Integration of photogrammetric DSM and advanced image analysis for the classification of urban areas

Mauro Dalla Mura\textsuperscript{a}, Francesco Nex\textsuperscript{b}, Fabio Remondino\textsuperscript{b} and Michele Zanin\textsuperscript{b}

\textsuperscript{a} GIPSA-lab, Grenoble Institute of Technology, Saint Martin D’Hères, France
mauro.dalla-mura@gipsa-lab.grenoble-inp.fr

\textsuperscript{b} Bruno Kessler Foundation, Trento, Italy
\{franex, remondino, mizanin\}@fbk.eu

ABSTRACT

In this paper, a technique for the integration of images and point cloud for urban areas classification is presented. A set of aerial RGB overlapping images are used as input. A photogrammetric Digital Surface Model (DSM) is firstly generated by using advanced matching techniques. Subsequently, a thematic classification of the surveyed areas is performed considering simultaneously the surface’s reflectance in the visible spectrum of the image sequence, the altitude information (provided by the generated DSM) and additional spatial features (Attribute Profiles). Exploiting the geometrical constraints provided by the collinearity condition and the epipolar geometry between the images, the thematic classification of the land cover can be improved by considering simultaneously the height information and the reflectance values of the DSM. Examples and comments of the proposed classification algorithm are given using a set of aerial images over a dense urban area.

Keywords: Urban remote sensing, Photogrammetry, Multi-view images, Digital Surface Model, Very high resolution images, Classification

1. INTRODUCTION

It is well known that remote sensing images can be an extremely useful for the monitoring of the evolution of the Earth’s surface.\textsuperscript{1,2} With particular regard to urban areas, the analysis of images acquired from satellite or airborne platforms can provide insights on the urban development, land cover characterization and evolution (e.g., construction of new buildings, monitoring of the green spaces).\textsuperscript{3} The monitoring of urban areas can be effectively performed through a thematic classification of the scene. In this scenario, it is possible to derive information on the evolution of the scene, since to each image pixel is associated a thematic class of land cover. When dealing with overlapping images acquired over the same scene (e.g. a photogrammetric block), it is possible to exploit the redundant information contained in the images in order to improve the results and accuracy of scene’s reconstruction and classification algorithms. Photogrammetric aerial cameras or the new very high resolution satellite sensors (e.g. GeoEye, Worldview, etc.) are all capable to collect multiple high spatial resolution imagery with different viewing angles on a single strip. An example of overlapping images acquired over the same scene is reported in Figure 1(a)-(c). It is possible to notice how rooftops are visible in all the images whereas, due to the acquisition geometry and platform’s movement, the building facades are only visible on some acquisitions. Overlapping image sequences can be processed with photogrammetric methods in order to produce Digital Surface Models (DSM) (Figure 1(d)), correlating homologous points in the images and deriving 3D object coordinates.

Altitude data, regardless their origin (LiDAR acquisitions, photogrammetry, cadastral maps), have already proven to provide valuable information for the characterization of the surveyed scene in many applications such as the extraction of urban 3D models\textsuperscript{4} and classification.\textsuperscript{5} With particular regard to classification, the coupling of the height data with the spectral information of optical images has proven to be capable of accurately characterizing urban\textsuperscript{6–11} and forest areas.\textsuperscript{12}

In this paper, a new classification technique which exploits both the spectral information of overlapping images and the elevation information provided by photogrammetrically generated DSM is presented. The use of photogrammetric processing can give a double contribution to the classification of urban area. On one hand, it allows to define the correspondences between pixels on different images: in this way, a “redundant” classification
Figure 1: Example of multiview acquisitions (a-c) and (d) the corresponding DSM generated with photogrammetry.

of corresponding points in several images can be obtained, improving the characterization of the scene with respect to the use of only a single image. On the other hand, the classification based only on multispectral pixel values can be integrated with 3D coordinates information imposing an additional information to the classification process: i.e. pixel classified as roofs in images could not be located on the ground, and vice versa. This work focuses on the former aspect. The presented technique, consider also spatial features computed with Attribute Profiles applied to the elevation data (DSM) for including the description of the spatial characteristics in the classification process.

The paper is organized in five sections. In the next section the generation of the DSM with a photogrammetric approach is given while in Section 3 the proposed classification technique is reported. The experimental analysis is described in Section 4. Conclusion and future perspectives are finally drawn in Section 5.

2. PHOTOGRAMMETRIC DSM

2.1 Generation of a photogrammetric DSM

Photogrammetry is the technique to reconstruct the geometry of a surveyed object visible in two or more images. Figure 2 depicts the scenario of the acquisition of multiple images acquired over the same area.

The photogrammetric processing can be mainly divided in two steps: the image orientation and the DSM generation (see flowchart of the generic photogrammetric pipeline in Figure 3). In the image orientation phase, the camera calibration is a fundamental prerequisite for any metric reconstruction as it allows to compensate for any systematic error as well as for lens distortions. The process to simultaneously orient all the images in the same reference system is called Bundle Block Adjustment, e.g. a non-linear optimization procedure to minimize
an appropriate cost function.\textsuperscript{16,17} This process firstly needs the collection of several correspondences (tie points) between the different images, normally found using automated feature extraction algorithms, such as the SIFT operator.\textsuperscript{18} In order to retrieve geo-referenced results, a set of 3D points of known coordinates, called Ground Control Points (GCP), must be used. Once the images have been oriented, a DSM can be generated. Starting from the known camera orientation parameters, automated image matching techniques are able to extract dense point clouds which describe the object’s surface and its main geometric discontinuities. The point cloud density must be adaptively tuned to preserve scene edges and, possibly, avoid too many points in flat areas.

An example of a DSM generated over the area in Figure 4(a) is shown in Figure 4(b). Different possible representations (color-coded, shaded and point cloud) are reported (Figure 4).

### 2.2 Exploitation of the DSM for scene classification

The exploitation of the generated DSM in the classification process can not be simply done by considering the altitude information as an additional feature. In particular, there are some issues related to the errors due to the
Figure 5: Issues related to the use of the generated DSM as feature for scene classification: (a) Inconsistent altitude information due to errors in the generation of the DSM through photogrammetry; (b) Height similarity between different thematic objects; (c) Effect of the terrain slope (areas in color are above the altitude of 240m, the areas of inferior elevation are set to black).

generation of the DSM with photogrammetric techniques. Such errors are related to mismatches between points in images due for instance to occlusions (e.g., points on the ground can only be visible on a subset of images in the block sequence) and lack of feature points (e.g., when the matches between points have to be found in homogeneous areas, shaded zones, areas showing highly repetitive texture patterns). In Figure 5(a) it is possible to notice that the altitude in the DSM of the shadowed areas is similar to the one of the adjacent buildings while their actual height should be close to the one of the neighboring terrain. Furthermore, there are intrinsic issues related to the height information provided by the DSM (for classification purposes) which do not depend on the technique used to generate the DSM. Specifically, there are some ambiguities in considering the height as discriminant feature: objects belonging to different semantic classes might have the same height (e.g., trees and houses as shown in Figure 5(b)). Moreover, if the terrain slope is still present in the DSM data (or not
completely subtracted) the heights of objects supposed to be equally height can be significantly different. For example, Figure 5(c) shows in color the areas of the scene whose altitude is greater than an arbitrary threshold. It is possible to notice that some buildings are below the altitude of the bare ground in the bottom part of the image.

3. METHODOLOGY

This section details the steps for the generation of the features subsequently considered for classification. When aiming at the generation of a classification map of the scene, it is mandatory to define in which reference system are the pattern defined. In greater detail, a pattern could be either related to a position on the ground (i.e., the reference system of the DSM) or in an image plane (i.e., a pixel). In the former case, the height information provided by the DSM can be directly considered as a feature, whereas the spectral values of the pixels of the overlapping images (or any other information expressed in the reference system of the image planes) should be projected on the ground reference system. This operation is necessary in order to link each sensed values from the different viewing angles to the same positions on the ground. Subsequent to the projection on the ground reference system, the information in the image planes can be converted as additional features for the characterization of each pattern (i.e., in this case, a position on the ground). Conversely, in the latter case, the patterns refer to pixel in a single image plane. Thus, if an image plane is considered as the reference plane, the elevation information in the DSM and the spectral values of the other image planes should be projected in the reference plane in order to be used as features. This setting, might lead to some issues related to the fact that parts of objects might not be visible from the point of view corresponding to the image plane considered as reference system and that some vertical objects only appear from this view (e.g., some facades). Since, the aim of the land cover classification is to provide information about the semantic of the coverage on the ground we have kept the ground coordinates as reference system. The flowchart of the proposed method is presented in Figure 6 and its blocks are detailed in the following.

3.1 Projection on the DSM

Once the DSM is generated via image matching (as detailed in Sec. 2.1), there is a match between each point on the ground (i.e., on the DSM) and the pixels in the images in which that given area on the ground is visible from the respective viewing points from which the images were acquired. The correspondence between pixels of an arbitrary image \( F \) in the multiview sequence and points on the DSM is defined because the absolute position (world coordinates) and the attitude (orientation) in space of \( F \) have been determined. Thus, thanks to this information, image coordinates \((\xi, \eta)\), in the arbitrary image plane \( F \), can be linked to object coordinates (i.e., on the ground) \((X, Y, Z)\) according to the equations of collinearity 1, 2:

\[
\xi = \xi_0 - c \frac{r_{11}(X - X_0) + r_{21}(Y - Y_0) + r_{31}(Z - Z_0)}{r_{13}(X - X_0) + r_{23}(Y - Y_0) + r_{33}(Z - Z_0)} = \xi_0 - \frac{Z_x}{N} \tag{1}
\]

\[
\eta = \eta_0 - c \frac{r_{12}(X - X_0) + r_{22}(Y - Y_0) + r_{32}(Z - Z_0)}{r_{13}(X - X_0) + r_{23}(Y - Y_0) + r_{33}(Z - Z_0)} = \eta_0 - \frac{Z_y}{N} \tag{2}
\]

\((X_0, Y_0, Z_0)\) define the coordinates of the image perspective centre, \((\xi_0, \eta_0)\) coordinates of the principal point of \( F \), the \( r_{ij} \) \((i, j \in [1, 2, 3])\) elements of the rotation matrix define the orientation of \( F \), \( c \) is the focal length of \( F \).
The point coordinates on the space $X, Y, Z$ can be determined when its position in two or more images are known. The link between object coordinates $(X, Y, Z)$ and the image coordinates $(\xi, \eta)$ in $F$ can be expressed by the following equations:

$$
X = X_0 + (Z - Z_0) \frac{r_{21}(\xi - \xi_0) + r_{22}(\eta - \eta_0) - r_{13}c}{r_{31}(\xi - \xi_0) + r_{32}(\eta - \eta_0) - r_{33}c} \tag{3}
$$

$$
Y = Y_0 + (Z - Z_0) \frac{r_{21}(\xi - \xi_0) + r_{22}(\eta - \eta_0) - r_{23}c}{r_{31}(\xi - \xi_0) + r_{32}(\eta - \eta_0) - r_{33}c} \tag{4}
$$

It is recalled that the aforementioned equations are only valid for central perspective images (i.e., the imaging sensors used for the acquisition can be modeled as pinhole cameras).

We denote the image $F(\xi, \eta)$ projected in the ground reference system as $I(X, Y)$. Image $I$ can be considered as ortho-rectified since it is showing only the areas that can be viewed in a nadiral acquisition. Since not all the pixels in $I$ map to $F$, we define $\Omega(X, Y) \rightarrow [0, 1]$ as a binary image representing the visibility map of the image $F$ on $I$ (pixels in $I$ that are not visible in $F$ will have value zero in $\Omega$ otherwise they will be set to one). An example of the result of a projection of an image $F$ on the ground coordinates is shown in Figure 7.

![Figure 7: Example of projection on ground coordinates. (a) Original image $F(\xi, \eta)$; (b) Same image projected on the ground $I(X, Y)$; (c) Visibility map of $F$ on $I$, $\Omega(X, Y)$.](image)

### 3.2 Attribute Profiles

Since the height information itself might not be enough discriminant for the classification of urban areas or some parts of the DSM might show spurious valued of altitude, some spatial features have been also computed on the DSM. The objectives of this operation are two-fold: i) perform a spatial regularization of the height information by exploiting the correlation of values of neighboring pixels; ii) characterize differently objects in the DSM according to their geometrical and textural characteristics. We have used Attribute Profiles (APs)\(^\text{14}\) for carrying out such task. APs perform a multilevel decomposition of a scalar image $F : \mathbb{Z}^2 \rightarrow \mathbb{Z}$ by applying a sequence of attribute filters\(^\text{19}\) defined in the framework of mathematical morphology.\(^\text{20}\) Attribute filters are connected filters and hence, they only operate on the flat zones (i.e., regions of iso-intensity pixels connected according to a rule of connectivity) of an image. Such filters are based on the computation of attributes (i.e., measures) on the regions and they produce a filtering by either keeping or merging the flat zones to their surrounding ones according to if the computed measure fulfills a predicate. Two operators are considered in an AP: attribute thinning and attribute thickening. A thinning is an anti-extensive and idempotent transformation\(^\text{20}\) that, roughly speaking, operates on regions that are brighter than their surrounding ones. By duality, a thickening transformation is extensive and idempotent and processes dark regions. The attribute considered in the filtering process can be any measure that can be computed on the image regions. Typically, measure related to size (e.g., area, length), shape (e.g., elongation, compactness, rectangularity) and contrast (e.g., standard deviation, range of intensity) can be considered. In the following we recall the definition of the AP on an image $F$ based on an arbitrary attribute $\alpha$ and associated predicate $P$ (e.g., $P_\alpha = \alpha > \lambda_i$, being $\lambda_i$ a reference value for the i-th filtering):
\[ AP_\alpha(F) = \{ \phi^{P_\lambda L}(F), \phi^{P_\lambda L-1}(F), \ldots, \phi^{P_\lambda 1}(F), F, \gamma^{P_\lambda 1}(F), \ldots, \gamma^{P_\lambda L-1}(F), \gamma^{P_\lambda L}(F) \}, \]

with \( \phi \) and \( \gamma \) an attribute thickening and thinning, respectively.

In Figure 8, an example of an AP computed on a DSM considering the standard deviation of the height values in the regions as attribute is shown. It is possible to notice how by increasing the “aggressiveness” of the filter (i.e., considering a more strict predicate) more and more objects in the scene are removed.

![Figure 8: Example of an AP with standard deviation attribute computed on the DSM.](image)

4. EXPERIMENTAL ANALYSIS

4.1 Data set

The data set considered is composed of three aerial RGB 8-bit images (\( F_i \) with \( i = 1, 2, 3 \)) acquired over an area of Torino (Italy) covering about 0.5×0.5 km. The images were acquired with a DMC camera and they are characterized by a spatial resolution (Ground Sampling Distance, GSD) of 12 cm. The surveyed scene is significantly complex featuring several high buildings, trees, roads and variation of the terrain height. A spatial subset of 2200×1540 pixels (about 264×185 m\(^2\)) of the original images was considered in the experiments (Figure 9(a)-(b) represents the images once reprojected on the DSM, \( I_i \)). Only the pixels in the ground reference system being actually visible from the three views were considered. Thus, the set of considered pixels is given by considering the logical AND between the visibility maps corresponding to \( I_i \) (i.e., \( \Omega = \Omega_1 \cap \Omega_2 \cap \Omega_3 \)).

The DSM (Figure 9(d)) generated from the images has a spatial resolution of 12 cm (1 GSD). The image orientation was achieved using the Apero tool. The dense matching reconstruction was carried out with an optimal flow algorithm, implemented in the MicMac tool. Two attributes were considered in the computation of the APs based on the DSM: area and standard deviation.

Seven values of the thresholding parameter were considered for both the attributes:

- Area attribute: \( \lambda_{area} = \{100, 500, 1000, 2500, 5000, 10000, 25000\} \);
- Standard deviation attribute: \( \lambda_{std} = \{5, 10, 15, 20, 25, 30, 35\} \).

Three land cover classes were identified in the scene: Buildings, Vegetation and Roads. A training (with 1000 samples per class) and a test sets (with Buildings: 813.014, Vegetation: 34.628, Roads: 132.982 samples) were generated through photo-interpretation. The test image is shown in Figure 9(e).

The classification was performed with a Random Forest (RF) classifier defined with 100 trees and number of variables equal to the square root of the feature number.

The accuracy of the classification results was assessed by computing Overall Accuracy (OA), Average Accuracy (AA) and Kappa statistics (\( \kappa \)).

4.2 Design of Experiments

Several experiments were carried out considering different combination of features, which are listed in the following:

1. single RGB image (\( I_i \) with \( i = 1, 2, 3 \))
2. all the RGB images stacked together (\( I_{all} = \{I_1, I_2, I_3\} \))
Figure 9: Data set considered in the experiments. (a) I1; (b) I2; (c) I3; (d) DSM; (e) Test image. Thematic classes: * buildings, * vegetation, * roads * no data.

3. DSM
4. Iall + DSM
5. AP computed on the DSM with area attribute (APa)
6. AP computed on the DSM with standard deviation attribute (APS)
7. APs computed on the DSM (APall = {APa, APS})
8. stack of Iall + APall

4.3 Experimental Results
The quantitative results obtained are reported in Table 1. From their analysis it is possible to state that by considering the images stacked together without any information related to the altitude accuracy slightly worse than the best case achieved considering each image separately was obtained. Considering only the DSM performed similarly with respect to when classifying the optical data. The inclusion of spatial features computed with the APs led to an increase in terms of OA of about 5% and 7% with respect to the single DSM when considering area and standard deviation as attribute, respectively. The joint combination of the spectral and altitude information (Iall + DSM) improved the accuracies with respect to considering only the spectral and elevation separately (of about 2% and 3%, respectively). The best results were obtained when considering as features all the images, the DSM and the APs (Iall + APall).

<table>
<thead>
<tr>
<th>Feats.</th>
<th>I1</th>
<th>I2</th>
<th>I3</th>
<th>Iall</th>
<th>DSM</th>
<th>APa</th>
<th>APS</th>
<th>APall</th>
<th>Iall + DSM</th>
<th>Iall + APall</th>
</tr>
</thead>
<tbody>
<tr>
<td>OA</td>
<td>64.58</td>
<td>74.37</td>
<td>77.84</td>
<td>76.62</td>
<td>73.37</td>
<td>78.73</td>
<td>80.25</td>
<td>78.53</td>
<td>79.34</td>
<td>81.66</td>
</tr>
<tr>
<td>AA</td>
<td>67.04</td>
<td>70.35</td>
<td>72.29</td>
<td>75.49</td>
<td>62.15</td>
<td>67.57</td>
<td>75.41</td>
<td>71.70</td>
<td>76.67</td>
<td>73.55</td>
</tr>
<tr>
<td>κ</td>
<td>30.27</td>
<td>40.40</td>
<td>45.16</td>
<td>44.30</td>
<td>48.90</td>
<td>51.99</td>
<td>49.60</td>
<td>50.62</td>
<td>54.65</td>
<td></td>
</tr>
</tbody>
</table>

Table 1: Classification accuracies obtained by considering different combination of features.

The classification maps are shown in Figure 10. By analyzing the obtained maps one can notice that by only considering the spectral information the maps appear noisy since the classification of pixels belonging to the same object in the scene might be classified differently. Considering only the height as feature is not also providing precise result since, even if the regions appears more homogeneous, in the upper part of the scene, some parts of the same rooftops are mis-classified. The best results are obtained by combining the spectral values, the elevation and the spatial features provided by the APs.
5. CONCLUSION

In this paper, a new technique suitable for the classification of urban areas was presented. The technique exploits the redundant information available in overlapping images and the geometric information from the corresponding DSM. The classification is performed by considering simultaneously (i) spectral values of reflectance, derived by all the available overlapping images (projected on the ground through the collinearity equations), (ii) elevation data of the DSM (derived using an image matching algorithm) and (iii) spatial features computed with spatial filters (Attribute Profiles) applied to the elevation data.

From the preliminary results obtained it is possible to state that:

- the DSM provides complementary information to the spectral one but it should not be used directly as feature in classification;
- the use of Attribute Profiles computed on the DSM provides discriminant features for the classification;
- by projecting the images on the DSM it is possible to jointly exploit the different scenes;
- the joint use of spectral values and DSM information (along with features extracted from it) provided the best accuracies.

Future developments will consider additional features that can be derived from the DSM such as a map reporting on-ground and off-ground points. At the same time, it is planned to use classifiers able to handle missing values in order to perform a classification on all the points on the ground regardless their visibility in the images.
REFERENCES


