the camera parameters are optimized, in order to yield the best performance. In the second part, a reconstruction algorithm extracts the 3D model from the two acquired images, maximizing the accuracy under the real-time constraint. A test was performed in monitoring a construction yard, obtaining very promising results. Highly realistic and easy-to-interpret 3D models of objects and areas of interest were produced in less than one second, with an accuracy of about 0.5m. For its characteristics, the designed method is suitable for video-surveillance, remote sensing and monitoring, especially in those applications that require intuitive and reliable information quickly, as disasters monitoring, search and rescue and area surveillance.

1. INTRODUCTION

In the last few years, the use of unmanned aerial vehicles (UAVs) has increased exponentially, involving a growing number of applications in different areas (Sauerbier 2010), (Manyoky, 2011), (Kanistras, 2014). One of the key factors of the success of UAVs lies in the possibility of equipping them with imaging sensors (Vasterling, 2013). This makes UAVs particularly effective in surveillance, monitoring and remote sensing applications, because of their capability of exploring areas inaccessible or dangerous for a human being (Adams, 2011).

Among the numerous applications, 3D reconstruction plays a crucial role, since it is capable to provide the spatial distribution of the information we are interested in. In particular, realistic reconstructions of wide areas, regions of interest, buildings, objects etc. have revealed to be a fundamental instrument of analysis and understanding of the phenomenon in exam.

Many techniques have been proposed in the literature for generating 3D models automatically by using an UAV properly equipped. Some of them are aimed to achieve very high accuracies, without (or partially) taking into account the aspect of the costs. In such techniques, LiDARs (Bisheng, 2015), laserscanners (Jutzi, 2013) or combinations of more different sensors (Masahiko, 2008) are generally used to improve the accuracy of the reconstruction. Also cameras can be successfully used for 3D reconstruction, permitting to limit the costs. However, often, for obtaining better accuracies, the aspect of computational time is neglected. For example, that is the case of reconstruction from multiple images (Mayer, 2008), or by using computationally costly algorithms (Huei-Hung, 2011), which can take minutes for generating the desired 3D model. In order to achieve good precision without sacrificing the computational time, many techniques rely on the help of auxiliary instrumentation (Wefelscheid, 2011), like inertial navigation systems (INSs) and global positioning systems (GPSs). However, these instruments can be quite expensive if they are required to be highly accurate. Some other techniques, instead, exploit some *a-priori* information like, for example, the

intrinsic parameters of the camera (Stephen, 2009) or the entire acquisition geometry, in case of UAV equipped with stereo cameras (Haubeck, 2013). In practical cases, this information is often unavailable, so that such techniques are not suitable for some contexts.

Unlike the approaches presented in the analysed literature, we tackled the problem of 3D reconstruction using an UAV, from the point of view of both computational time and costs, trying to achieve the best accuracy under these constraints. Thus, we designed a method for automatically extracting realistic 3D models of objects and regions of interest in real-time, using a very cheap instrumentation. Only a single cheap compact camera is mounted on the UAV, and no additional instrumentation (INSs, GPs etc.) is needed to achieve reconstruction. In addition, the *a-priori* knowledge of the intrinsic parameters is not strictly required.

The designing of the method takes into account the whole process of creating the 3D model, from the choice of the flight and camera setup, to the presentation of the final output. Two main parts can be recognized within our method. The first one is a preliminary phase, which aims to correctly set the camera parameters and define the acquisition geometry, in order to acquire the best images for maximizing the global performances. The second one is the 3D reconstruction algorithm, which permits to extract the model of the scene in real-time, by using a couple of images of it and no additional auxiliary data. All the steps of the proposed algorithm were selected among the most common routines for 3D reconstruction present in literature, in order to achieve the best compromise between accuracy and computational time.

The generated 3D model is a dense point-to-point, true-colour map of the scene, which gives the impression to be physically present within the reconstructed area. In addition, the model can be presented to the operator by exploiting whichever one of the main 3D visualization software, so that the point of view of the scene can easily be changed according to the specific needs.

The designed 3D reconstruction method is suitable for a large number of fields, like urban planning, video surveillance, search and rescue, natural disasters monitoring, archaeology, etc.

Its effectiveness was tested in a case of human work monitoring, and in a case of object recognition, within a construction yard. Very realistic and easy-to interpret 3D reconstructions were produced in real-time, demonstrating the efficiency of the proposed method.

2. 3D RECONSTRUCTION METHOD

The 3D reconstruction method we propose, is based on the triangulation of the objects/areas we are interested in reconstructing, starting from two images, acquired from different points of view. The images are taken by a camera, mounted on an UAV and guided over the region of interest (ROI). No additional equipment is necessary. The camera should be adjustable in focal length (f), aperture (a), exposure time (t_{exp}) and ISO. Nowadays, also cheap compact cameras fit these characteristics, so that an expensive imager is not needed. In addition, the limited weight of compact cameras and the absence of other instrumentation make small and cheap UAV perfectly suitable for our purpose.

The geometry of the problem is shown in Figure 1, where h represents the flight altitude, f the focal length of the camera and b the baseline, i.e. the distance covered by the camera focus between the two acquisition instants, obtained through:

$$b = V_{UAV} \cdot r_{fr} \qquad (1)$$

where $\qquad V_{UAV}$ = UAV velocity.
$\qquad\qquad\quad r_{fr}$ = camera frame rate.

Note that V_{UAV} can be assumed constant between the two acquisitions, since it is reasonable that the UAV does not change its velocity/trajectory significantly in the frame interval.
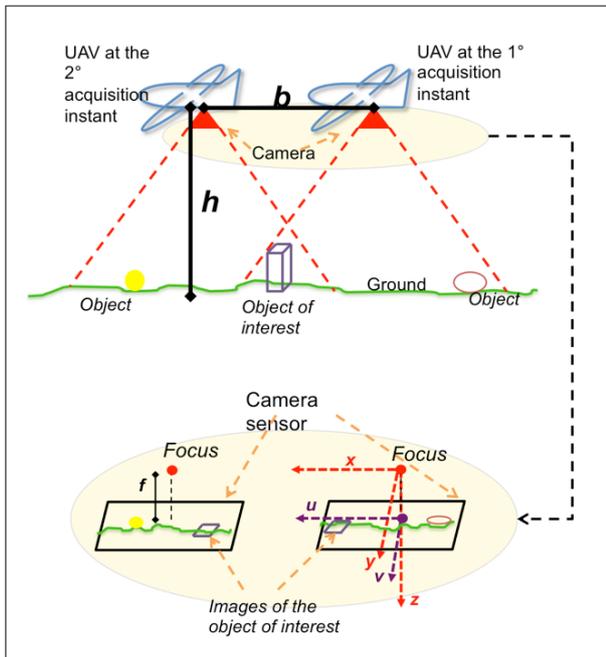
Figure 1. Geometry of the problem.

We define a 3D reference system for the points in the real world, which coincides with the camera system at the first acquisition instant. Such a system, depicted in red in Figure 1, is centered in the camera focus and has the x- and the y-axis directed along the largest and the shortest dimension of the camera sensor, respectively, and the z-axis coinciding with the optical axis, following the right-hand rule. The coordinates of the reconstructed 3D points will be relative to this reference system. We also define a 2D reference system (u,v) for the points in the first (second) acquired image, which we refer to as *image system 1 (2)*. It is centered in the projection of the focus on the image plane and has the u-axis and v-axis directed as the camera x-axis and y-axis of the camera system, as depicted in Figure 1. Hereinafter, we maintain this notation for the reference systems.

The proposed method can be divided in two parts: the design of the acquisition system and the 3D reconstruction algorithm.

2.1 Design of the acquisition system

Camera and acquisition geometry parameters influence the efficiency and fix the limitations of the 3D reconstruction methodologies. Thus, correctly setting them is a fundamental issue. Obviously, the correct choice strictly depends on the application, on the type of aircraft used, and on the available camera. We use some performance indicators, typical of 3D reconstruction, to choose the setting that maximizes the efficiency of the proposed method and to spot the main limitations.

A list of the considered performance indicators, with the respective parameters involved in their calculation, is presented in Table 1. Note that computational time depends more on the reconstruction algorithm than on the camera setting, thus it is not treated in this section.

Performance Indicator	Acronym	Parameters
Ground resolution	G_{res}	Flight altitude Sensor resolution Focal length Exposure time Aperture
3D model resolution	$3D_{res}$	Flight altitude Focal length Frame rate UAV velocity Disparity resolution
Maximum reconstructable area	A_{max}	Flight altitude Sensor size Focal length Frame rate UAV velocity
Maximum reconstructable height	H_{max}	Flight altitude Sensor size Focal length Frame rate UAV velocity
Computational time	T_{com}	Sensor size Number of images used

Table 1. Performance indicators considered in 3D reconstruction and parameters involved in their calculation

Ground resolution fixes the minimum recognizable area at the lowest point in the scene. Naming Δx and Δy the G_{res} components along the largest and the shortest dimension of the camera sensor, respectively, they can be expressed in formula as:

$$\Delta x = \frac{h}{f} \cdot Du \quad , \quad \Delta y = \frac{h}{f} \cdot Dv \qquad (2)$$

where f = focal length
h = flight altitude (with respect to the lowest point in the scene)
Du, Dv = horizontal and vertical dimensions of a single photodetector

It is worth noting that the actual G_{res} can be worse than the one calculated with (2), since the acquired images can be blurred. Two sources of blurring are present. The first one is the motion blur, generated by the movement of the camera during the exposure time and quantifiable as:

$$G_{res}^{mb} = V_{UAV} \cdot t_{exp} \qquad (3)$$

where: t_{exp} = exposure time.

The second one is the optical blur, generated by the aperture of the diaphragm of the camera. Such quantities must be added to G_{res} for determining the ground resolution appropriately.
3D model resolution, instead, defines the minimum reconstructable distance in the three dimensions. It means that two points, which are distant more than $3D_{res}$ from each other, are mapped in the same point in the extracted 3D model. $3D_{res}$ is defined for each spatial coordinate (Δx_{3D}, Δy_{3D} and Δz_{3D}) as:

$$\begin{cases} \Delta x_{3D} = \dfrac{u \cdot (h - H_{obj})^2}{b \cdot f^2} \cdot \Delta d \\[2ex] \Delta y_{3D} = \dfrac{v \cdot (h - H_{obj})^2}{b \cdot f^2} \cdot \Delta d \\[2ex] \Delta z_{3D} = \dfrac{(h - H_{obj})^2}{b \cdot f} \cdot \Delta d \end{cases} \qquad (4)$$

where u,v = pixel coordinates in the image chosen as reference.
H_{obj} = object height from the ground.
Δd = disparity resolution (explained in Section 2.2).
b = baseline.

3D model resolution should not be confused with ground resolution, since, for example, two points that are resolved in terms of G_{res}, but whose actual distance is less than $3D_{res}$, will still be considered as two different points in the extracted 3D model, but they will be placed in the same position, so that they will be no longer resolvable.
The maximum reconstructable area corresponds to the overlap area of the two images, expressed by the formula:

$$A_{max} = \left(\frac{U \cdot h}{f} - b \right) \cdot \frac{V \cdot h}{f} \qquad (5)$$

where: U, V = largest and shortest dimension of the camera sensor.

Finally, the maximum reconstructable height corresponds to the highest distance from the ground at which the two images overlap and can be calculated as:

$$H_{max} = h - \frac{f \cdot b}{U} \qquad (6)$$

Note that in (5) and (6) we are implicitly making some assumptions about the acquisition geometry, which will be presented below.
In order to maximize the efficiency of the proposed method, we preliminarily set those parameters that permit us to improve some indicators, without affecting the others. First of all, we mount the camera on the UAV so that it looks at nadir. Such a choice not only simplifies the geometry of the problem considerably, but also ensures that the ground samples size, within the area of interest, is the same in the two acquired images. This improves both the average G_{res} and the accuracy of the reconstruction algorithm, since this is based on the recognition of similar features in both images, as we will see in Section 2.2. In addition, we orient the camera so that the largest dimension of the sensor is parallel to the UAV flight direction. This simple expedient permits one to maximize either A_{max} or H_{max} without modifying any parameter.
Then, once observed the scene illumination, determined the height range of the objects to reconstruct, and decided the flight altitude, we reduce the optical blur by fixing a large depth of field. This can be done independently from f and h, by decreasing a, even if it results in darker images, a problematic factor in presence of poor illumination. However, this problem can be overcome by increasing t_{exp} or ISO. The first choice is not advisable, since it also increases motion blur -as it is evident from (3)-, worsening either G_{res} or $3D_{res}$. The second choice, instead, only produces some noise, which can be filtered in the reconstruction algorithm. Thus, increasing ISO is the best choice to maintain a good illumination while decreasing a.
The main limitations in 3D reconstruction are due to the fact that, if f increases and h decreases, G_{res} and $3D_{res}$ improve, while H_{max} and A_{max} decrease. Moreover, $3D_{res}$ also improves when b increases, while H_{max} and A_{max} worsen. Since a high A_{max} is not necessary for reconstructing objects or portions of the entire scene, while good resolutions are desirable, higher values of f and b and lower values of h should be set. In particular, h imposes the more stringent constraint, since each coordinate of $3D_{res}$ depends from it as a quadratic function. However, we cannot increase f and b and decrease h freely. In fact, a low A_{max} could result in the total or partial lack of overlap between the ROIs in the two images, making a complete reconstruction impossible. At the same time, a low H_{max} could not permit the reconstruction of the highest objects.
Moreover, some physical limitations occur in the choice of these parameters. In particular, b is constrained by the camera frame rate and by the UAV velocity (1), while h is prone to the aerodynamic laws.
Summarizing the above-mentioned considerations, we suggest a robust empirical method for setting f, b and h correctly. First, we choose the lowest permitted h, to obtain a sufficient A_{max} and to equals H_{max} to the maximum height we are interested in reconstructing. Then, we acquire the two images with the highest b, respecting the A_{max} constraint and according to the limitations imposed by V_{UAV} and r_{fr}. Finally, we calculate the highest value of f that results in the A_{max} and H_{max} nearest to the fixed ones. This automatically maximizes G_{res} and $3D_{res}$.

2.2 3D Reconstruction Algorithm

The 3D reconstruction algorithm we implemented is capable to create a 3D model from two images, acquired as specified in the previous section, without knowing the position and the orientation of the camera at the images acquisition instants (extrinsic parameters). It is based on the normalized 8-point algorithm [1], because this approach is a good compromise between accuracy and computational efficiency, in the absence

of the extrinsic parameters. The block diagram of the algorithm is shown in Figure 2.
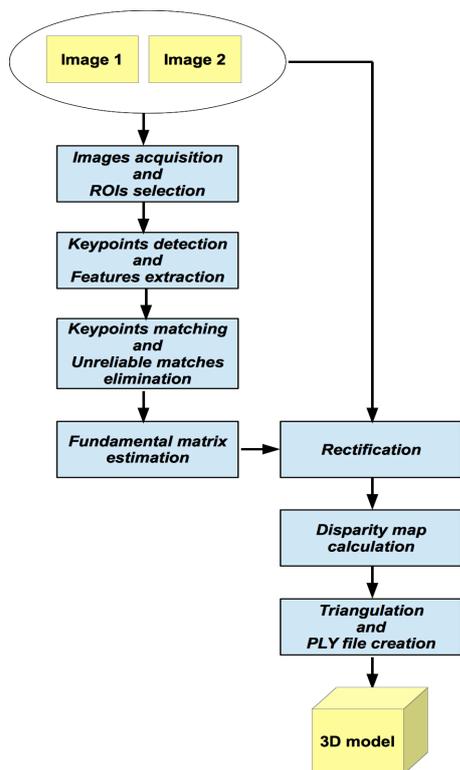
Figure 2. Block diagram of the proposed 3D reconstruction algorithm

After acquiring the couple of frames, the operator selects in both images the object or the area we are interested in reconstructing. A portion of each image, centred on its respective object/area of interest is isolated (we will refer to them as P_1 and P_2, respectively). P_1 and P_2 should be rich of easily detectable features (like corner and edges), thus, their size can be small in case of highly-textured images, but it has to increase in case of poorly-textured images.

Within P_1 and P_2, the points richest in features (*keypoints*, KPs) are detected, using the scale-invariant feature transform (SIFT) algorithm [2]. Then, the features of every KP are extracted and parameterized in form of a vector, again through the SIFT algorithm. Even if other typical feature extraction algorithms are computationally less expensive, like, for example, the speeded up robust features (SURF) (Bay, 2006), we preferred to use a slightly slower but more accurate algorithm. This is because the errors committed in detecting the KPs and extracting their features weight considerably on the successive step, i.e. the KPs matching, which, in turn, directly affects the computation of the fundamental matrix (Prince, 2012), which is the pivotal operation of the entire reconstruction algorithm. In addition, if we rely more on the extracted features, we can relax the point matching refinement phase -which can be quite expensive in terms of computational time- anyway achieving the required precision.

Subsequently, the analogous KPs in the two image portions are matched, by searching, for each KP features vector of the first image, the nearest KP features vector of the second image. For the motivation explained above, we again sacrificed the computational cost for the sake of accuracy. In fact, a slower but exhaustive search of the best matches was chosen, instead of

a faster but approximate search, like the one performed by the algorithms of the fast library for approximated nearest neighbors (FLANN) (Muja, 2013), commonly used in this kind of application. Then, the matched KPs whose feature vectors are more distant than a certain threshold are removed. The threshold is set between 0.9 and 1.5 times the average distance over all the couples of analogous KPs features vectors, depending on the total number of matched KPs. This step permits us not only to refine the found matches, but also to reduce the computational cost of the subsequent step, i.e. the outliers removal, since it will be performed on less couples of KPs. To remove outliers, the randomized RANSAC with sequential probability ratio test (R-RANSAC SPRT) algorithm (Matas, 2005) is used, since it is one of the speediest (Sunglok, 2009).

With the remaining matched KPs as input, the normalized 8-point algorithm is run. It gives as output an estimation of the fundamental matrix (F). It is worth noting that we have neglected the effect of lens distortion, since this phenomenon is very limited in compact cameras. This permits one to considerably simplify the calculation of F.

The estimated value of F is used to rectify the two images, i.e. to transform them, through a projective homography, so that the rows of the first image are aligned with the analogous rows of the second one. Note that rectification can be precise (up to a scale factor) only if the photodetectors size and the offset between the optical axis and the image centre, which compose the camera intrinsic parameters together with the focal length, are known. Conversely, rectification is precise up to a projective transformation. To limit computational time, rectification is performed only for the area of interest.

On the two rectified sub-images, the disparity map is calculated. With the word *disparity* we refer to the horizontal displacement between the analogous points of the two considered sub-images -note that no displacement is present along the vertical dimension, after rectification. The first step for calculating the disparity map is to match the analogous points. Since we are interested in a realistic reconstruction, this step must be performed for all the points of the two sub-images, to avoid "holes" in the 3D model and, thus, to obtain a dense reconstruction. The point matching is executed by means of the semi-global block-matching algorithm (SGBM) (Hirschmuller, 2007), which has proven to be one of most reliable solutions, among the real-time algorithms capable to provide dense disparity maps (Scharstein, 2002).

Finally, the 3D reconstruction of the area of interest is obtained, from the disparity map, by triangulation:

$$x = \frac{b \cdot u}{d} \quad , \quad y = \frac{b \cdot v}{d} \quad , \quad z = \frac{b \cdot f}{d} \qquad (7)$$

where   u, v = 2D coordinates of the point to be reconstructed, expressed in the image system 1.

          x, y, z = coordinates of the reconstructed point, expressed in the chosen 3D reference system.

The algorithm also creates a PLY file of the 3D model. This format permits us to associate the coordinates of every reconstructed point to its colour intensity values (RGB). This makes the 3D model more realistic. The PLY format can be read by all the most common 3D visualization software, so that the 3D model can be easily presented and used by the operator.

3. RESULTS

In order to illustrate the potential of the proposed method, we report an example of application in the monitoring of a

construction yard. The aim of the activity was to extract the 3D model of some areas to monitor the progress of the works, and the 3D model of some objects present within the yard, to recognize them.

The camera we mounted on the UAV is a Canon IXUS 220HS, a cheap compact camera with 4000x3000 pixel resolution, that permits to adjust the main acquisition parameters.

To evaluate the performance of the method in a critical case, we chose a high value of h (120m). Since we did not know exactly where the objects of interest were, we desired an A_{max} of about 75%. b is constrained to about 10m from the minimum UAV velocity (about 35m/s) and from the frame rate, which is not adjustable and equal to 3.4Hz. Once b and h were fixed, we chose the highest f compliant with the required A_{max}, i.e. capable to ensure the 75% of overlap between the two acquired images. Considering that the camera sensor is a CMOS 1/2.3, i.e. it is 1/2.3 inches large on the diagonal, it is easy to find that $U \approx 6.6mm$. Therefore, from Equation (5) we found $f \approx 20mm$, which is less than the maximum focal length of the camera and, thus, it could be chosen. No problems arise from the choice of H_{max}, since, for $h = 120m$, all the objects in the overlapped area can be entirely reconstructed. a and t_{exp} were set to the minimum (F# = f/5.9 and 1/2000s, respectively) for limiting optical and motion blur, while the ISO was set to 800 to enhance the illumination.

3.1 Qualitative analysis of the results

In Figure 3(a) and 3(b), two images of an area of interest are shown. Note that it is impossible to determine which are the highest/lowest parts of the scene, and, thus, to evaluate the advances in excavations. We gave them as input to the reconstruction algorithm, obtaining the disparity map, shown in Figure 3(c), where higher values tend to darker shades of red and lower values to darker shades of blue. The occlusions, i.e. those points for which the disparity cannot be calculated because they are covered in one of the two images, are represented in black. From the disparity map, the algorithm extracted the 3D model and stored it in a PLY file. In Figure 4 the model is presented by using a 3D visualization software. The 3D reconstruction is very realistic, allowing the operator to evaluate the state of progress of the work rapidly and clearly, as if he were physically present within the yard.

(a)   (b)

(c)
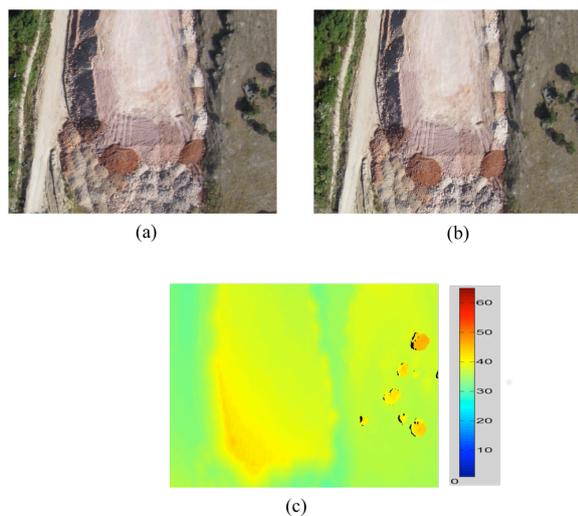
Figure 3. First (a) and second (b) image of the area of interest, and disparity map (c) -lower values tend to dark blue, higher values to dark red and occlusions are depicted in black
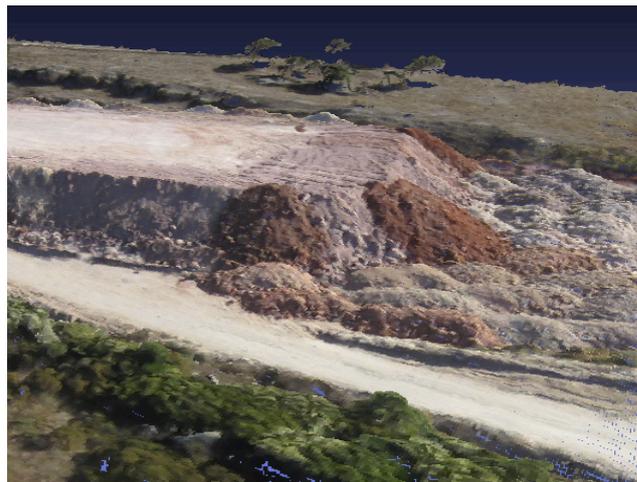
Figure 4. 3D model of the area of interest

In Figure 5(a) and 5(b) two isolated portions of a couple of images of the yard, containing an object of interest, are shown. We ran the reconstruction algorithm on the two image portions, to obtain more information about such an object. The disparity map of the examined portion of image is shown in Figure 5(c), while the 3D model is shown in Figure 6. The level of understanding is highly enhanced and the object is easily recognizable as an excavator, from the wide cockpit and from the mechanical arm. Also the position of the object, on the edge of a dig, stands out clearly.

(a)   (b)

(c)
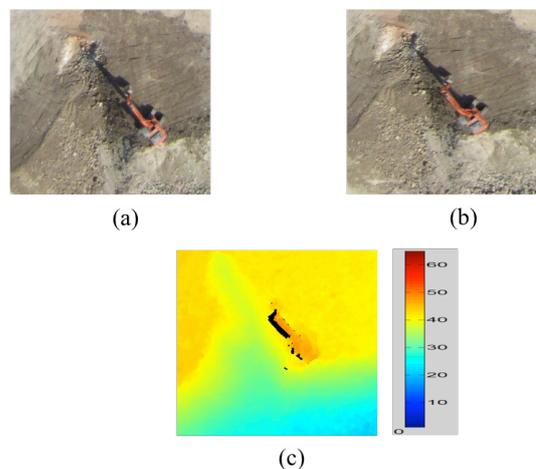
Figure 5. First (a) and second (b) image of the object of interest, and disparity map (c) -lower values tend to dark blue, higher values to dark red, and occlusions are depicted in black

3.2 3D reconstruction accuracy

The absence of additional instrumentation on board the UAV and, in particular, of a GPS, did not permit us to exhaustively evaluate the accuracy of the extracted 3D model, as the coordinates of the 3D reconstructed points are expressed in the camera reference system, but the position of the UAV was not known. However, we knew the flight altitude, so that we were able to calculate the error committed in calculating the z-coordinate, which is also the error committed in determining the altitude.

Hence, we computed the average absolute error on the altitude (E_A), over a group of control points at ground level, whose actual altitude is zero. We obtained $E_A = 0.56m$. Such a result exceeds almost 5 times the theoretical resolution $\Delta z_{3D} = 0.12m$, calculated from (4) mostly because of the errors in measuring the baseline and the flight altitude. However, it is sufficient enough for the case in question. The accuracy can be further improved by reducing the above-mentioned errors.

Figure 6. 3D model of the object to recognize

3.3 Computational time

Finally, we evaluated if the real-time requirement is satisfied. In critical scenarios, an operator should have the 3D model available almost instantaneously, after selecting the area of interest. Thus, we decided to tolerate a delay of at most 1sec in order to affirm that the algorithm is working (nearly) in real-time.

Obviously, the computational time increases with the size of the area of interest. Thus, a correct evaluation should aim to determine for which sizes of the area of interest the real-time requirement is achieved. For this purpose, we ran the reconstruction algorithm several times, for different sizes of P_1 and P_2. The machine on which we performed the evaluation is a PC, equipped with an Intel Core i7 CPU at 2GHz and 8GB RAM, which use Windows 7 at 64bit as operative system.

We observed that, for areas of at most 1000x750 pixels, the 3D model was produced in less than 1sec, so that we can consider the real-time performance achieved. It is worth noting that, with the chosen acquisition geometry parameters, a ROI of 1000x750 pixels corresponds to an area of about 10x7.5 meters at the ground level, which is large enough to focus on man-made objects or small specific areas.

4. DISCUSSION

The obtained results have proven that the proposed method is very effective, both in terms of reliability and interpretability of the produced 3D models, and in terms of computational time. Nevertheless, some inaccuracy factors are present. One of these factors is the presence of some occlusions, which results in "holes" within the 3D model. Occlusions can be avoided directly, by reconstructing the scene from more than two images, or indirectly, by interpolation. Unfortunately, both methods are computationally demanding and, thus, not suitable for real-time applications. However, when the flight altitude is quite high, like in the case we examined, the phenomenon of occlusion is not so significant, since the two images are less prone to parallax, so that the amount of "holes" in the 3D model is low.

Another inaccuracy factor is the not perfect flatness of low-texturized flat surfaces of the scene, which is due to mismatches in the disparity map calculation, generated by the SGBM algorithm. More reliable algorithms exist [4], but the larger amount of computational time they require does not compensate the improvement in matching precision.

The achievement of real-time performance, jointly with the good accuracy of the 3D model, demonstrates that the choices made in implementing the reconstruction algorithm are appropriate. Thus, the approach of stressing the procedures for features extraction and matching of the KPs in terms of accuracy, while relaxing the outliers elimination and the disparity map calculation, can be used as a guideline when 3D reconstruction has to be performed accurately but under the real-time constraint.

5. CONCLUSION

A method for creating realistic 3D models from aerial images, acquired from an UAV, was presented. It is capable to produce a dense and true-colour reconstruction of an object or a region of interest, from a couple of images, automatically and in real-time. In addition, the method is suitable to work with very cheap instrumentation, consisting uniquely of a compact camera, and without knowing the intrinsic parameters *a-priori*.

The method involves the entire process that leads to the creation of the 3D model, so that every factor that can improve the performance is adjusted. In fact, either the design of the acquisition system or the 3D reconstruction algorithm are taking into account. The former to acquire the best couple of images from which generating the 3D model, the latter to output it with the maximum accuracy achievable under the real-time constraint.

The effectiveness of the method was illustrated by testing it on an experimental data set acquired over an area interested by a construction yard. The produced 3D reconstructions of both objects and areas of interest are very realistic and permitted us to recover additional information, not directly obtainable from the 2D images. In addition, the real-time requirement is satisfied, since 3D models of quite large areas can be generated in less than one second.

Its characteristics make the designed method particularly suitable for remote sensing and video-surveillance applications, especially in those contexts where easy-to-interpret models of areas, that are not directly explorable by human being, are needed in a short time, as, for example, in natural disaster monitoring, in search and rescue and in automatic area surveillance.

Future studies will aim to solve the problem of occlusions, by implementing a reliable real-time reconstruction algorithm based on more than two images, and to further improve the accuracy of the 3D model.

6. ACKNOWLEDGEMENTS

REFERENCES

Adams, S. M., Friedland, C. J., 2011. A Survey of Unmanned Aerial Vehicle (UAV) Usage for Imagery Collection in Disaster Research and Management. *Proceedings of the 9th*

International Workshop on Remote Sensing for Disaster Management, Stanford University, Stanford, CA, USA.

Bay, H., Tuytelaars, T., Van Gool, L., 2006. SURF: Speeded Up Robust Feature. *Proceedings of the 9ᵗʰ European Conference on Computer Vision*, Graz, Austria, Volume 3951, pp. 404-417.

Bisheng, Y., Chi, C., Automatic registration of UAV-borne sequent images and LiDAR data. *ISPRS Journal of Photogrammetry and Remote Sensing*, Volume 101, pp. 262-274.

Hartley, R. I., 1997. In defence of eight-point algorithm. *IEEE Transaction on Pattern Recognition and Machine Intelligence*, 19 (6), pp. 580-593.

Haubeck, K., Prinz, T., 2013. A uav-based low-cost stereo camera system for archaeological surveys - experiences from Doliche (Turkey). *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Volume XL-1/W2, pp. 195-200.

Hirschmuller, H., 2007. Stereo processing by semiglobal matching and mutual information. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 30 (2), pp. 328-341.

Huei-Hung, L., Yuping, L., Medioni, G., 2011. Aerial 3D reconstruction with line-constrained dynamic programming. *Proceedings of the 2011 IEEE International Conference on Computer Vision (ICCV)*, Barcelona, Spain, pp. 1855-1862.

Kanistras, K., et al., 2014. Survey of Unmanned Aerial Vehicles (UAVs) for Traffic Monitoring. *Handbook of Unmanned Aerial Vehicles*, pp. 2643-2666.

Lowe, D. G., 1999. Object recognition from local scale-invariant features. *Proceedings of the 1999 IEEE International Conference on Computer Vision (ICCV)*, Corfu, Greece, pp. 1150-1157.

Manyoky, M., et al., 2011. Unmanned Aerial Vehicle in Cadastral Applications. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Volume XXXVIII-1/C22.

Masahiko, N., et al., 2008. UAV Borne Mapping by Multi Sensor Integration. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Volume XXXVII, Part B1, pp. 1215-1222.

Matas, J., Chum, O., 2005. Randomized RANSAC with Sequential Probability Ratio Test. *Proceedings of the 10ᵗʰ IEEE International Conference on Computer Vision (ICCV)*, Beijing, China, Volume 2, pp. 1727-1732.

Mayer, H., Bartelsen, J., 2008. Automated 3D Reconstruction of Urban Areas from Networks of Wide-Baseline Image Sequences. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Volume XXXVII, Part B5, pp. 633-638.

Muja, M., Lowe, D. G., 2013. FLANN - Fast Library for Approximate Nearest Neighbors. www.cs.ubs.ca/research/flann/

Prince, S. J. D., 2012. Computer Vision. Models, Learning and Inference. Section 16.3, pp. 361-363.

Sauerbier, M., Eisenbeiss, H., 2010. UAVs for the documentation of archaeological excavations. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Volume 38(5).

Scharstein, D., Szeliski, R., 2002. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision*, 47 (1-3), pp. 7-42.

Stephen, S., et al., 2009. Automated UAV-based mapping for airborne reconnaissance and video exploitation. *Proceedings of SPIE 7307, Airborne Intelligence, Surveillance, Reconnaissance (ISR) Systems and Applications VI*, Orlando, FL, USA, Volume 7307.

Sunglok, C., Taemin, K., Wonpil, Y., 2009. Performance Evaluation of RANSAC Family. *Proceedings of 2009 British Machine Vision Conference (BMVC)*, London, UK.

Triggs, B., 1999. Bundle Adjustment - A Modern Synthesis. *Proceedings of the International Workshop on Vision Algorithms (ICCV '99): Theory and Practice*, Corfu, Greece, pp. 298-372.

Vasterling, M., Meyer, U., 2013. Challenges and Opportunities for UAV-Borne Thermal Imaging. *Remote Sensing and Digital Image Processing*, Volume 17, pp. 69-92.

Wefelscheid, C., Hänsch, R., Hellwich, O., 2011. Three-dimensional building reconstruction using images obtained by unmanned aerial vehicles. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Volume XXXVIII-1/C22, pp. 1-6.