Retrieval of the Height of Buildings From WorldView-2 Multi- Angular Imagery Using Attribute Filters and Geometric Invariant Moments

Giorgio Licciardi, Alberto Villa, Mauro Dalla Mura, Lorenzo Bruzzone, Jocelyn Chanussot, Jon Atli Benediktsson

RETRIEVAL OF THE HEIGHT OF BUILDINGS FROM WORLDVIEW-2 MULTI-ANGULAR IMAGERY USING ATTRIBUTE FILTERS AND GEOMETRIC INVARIANT MOMENTS

G. A. Licciardi (IEEE Member)[1], A. Villa [2], M. Dalla Mura [3], L. Bruzzone (IEEE Fellow member)[4]
J. Chanussot (IEEE Fellow member)[1], J. A. Benediktsson (IEEE Fellow member)[5]

(1) GIPSA-Lab, Grenoble Institute of Technology, France
(2)Dipartimento Elettronica ed Informazione (DEI) - Politecnico di Milano - Aresys
(3) Technologies of Vision, Fondazione Bruno Kessler, Italy
(4) Department of Information Engineering and Computer Science, University of Trento, Italy
(5)Department of Electrical and Computer Engineering. University of Iceland, Reykjavik, Iceland
e-mail: Giorgio-Antonino.Licciardi@gipsa-lab.grenoble-inp.fr

ABSTRACT

This paper proposes a novel approach to the retrieval of buildings' height from multi-angular high spatial resolution images. To achieve this task, we combined two main concepts: multilevel morphological attribute filters, used for the definition of the objects in the image, and geometric invariant moments exploited for the characterization of the spatial properties of the previously detected shapes. The main concept of this study relies on the spatial properties of very high resolution images acquired from different angles of view. In particular, if we model the urban environment as an ensemble of vertical and horizontal surfaces, we can assume that the shapes related to the horizontal surfaces (i.e. the top of the buildings) do not suffer any relevant spatial distortion if detected from two angles of view, while vertical surfaces present strong changes in shape. Starting from this assumption, once each shape in each angular images has been spatially characterized, it is possible to identify univocally the same horizontal surface (i.e. the roof of a building) in each angular image. Finally, the knowledge of the acquisition angles permits the retrieval of the buildings height using simple trigonometric calculations. In this paper the proposed approach has been successfully applied to a WorldView-2 (WV2) very high resolution dataset composed by 5 angular images.

Index Terms— Morphological attribute filters, Invariant moments, Buildings height retrieval, Multiangular imaging

1. INTRODUCTION

Building height retrieval has become one of the hot topics in the remote sensing community. In particular most of the researches in this field are mainly focused on retrieving the height of the buildings from SAR images [1][2][3], or from the fusion of SAR and high resolution optical images [4]. Besides the use of SAR images, retrieving the height of the buildings from optical data is not a trivial task, because the height information can be extracted only by investigating the effects of the buildings on the surroundings, such as by measuring the shadows produced by the buildings themselves [5]. However this kind of approaches cannot be accurate because they do not take into account the relative height of the surrounding environment.

In general, before measuring the height of a building it is necessary to uniquely identify the building itself, separating it from the surrounding objects. This preprocessing step can be performed by analyzing the geometrical information contained in the image. In particular, one of the most important approaches to the analysis of very high resolution (VHR) remote sensing images

is based on the exploitation of the spatial information. The relevance of geometrical features increases in the analysis of urban areas, where they are usually more representative than the spectral information.

In literature, one of the most commonly used approaches to the extraction of the geometrical information from a remotely sensed image is based on the Morphological Profiles (MPs) and Attribute Profiles (APs) [6]. The use of morphological profiles is a common approach for the analysis of the spatial inter-pixel dependency, and has been successfully used to extract spatial information from high spatial resolution images. In particular in [7] the use of the MP has been applied to high resolution satellite images. The idea at the base of the MP is to apply geodesic closing/opening transformations of increasing size with the purpose to build a set of opening profiles (OPs) and closing profiles (CPs). The MPs are then obtained by grouping the two sets of OP and OC with the original image.

APs can be defined as a generalization of the MPs, providing a multilevel characterization of an image by the sequential application of morphological attribute filters (AF), that depending on the type of the attributes , can be used to model different kinds of structural information. In this way, AP are capable of extracting spatial features that can better model the spatial information, with respect to conventional morphological profiles. Filters based on the geodesic reconstruction (e.g. opening and closing by reconstruction), in contrast with standard morphological operators, first transform the input image by erasing those regions that do not enclose a set of pixels of a certain shape and size, called structuring element (SE); and after reconstruct the filtered image with an iterative process. The features extracted by the extended profiles were then considered in order to detect the shapes of the objects present in the image. In [8] different APs were used to model the spatial informations in satellite images, and produce better classification maps. Starting from these assumptions, it is possible to characterize the spatial properties of the shapes of the objects obtained using the APs, using the Hu set of invariant moments [9], which are invariant under translation, changes in scale, and also rotation.

To retrieve the heights of the buildings, in this paper we propose a novel method based on the application to multiangular images of the aforementioned approach. However, the combined use of APs and invariant moments, even if permits to identify and characterize the buildings in a image, do not provides any information on their heights. This information can be extrapolated by measuring the relative position of the rooftop of a building under different angles of view. Under different acquisition angles horizontal surfaces do not present great variations of the spatial properties, while, on the other hand, vertical surfaces show strong distortions even if detected under slightly different angles of view. Starting from this assumptions it is possible to match two objects, in two images acquired under different angles, that are related to the same horizontal surface. Once two objects are matched it is possible to retrieve the height of that object, using trigonometric calculations by measuring the relative shift of the two objects. Even if the combined use of attribute vectors with Hu invariant moments was firstly presented by Urbach et al. in [10], its use to do stereo matching is novel. Aim of this paper is to present a new instrument for the retrieval of building heights in urban environment.

The paper is organized as follows. Sections 2 and 3 present the AFs and the geometric invariant moments, respectively. Section 4 describes the proposed method while section 5 presents a practical application to the WorldView-2 dataset provided for the DFC2011 [1]. Finally, conclusions are drawn in Section 6.

2. ATTRIBUTE FILTERS

Let us consider a discrete scalar image f defined by a mapping function $f : E \to \mathbb{Z}$ with discrete domain $E \subset \mathbb{Z}^2$. Connected operators are filters that operate on a grayscale image f, by only merging its "flat zones" [11, 12]. We recall that a flat zone or connected component (CC) is a region of connected pixels (according to a connectivity rule, e.g., 4- or 8-connected [13]) of iso-intensity values. The results of a filtering with connected operators show a reduction of the complexity of the scene, due to the cancellation of some details present in the original image. Since only connected components are processed, the

[1] http://www.grss-ieee.org/community/technical-committees/data-fusion/data-fusion-contest/

filtering effect does not distort the geometrical characteristics (e.g., the borders) of those regions which are not canceled by the filtering. This is a very important feature of this family of filters especially when processing images in which the spatial features are informative components (e.g., VHR images). Among connected operators, opening and closing by reconstruction have been widely used in remote sensing applications [14]. Operators by reconstruction process an image through a window of given shape and size (parameter that defines the characteristics of the filter) called structuring element (SE) which is adapted to the image content (referred as "mask" in the geodesic reconstruction operation), [6]. By varying the characteristics of the SE different filtering effects are achieved. The application of a sequence of opening and closing by reconstruction with SE of fixed shape and increasing size led to the definition of Morphological Profiles (MPs) [7], which have been successfully exploited for different remote sensing applications [15]. We recall that the concept of MP, i.e., the application of morphological operators in a multiscale setting, was first introduced by Maragos [16].

As for operators by reconstruction, attribute filters (AFs) are connected operators [17]. Actually, the operators by reconstruction are a subset of the AFs [17]. The attribute filters process an image according to a criterion that is evaluated on the image f. A generic criterion T can be defined as a mapping $T : S \rightarrow \Omega$ with S the input domain of the criterion and Ω the set made up by the couple of Boolean values $\{false, true\}$. The criterion, which is evaluated on each connected component of the image, performs a comparison between an arbitrary measure α computed on each CC and a reference value λ, with $\alpha, \lambda \subset S$. For an arbitrary connected component C an example of criterion might be $T(C) = \alpha(C) \geq \lambda$. Any measure which is computable on the image regions can be considered in the analysis. This feature leads to a great flexibility in the definition of the attribute filters, which consequently improves their capability in the spatial information when compared with operators based on fixed SEs (e.g., operators by reconstruction). For example, the attributes considered can be related to features such as the shape and size of the regions, texture, contrast, etc. A very important property of the criterion considered in the transformation is the *increasingness*. A criterion is said to be increasing when, if it is verified for a connected component C_j (i.e., $T(C_j) = true$), then it will be also true for each component C_i that includes C_j (i.e., $T(C_i) = true \, \forall C_i \supseteq C_j$). Examples of increasing criteria involve increasing attributes (e.g., area, volume, size of the bounding box, etc.) and an inequality relation (e.g., $\geq$). In contrast, non increasing attributes, such as scale invariant measures (e.g., homogeneity, shape descriptors, orientation, etc.), lead to non increasing criteria.

In order to introduce the definition of AF for grayscale images, the formulation of AF for binary images is first recalled. Attribute openings for binary images consider an increasing criterion T. According to [17], they are obtained by computing a trivial opening, Γ^T, on the output of a connected opening, Γ_F, applied to all the connected components of a binary image F. Given a pixel $p \in E$ and a connected component C, the connected opening is computed as:

$$\Gamma_F(p) = \begin{cases} C & if \quad p \in C; \\ \varnothing & otherwise. \end{cases} \tag{1}$$

The trivial opening keeps the regions for which the increasing criterion T holds. This can be expressed as:

$$\Gamma_T(C) = \begin{cases} C & if \quad T(C) = true; \\ \varnothing & otherwise. \end{cases} \tag{2}$$

Thus, an attribute opening is then given by:

$$\Gamma^T(f) = \bigcup_{p \in F} \Gamma_T(\Gamma_F(p)). \tag{3}$$

If the criterion considered is increasing, the resulting transformation is increasing, idempotent and anti-extensive (i.e., it is an opening). In contrast, if the increasingness property is not fulfilled by the criterion, the filter remains idempotent and anti-extensive but not increasing. For this reason, the transformation based on a non-increasing criterion is not an opening, but a

thinning.
Analogous considerations can be made for the dual transformation. If the criterion is increasing, the transformation is actually a *closing* otherwise it is a *thickening*.

The extension of the operators from binary to gray-scale images is straightforward when the criterion is increasing because of the principle of threshold superposition [18]. Since a grayscale image can be equivalently be expressed as the sum of all its binary thresholds, then the output image of these filtering transformations is the sum of all the filtered input threshold images, i.e.,

$$\gamma^T(f) = \sum_{k=0}^{K} \Gamma^T(F_k) \tag{4}$$

with F_k the binary threshold image f at graylevel $k \in [0, K]$ the destination domain of the grayscale values. Equation (4) can also be expressed as:

$$\gamma^T(f)(p) = \max\{k : \Gamma^T(F_k)(p) = 1\} \quad p \in E. \tag{5}$$

When the attribute criteria are not increasing, the extension to numerical functions is not straightforward anymore.

For example, let us consider a numerical function f and a binary criterion $T(F_k)$ that acts on the binary sections F_k of f at successive thresholds $k_1 < k_2 < k_3$. For a given position p in the image domain, we may have $T(F_{k_2})(p) = false$, whereas $T(F_k)(p) = true$ for $k = \{k_1, k_3\}$. Thus, the results of the transformation applied to successive sections of the image could not decrease as k increases. Therefore, they cannot be considered as the stack of sections of a function. A way to force the decreasingness of the sequence is to consider as criterion applied to the section at the level k and at the position p the union of the criteria computed on the top sections with respect to k, i.e., $T'(F_k)(p) = \cup\{T(F_i)(p), i \geq k\}$. This leads to the following definition of grayscale attribute thinning (by using the criterion T'):

$$\gamma^{T'}_{max}(f)(p) = \max\{k : \Gamma^{T'}(F_k)(p) = 1\}. \tag{6}$$

This solution leads the grayscale attribute thinning (see (5)) that operates on an image by removing a connected component C if its attribute does not fulfill the criterion and nor holds for the regions included by C. In [19] this filtering strategy was referred to as *max rule*. However, other arbitrary filtering strategies can be implemented in order to achieve different output effects when extending the binary thinning and thickening to numerical images [19, 20]. For example, Urbach *et al.* found that the so-called *subtractive rule* is particular suitable when considering shape descriptors as attributes [20]:

$$\gamma^T_{sub}(f)(p) = k_{min} + \sum_{k=k_{min}+1}^{K} \Gamma^T(F_k)(p), \tag{7}$$

being k_{min} the minimum greylevel mapped by f. Moreover in [20], the concept of pattern spectrum was extended by defining scale invariant pattern spectra obtained with shape attributes. If the criterion is increasing, then (6) and (7) are equal to (5). Similar conclusions can be drawn for attribute closing and thickening.
The application of attribute filters in a multilevel architecture was introduced in [10, 20]. In remote sensing, Attribute Profiles (APs) were defined in [8] as a generalization of the MP concept based on attribute filters.

3. GEOMETRIC INVARIANT MOMENTS

The definition of an efficient feature-based object-recognition tool, independent of object position, size and orientation has been extensively investigated in literature [21][22][23]. In order to characterize the shape of the regions a set of efficient invariant features are needed. These features can be of different types, such as visual features (edges, textures and contours), algebraic features (based on matrix decomposition of the image), statistical features and transform coefficient features. In this paper

we use the method of invariant moments. Geometric invariant moment (GM) was first introduced by Hu [9] and has become a classical tool for feature-based object recognition in remote sensing [22]. The GM technique, derived from the theory of algebraic invariant, is used to extract image features since the features generated are Rotation Scale Translation (RST)-invariant. For a 2-D continuous function *f(x,y)* $x, y \in \mathbb{R}$ the moment of order *(p+q)* is defined as:

$$M_{pq} = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x^p y^q f(x,y) dx dy. \tag{8}$$

Considering a discrete grayscale image with pixel intensities *I(x,y)* (where x, y are the indexes of the rows and columns of the image), the image moment M_{ij} can be derived from 8 by:

$$M_{ij} = \sum_i \sum_j x^i y^j I(x,y) \tag{9}$$

The central moments $\mu_{I(x,y)}$ for a grayscale image are defined as:

$$\mu_{ij} = \sum_i \sum_j (x - \overline{x})^i (y - \overline{y})^j I(x,y) \tag{10}$$

where $(\overline{x}, \overline{y})$ are the components of the centroid with coordinates determined using 11:

$$\begin{cases} \overline{x} = \frac{M_{10}}{M_{00}} \\ \overline{y} = \frac{M_{01}}{M_{00}} \end{cases} \tag{11}$$

Hu defined seven moments derived by normalizing central moments of order three, obtaining invariance under translation, changes in scale, and also rotation:

$$M_1 = \eta_{20} + \eta_{02} \tag{12}$$

$$M_2 = (\eta_{20} + \eta_{02})^2 + 4\eta_{11}^2 \tag{13}$$

$$M_3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2 \tag{14}$$

$$M_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2 \tag{15}$$

$$\begin{aligned} M_5 = (\eta_{30} - \eta_{12})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 \\ -3(\eta_{21} + \eta_{03})^2] + (3\eta_{21} + \eta_{03})(\eta_{21} + \\ \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \end{aligned} \tag{16}$$

$$\begin{aligned} M_6 = (\eta_{20} + \eta_{02})[(\eta_{30} - \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \\ +4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03}) \end{aligned} \tag{17}$$

$$\begin{aligned} M_7 = (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 \\ -3(\eta_{21} + \eta_{03})^2] + 33(\eta_{12} + \eta_{30})3(\eta_{21} + \\ \eta_{03})[3(\eta_{12} + \eta_{30})^2 - (\eta_{21} + \eta_{03})^2] \end{aligned} \tag{18}$$

The first six moments are independently of position, size and orientation, while the last is skew orthogonal invariant, useful to distinguish mirror images. The moments described above can be calculated either from a segmented binary image or from a shape's boundary, although it has been demonstrated that the binary image calculation is less susceptible to noise [24].

4. HEIGHT RETRIEVAL

The main idea of the proposed approach is based on the geometric properties of optical images on urban environments. In very high resolution images, a dense urban environment can be modeled as a projection of vertical and horizontal surfaces. From this point of view, the horizontal surfaces preserves their spatial characterizations if viewed from slightly different angles of view. In the other hand, vertical surfaces present strong geometrical distortions, even if the difference between the two angles of view is not relevant. Fig. 1 shows an example where the same building is represented under different angles of view: the shape of the roof of the building (in red) remains the same in both images while the shape related to the facade of the building (in yellow) presents great differences in the two images. This means that shapes of the same horizontal surface (i.e. the rooftop of a building) viewed from different angles, should have values very similar to the invariant moments (Fig.2). Starting from this assumption it is possible to match the most similar shapes in two images acquired under different angles of view, based on the values of the invariant moments. Once two shapes are matched it is then possible to evaluate the height of the related building from the relative distance of the centroids of the shapes. This assumption is valid only if the difference between the angles of view is not relevant, otherwise the assumption of shape invariance in the horizontal component is not preserved, and also horizontal surfaces present strong distortions.

Fig. 5 presents the complete scheme of the proposed method. In a first step, each angular image is processed with different attribute filters in order to characterize and extract the region candidates for representing the buildings. Each of the resulting image from the filtering phase, presents homogeneous surfaces that are mainly composed by iso-intensity pixels (i.e. having the same digital number (DN). Now connected pixels having the same DN values can be straightforwardly grouped and considered as a unique entity, (or shape), which can be characterized by the DN and by its spatial characteristics. From this point of view it is possible to consider the filtered image as an initial set of shapes. Since it can happen to have some relatively small shapes which can be irrelevant to the analysis, we considered to merge these regions to their adjacent ones. In particular, each shape composed by a number of pixels lower than a certain threshold has been merged with one of the connected shapes that have the most similar DN value. Then the seven Hu invariant moments are computed on each shape after the merging process. At this point, each shape, in each angular images, is geometrically characterized. This means a match between two similar shapes in two different images can be achieved. In particular, this task is performed using a neural network algorithm (NN), whose inputs are two sets of seven Hu invariant moments, identifying a reference shape, extracted from the nadir image, and the investigated shape from one of the other angular images. The used NN was a multilayer perceptron network trained by using a Scaled Conjugated Gradient (SCG) algorithm and having the logistic activation function in each node [25]. The matching task can be achieved using 14 inputs, corresponding to two sets of seven Hu invariant moments, and one output having values between 1 and 0. The training of the network was performed so to have output 1 if the input couple is referred to the same shape in two images, or 0 if the two shapes do not present any geometrical relation. A grid search has been applied to detect the best number of hidden layers and of nodes in the hidden layers, resulting in a topology of 14 input nodes, two hidden layers of 10 nodes each and one output node.

The computational load of the entire process depends on the number of shapes. For each set of images obtained from the four different filters, the process compares each shape of the reference image with all the shapes in the other angular images. Obviously, this approach can greatly increase the time requirement for the process, so to avoid comparison of shapes of moving objects (i.e. airplanes) or shapes too far each other, a constraint on the shapes to compare was set. In fact, In an optical image, high buildings are projected in the far side from the sensor, if the acquisition angle is different from the nadir. In this way it is possible to ease the computational load by choosing in the angular image only that shapes which position from the sensor is far from the reference shape, and that are close to the line of sight of the sensor.

Once trained, the NN algorithm has been applied to each couple of shapes. The output is an estimation of the percentage of similarity between one shape in the reference image with a shape of the considered image. For each couple of images, a file containing a list of the similarity values evaluated comparing each shape in the nadir image to each shape in the angular one

is generated. Having to compare the nadir image to two angular ones for each of the four different filters, a total of 8 lists of similarity have been produced. From these files, only shapes having a similarity over 90% can be matched with the reference shape. If more than one shape per image goes over this threshold, the shape with the highest value will be matched to the reference one. For each match, the coordinates of the two centroids are then used to determine the shift derived from the angle of view. The height of the shapes of horizontal surfaces is then computed by:

$$H = D\frac{Tan(\alpha)Tan(\beta)}{Tan(\beta) - Tan(\alpha)} \tag{19}$$

Where D is the distance between the centroids of the two shapes, while α and β are the angles of view of the off-nadir and nadir image, respectively (Fig.3).

Comparing different angular images, it can happen, for some shapes in the reference image, that a matching can be found in more than one angular image. In this case the average value is selected for the height.

5. RESULTS

The proposed method has been applied to the WorldView-2 dataset provided for the 2011 Data Fusion Contest. The dataset is composed by 5 images of the city of Rio acquired under 5 different angles of view (44.7°, 56.0° and 81.4° in the forward direction, 59.8° and 44.6°) both in PAN and multispectral mode, having 0.5 mt and 2 mt spatial resolution, respectively. For sake of uniformity the resolution of the panchromatic images have been reduced to match the multispectral one.

The attribute filters have been applied only to the 56.0°, 81.4° and 59.8° PAN acquisitions, because the other two images introduced strong distortions also in the shape related to the horizontal surfaces. Being derived from PAN images, characterized by a single band covering the 450-800 nm, the shapes extracted by AFs can represent something different from the real objects in the scene. In particular, two adjacent surfaces can have very similar values, leading the AFs to process them as a unique shape. To overcome this problem and better detect the shapes in the images, the AFs were applied to four different attributes: area, standard deviation of the values of the pixels, moment of inertia (which corresponds to the first invariant moment of Hu, (12)) and the length of the diagonal of the region bounding box, as shown in Fig. 6. Each filter has been applied to the three acquisitions, in order to have 12 resulting filtered images. The combined use of these filters permits to achieve a better detection of the shapes. In particular, in each of the 12 images obtained from these four filters, the connected pixels having the same DN values have been grouped to form a shape. Each shape composed by a number of pixels lower than a threshold of 100 has been merged with one of the connected shapes that have the most similar DN value. This processing step can produce more than one shape related to the same building. In this case the resulting height of the building will be the average value of the different heights. This step became necessary because, as shown in (10), the calculation of the Hu invariant moments are very sensitive to the position of the shape's centroid, and this can change depending on random noise or poor segmentation, resulting in unexpected height values.

In order to further reduce the computational load, water, vegetation and shadow areas were masked. The masking processing has been performed by a classification of the multispectral images associated to the panchromatic ones. In particular, a Support Vector Machines classification algorithm (SVM) has been applied for the detection of 4 different classes: i.e. water, vegetation, shadows and manmade structures. [26]. Starting from the classification result, the pixels related to the water, vegetation and shadow classes have been used to build a mask evidencing only the manmade structures in the scene. Thus only manmade objects were considered in the process. Moreover shapes located at the border of the image were discarded to ensure a better result. For each resulting shapes in each images the seven moments of Hu were computed. Then each shape in the 81.4° image was compared with each shapes in the 56.0° and 59.8° images using the NN retrieval algorithm. In particular the network has been trained in 500 epochs, using 100 training samples and 50 test samples randomly selected from both coupling of reference image with angular ones. Fig. 4 shows the original 81.4° image, and the cleaned image representing only the matched objects.

The use of three angular images became necessary in order to avoid ambiguity between similar shapes and also to detect the shapes of those buildings that are covered by higher ones.
Not all the shapes detected in the 81.4° image have been matched to the other two images because in some cases trees too close to low buildings affected negatively the invariant moments. In other cases some shapes were covered by higher buildings in both 56.0° and 59.8° images. The value of the retrieved height, being the height calculated from the distance of two centroids, can be affected by some errors due to slight differences in the matched shapes. To evaluate the accuracy of the proposed method, the retrieved heights of 30 randomly chosen buildings have been compared with real measurements of the corresponding buildings obtained from Google Earth 3D map. Analyzing the values in Table 1 one can note that almost all the measurements approximate the real heights, except for two buildings, for which the proposed technique retrieves wrong values. These errors, related to the stadium and to the airport terminal, are produced by two different causes. In detail, the error on the stadium was due to a mismatch of the shapes from the 56.0° and 81.4° images, as depicted in Fig. 7. In fact, the NN related the shape of the inner track to the shape derived from the terraces.
As for the airport, the error was introduced by the different shapes of the building in the two angular acquisitions. These differences result in a different relative positions of the centroids. Even if the neural network algorithm was able to correctly match the two shapes, the distance between the two shapes was strongly influenced by the position of the two centroids, producing an incorrect height value. The final number of matched shapes were 338 and the resulting 3D renderings is depicted in fig. 9 and 8.

6. CONCLUSIONS

In this paper we presented a novel approach to the retrieval of the building height using multi-angular optical images. The proposed method aims to detect the buildings footprint, exploiting two main concepts: attribute filters and geometric invariant moments. The attribute filters were used to extract regions from each of the multiangular images, while the invariant moments were computed on each regions to geometrically characterize them. This proposed approach is based on the assumption that a horizontal surface does not change its shapes if looked from slightly different angles of view, while vertical surfaces are significantly distorted. In this way, modeling an urban environment as a sequence of vertical and horizontal surfaces, it is possible to uniquely identify the shapes of horizontal surfaces associated to the rooftop of the buildings, roads and flat fields in all the multiangular images. Exploiting the relative coordinates of the centroids of the matched shapes in each image it was possible to quantify the displacement of the horizontal surfaces due to the different angles of view. Finally, it is possible to retrieve the heights of the buildings from trigonometric triangulations of the displacement of each matched shape.
This approach has been successfully applied to a WV2 multiangular dataset, considering only three out of five very high resolution images acquired with slightly different angles of view (i.e. at-nadir, forward and backward). The other two images have not been used because they were acquired with angles of view that introduced distortions also in the horizontal surfaces. A comparison with ground truth heights reported an average error of about 3m for the retrieved surfaces. Nevertheless it was not possible to extract all the horizontal surfaces in the images, because of the presence of trees close to some shapes that influenced the values of the invariant moments or because some shapes were not detected because they were occluded by the surrounding buildings. Moreover, being the height retrieval process based on the distance between the centroids of two matched shapes, the height of two shapes have been affected by errors in the position of the centroids, resulting in incorrect retrieved height values.

7. REFERENCES

[1] S. Guillaso, L. Ferro-Famil, A. Reigber, and E. Pottier, "Building characterization using l-band polarimetric interferometric SAR data," *IEEE letters on Geoscience and Remote Sensing*, vol. 2, pp. 347–351, 2005.

[2] R. Guida, A. Iodice, and D. Riccio, "Height retrieval of isolated buildings from single high-resolution SAR images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, pp. 2967–2979, 2010.

[3] D. Brunner, G. Lemoine, L. Bruzzone, and H. Greidanus, "Building height retrieval from VHR SAR imagery based on an iterative simulation and matching technique," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, pp. 1487–1504, 2010.

[4] H. Sportouche, F. Tupin, and L. Denise, "Extraction and three-dimensional reconstruction of isolated buildings in urban scenes from high-resolution optical and SAR spaceborne images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, pp. 3932–3946, 2011.

[5] A. Massalabi, H. Dong-Chen, G. B. Benie, and E. Beaudry, "Detecting information under and from shadow in panchromatic ikonos images of the city of sherbrooke," *Geoscience and Remote Sensing Symposium,2010 IEEE International,IGARSS 2004*, vol. 3, pp. 2000–2003, 2004.

[6] P. Soille, "Morphological image analysis, principles and applications," *2nd ed. Berlin, Germany: Springer-Verlag*, 2003.

[7] M. Pesaresi and J. A. Benediktsson, "A new approach for the morphological segmentation of high-resolution satellite imagery," *IEEE Trans.Geosci. Remote Sens.*, vol. 39, no. 2, pp. 309–320, 2001.

[8] M. Dalla Mura, J. A. Benediktsson, B. Waske, and L. Bruzzone, "Morphological attribute profiles for the analysis of very high resolution images," *IEEE Transactions on Geoscience and Remote Sensing.*, vol. 48, no. 10, pp. 3747–3762, 2010.

[9] M. K. Hu, "Visual pattern recognition by moment invariants," *IRE Trans. Info. Theory*, vol. IT-8, pp. 179–187, 1962.

[10] E. R. Urbach, N. J. Boersma, and M. H. F. Wilkinson, "Vector-attribute filters," *Mathematical Morphology: 40 Years On, Proc. Int. Symp. Math. Morphology (ISMM) 2005*, 2005.

[11] P. Salembier and J. Serra, "Flat zones filtering, connected operators, and filters by reconstruction," *IEEE Transactions on Image Processing*, vol. 4, no. 8, pp. 1153 –1160, Aug. 1995.

[12] P. Salembier and M.H.F. Wilkinson, "Connected operators," *IEEE Signal Processing Magazine*, vol. 26, no. 6, pp. 136 –157, november 2009.

[13] J. Serra, "Connectivity on complete lattices," *Journal of Mathematical Imaging and Vision*, vol. 9, no. 3, pp. 231–251, 1998.

[14] P. Soille and M. Pesaresi, "Advances in mathematical morphology applied to geosciences and remote sensing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 40, pp. 2042–2055, 2002.

[15] M. Dalla Mura, J.A. Benediktsson, J. Chanussot, and L. Bruzzone, *Optical Remote Sensing - Advances in Signal Processing and Exploitation Techniques*, chapter The Evolution of the Morphological Profile: from Panchromatic to Hyperspectral Images, Springer Verlag, 2011.

[16] P. Maragos, "Pattern spectrum and multiscale shape representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 7, pp. 701–716, 1989.

[17] E. J. Breen and R. Jones, "Attribute openings, thinnings, and granulometries," *Comput. Vis. Image Underst.*, vol. 64, no. 3, pp. 377–389, 1996.

[18] P. Maragos and R.D. Ziff, "Threshold superposition in morphological image analysis systems," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 5, pp. 498–504, 1990.

[19] P. Salembier, A. Oliveras, and L. Garrido, "Antiextensive connected operators for image and sequence processing," *IEEE Transactions on Image Processing*, vol. 7, no. 4, pp. 555–570, 1998.

[20] E. R. Urbach, J. B. T. M. Roerdink, and M. H. F. Wilkinson, "Connected shape-size pattern spectra for rotation and scale-invariant classification of gray-scale images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 2, pp. 272–285, 2007.

[21] J. Flusser and T. Suk, "Rotation moment invariants for recognition of symmetric objects," *IEEE Trans. Image Proc.*, vol. 15, pp. 3784–3790, 2006.

[22] A. Khotanzad and J. Lu, "Classification of invariant image representations using a neural network," *IEEE Trans. Acoust., Speech Signal Process.*, vol. 38, pp. 1028–1038, 1990.

[23] H. Wechsler and G. Zimmerman, "2-d invariant object recognition using distributed associative memory," *IEEE Trans. Pattern Anal. Mach. Intel.*, vol. 10, pp. 811–821, 1998.

[24] S. A. Dudani, K. J. Breeding, and R. B Mcghee, "Aircraft identification by moment invariants," *IEEE Transactions on Computers*, vol. C-26, no. 1, 1977.

[25] C. Bishop, "Neural networks for pattern recognition," *Oxford Univ. Press, London, U.K.*, 1995.

[26] C. C. Chang and C. J. Lin., "Libsvm : a library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, pp. 1–27, 2011.

Retrieved height [m]	Real height [m]	Difference
20	16	4
31	35	4
61	65	4
72	78	6
50	**7**	**43**
146	139	7
70	73	3
10	10	0
14	20	6
8	10	2
68	63	5
18	24	6
71	74	3
19	19	0
12	12	0
9	8	1
40	44	4
12	12	0
11	15	4
49	45	4
23	30	7
65	**13**	**52**
10	13	3
41	42	1
17	16	1
6	10	4
15	14	1
20	22	2
15	18	3
14	15	1
	Average error [m]	6.03

Table 1. The table reports the errors between the estimated heights of 30 known heights taken as reference. The results related to the Airport and the Stadium are underlined. The average error computed by excluding them is 3.23 m.

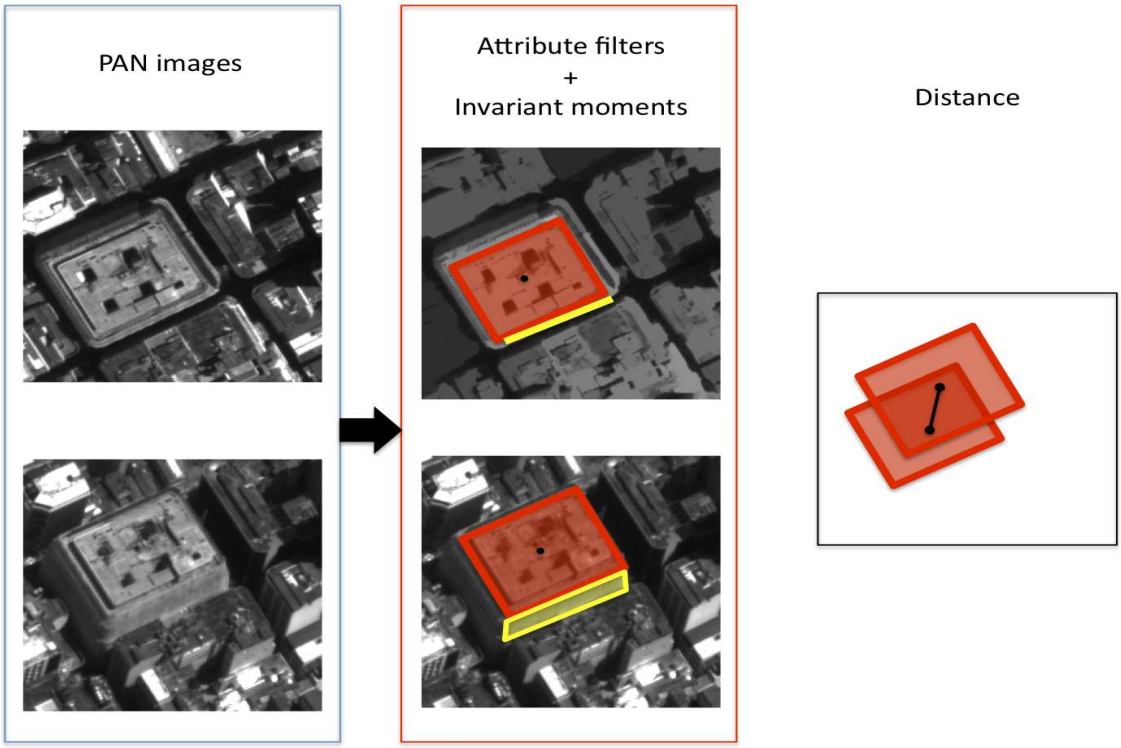
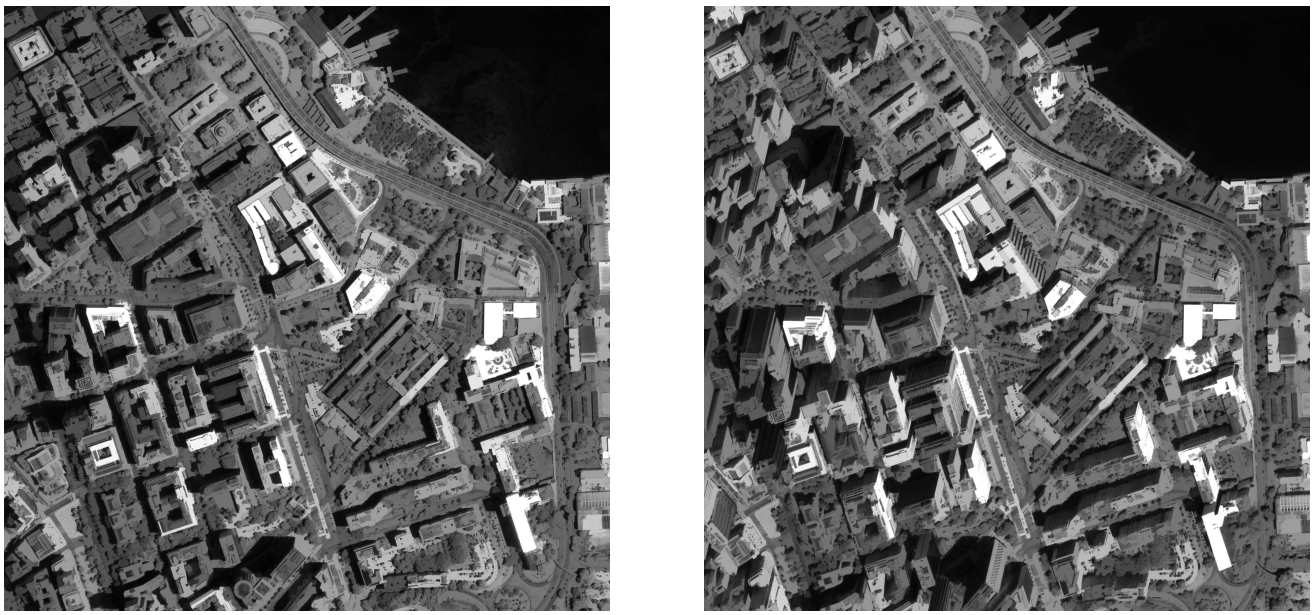


Fig. 1. *Example showing how the different angles of view differently influence the geometrical shape of vertical and horizontal surfaces in urban environment.*



(a) 81.4° (b) 44.7°.

Fig. 2. Filtered PAN images acquired at different angles.

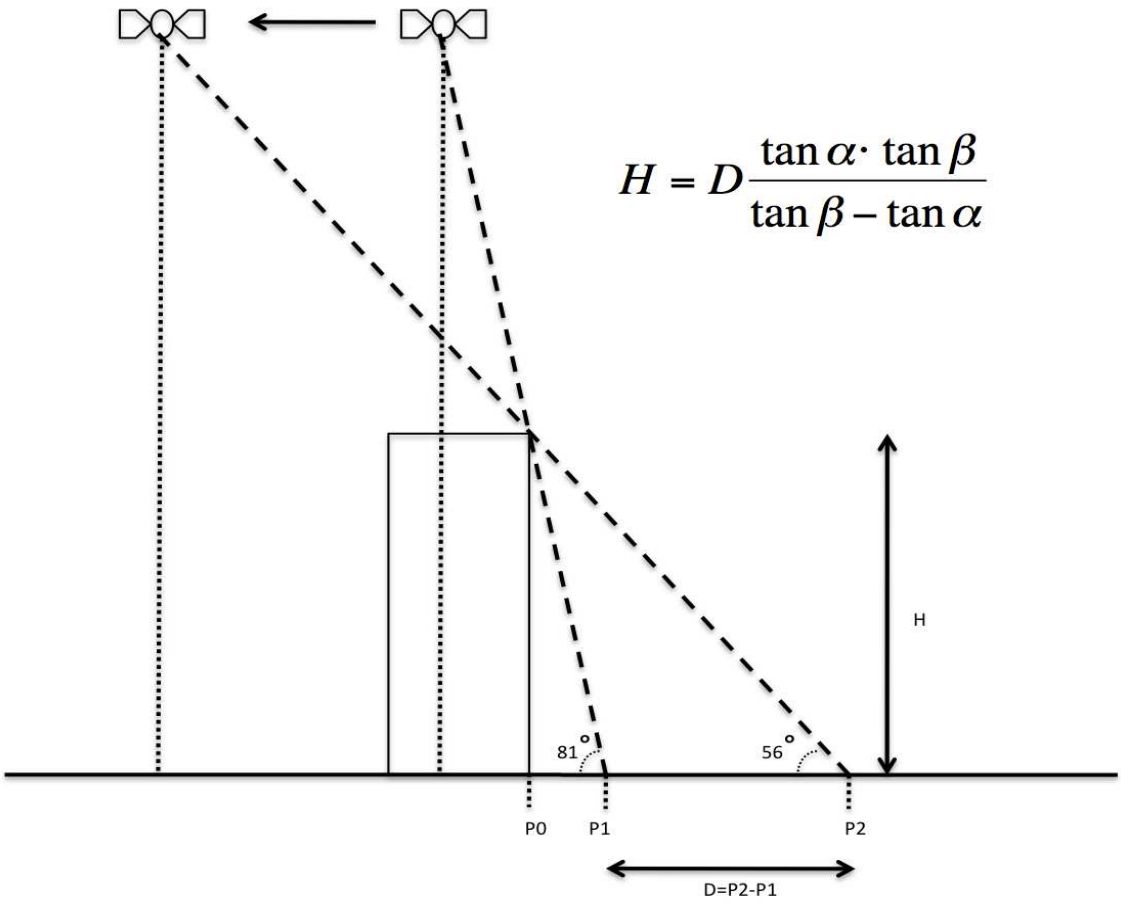
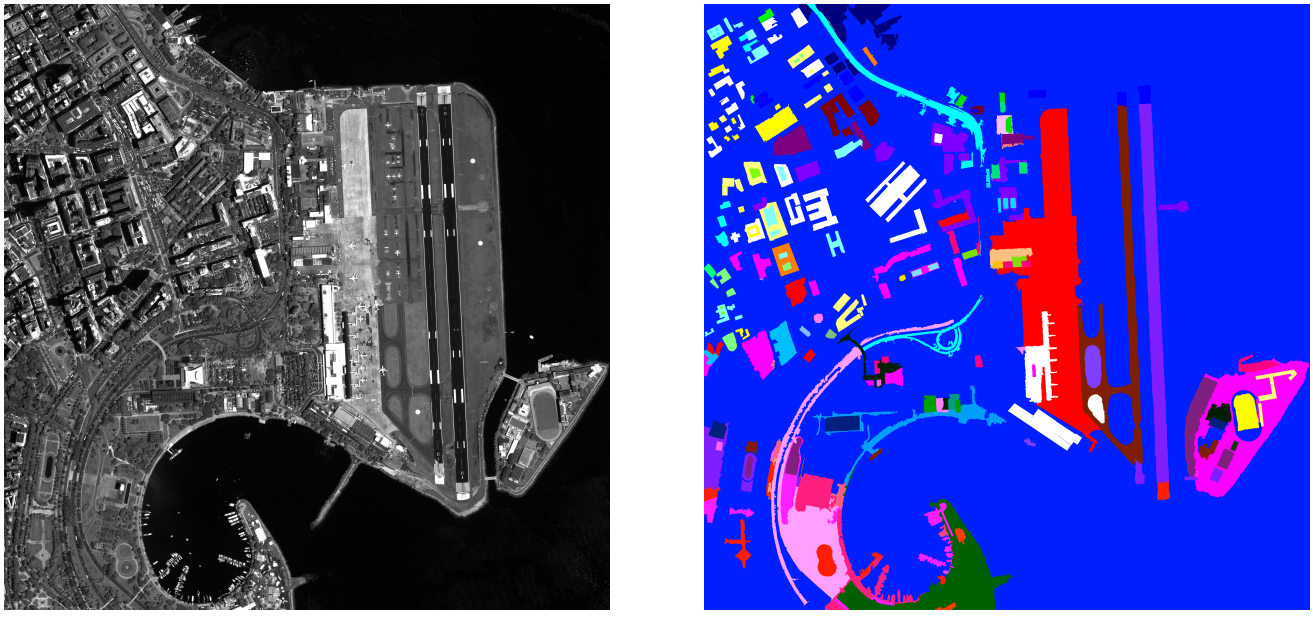


Fig. 3. *Schema representing the height retrieval method from two angular images.*



(a) Original PAN image

(b) Filtered image.

Fig. 4. Original 81.4° PAN image (a) and the relative filtered image base on the Area attribute, representing only the matched shapes (b).

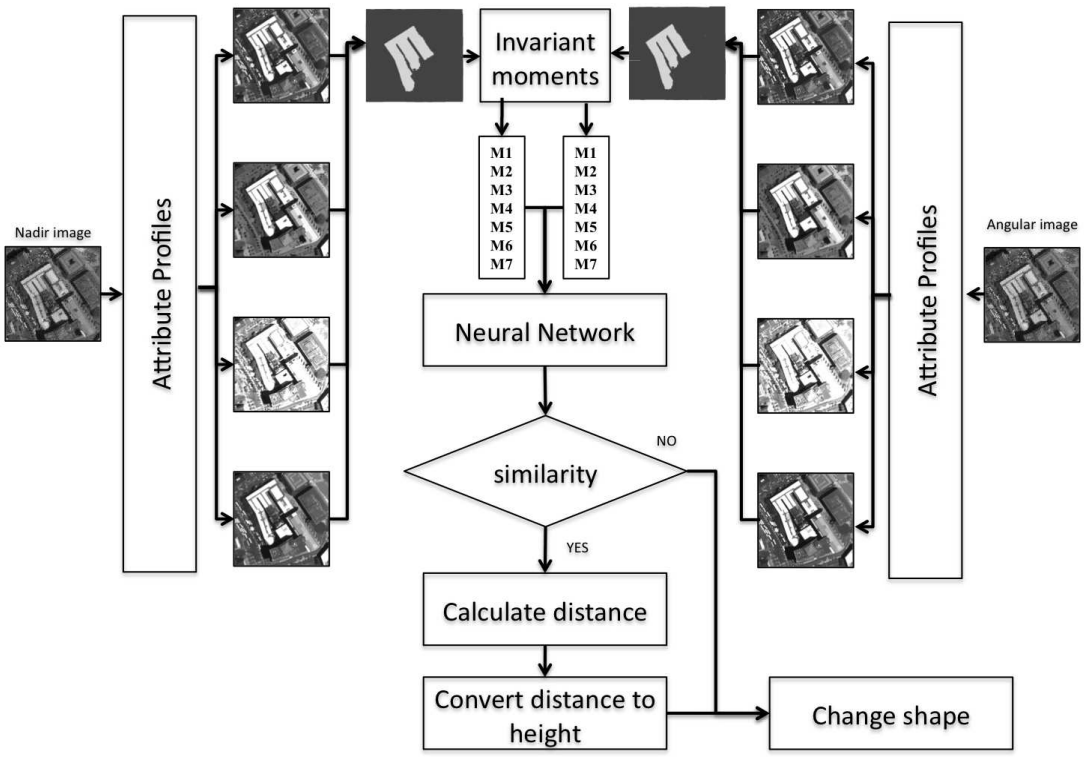
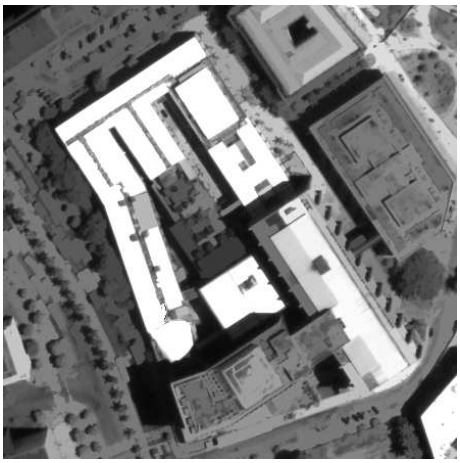
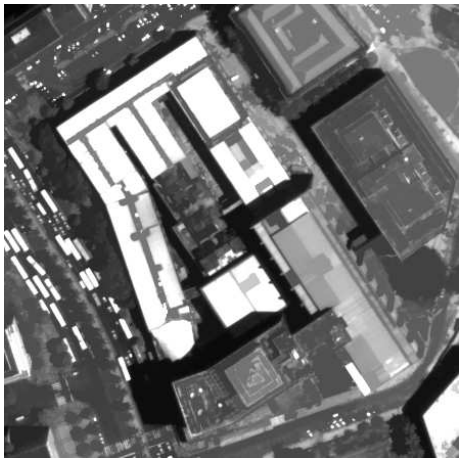


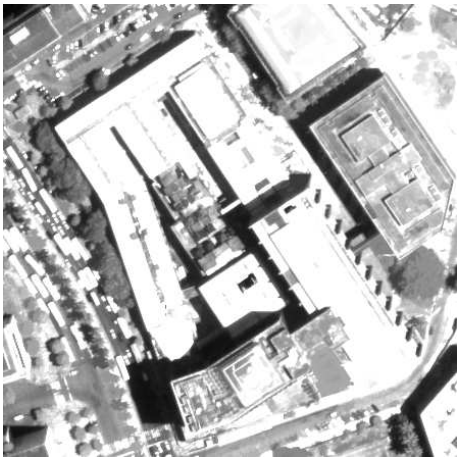
Fig. 5. *Complete scheme of the proposed technique. The attribute filters are applied to the PAN images and the invariant moments are computed on the resulting shapes. A matching between objects in the two images, based on the moments values similarities, is performed by a neural network algorithm. Finally the distance between the centroids of two matched shapes is used to derive the object height.*



(a) Area filter



(b) Diagonal filter.



(a) Inertia filter

(b) Standard deviation filter.

Fig. 6. Samples of the four different Attribute filters used in the proposed methos.

Fig. 7. *Example of two shapes that have not been matched correctly.*

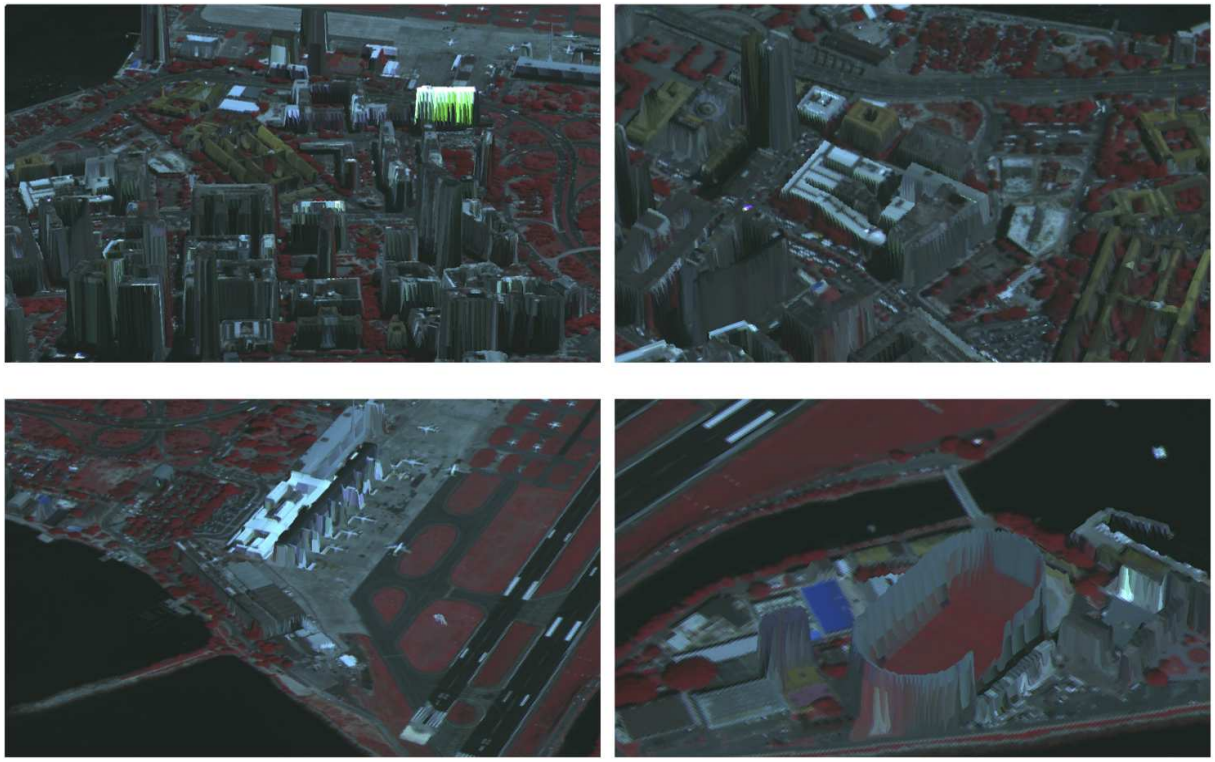


Fig. 8. *Detail of a 3D rendering representing the buildings height retrieved using three angular images.*

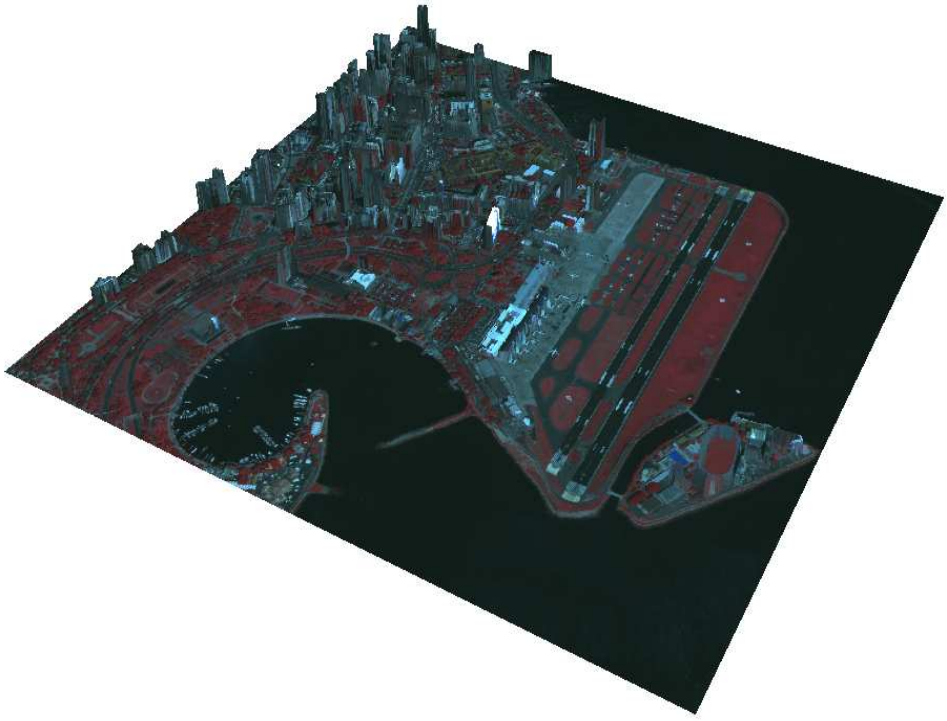


Fig. 9. *3D rendering representing the buildings height retrieved using three angular images.*