Building Change Detection by Histogram Classification

Charles Beumier, Mahamadou Idrissa
Signal and Image Centre, Royal Military Academy, Brussels, Belgium
{beumier, idrissa}@elec.rma.ac.be

Abstract

This paper presents a supervised classification method applied to building change detection in VHR aerial images. Multi-spectral stereo pairs of 0.3m resolution have been processed to derive elevation, vegetation index and colour features. These features help filling a 5-dimensional histogram whose bins finally hold the ratio of built-up and non built-up pixels, according to the vector database to be updated. This ratio is used as building confidence at each pixel to issue a building confidence map from which to perform building verification and detection. The implementation based on histogram is very simple to code, very fast in execution and compares in this application to a state-of-the-art supervised classifier. It has been tested for the Belgian National Mapping Agency (IGN) to identify areas with high probability of change in building layers.

1. Introduction

In our rapidly evolving world, up-to-date topo-geographical information is of prime importance for applications such as navigation, environmental planning and risk assessment. Although information technology has been rapidly developing the topo-geographical data production is only partially handled by automatic procedures [1]. Much more, the effort needed to update topo-geographical information may be as high as half the effort required to complete the first release [2]. No need to say that with the increasing desire to maintain information up-to-date, mapping agencies are looking for fast solutions.

Change detection helps focusing the update work where modifications are most probable. Up to now, the update efforts are measured in man month since the procedures largely involve human intervention. The current trend is to consider image processing as a way to automatically detect candidates for changes: the operator can then concentrate on those candidates and avoid the time-consuming and error-prone task of scrutinizing large images. In some applications, a simple image difference can reveal most of the modified areas, but in aerial images the influences of the sun, vegetation and angle of view make the relation 'image to object' very complex.

To circumvent the complexity of images, one has to find robust and discriminative features on which to base change detection: robust to offer reliable values, as weakly dependent as possible on influences external to the objects of interest and discriminative in order to distinguish real targets from false alarms. Common imagery used by mapping agencies nowadays includes stereoscopic pairs of multi-spectral Very High Resolution images (ground resolution less than 0.5m). Such images can be processed to extract a Digital Surface Model from which to look for elevated structures; to apply colour analysis to isolate red or grey roofs; and to estimate from the near infrared band a vegetation index from which to localize non-vegetation objects. The combination of those features to detect built-up areas brings robustness to the detection problem as the chance to miss targets for every feature dimension is low. However, the fusion of the features is not obvious as several measures with different ranges and importance have to be mixed to issue one decision value.

The problem of deciding if a given image area is built-up or not relates to classification. In our case, the desired output is binary and the input is a vector consisting of 5 dimensions: 3 for colours, the local elevation and a vegetation index. As the database contains examples with known class, the problem reverts to supervised classification for which examples feed the learning phase to let the classification engine adapt its output. For the quality of results, the generalization of the classifier towards new (not-learnt) examples is crucial. In our application, database errors only concern a small percentage and relatively few new feature values appear.

We propose to handle supervised classification with one (or several) histogram. This histogram holds for each feature vector (derived from the recent images) the ratio of built-up and non built-up pixels, according to the old database. A building confidence map is derived by replacing each pixel by the ratio associated to its feature vector. Change detection is achieved from the comparison of the database with the building confidence map. This approach is direct and supposes that the building types are sufficiently represented in the database, what is commonly the case for a given geographical area and a usual time span of a few years. A feature vector not encountered during learning will be classified as unknown and should call for human supervision.

Our development fits with some of the trends of building change detection from aerial images [3]. It processes a pair of multispectral images to derive 3D and spectral cues. These are combined by a supervised classifier exploiting the a priori class belonging contained in the old database to detect inconsistencies like building disappearance, modification or appearance. The implementation is simple to code, runs very fast and compares to state-of-the-art supervised classification in our specific case.

This paper is organized as follows. Section 2 presents the evolution of our developments and the related image features used in this work. Section 3 describes supervised classification with histogram and section 4 shows the change detection results obtained with that classifier. Section 5 concludes the paper.

2. Feature Extraction

2.1. Our progress in Change Detection for Buildings

The work presented in this paper is the continuation of developments for building change detection areas started more than five years ago. This subsection presents a summary whose details can be found in mentioned papers.

Geometry and Shadow

Our first building detection study [4] showed that two pieces of information are highly important for building verification in panchromatic monoscopic images: the presence of linear segments and the presence of shadow that attests the prominence of buildings relative to their neighbourhood.

Although the developed system correctly confirmed most individual or groups of buildings described by polygons in the database, it was not designed to find new constructions. The presented approach would have to look for areas with linear segments and shadow, resulting in numerous candidates and many false alarms as the constraints on building size, shape and orientation are weak. Even if it is practically grounded to keep human supervision in the process for new building detection, a solution with too many false alarms reduces the advantage of the automatic image processing part.

Disparity

A more elegant solution exists when considering the elevation clue obtained from stereo couples of images. In a first attempt, we considered the information contained in the disparity map which is the distance of corresponding points in the left and right images of a stereo pair. Disparity values are related to the distance of the object to the camera and thus reveal the relief for airborne or space-borne imagery.

Looking for buildings in the disparity map consists in highlighting locally elevated objects [5]. Simply setting a threshold is not applicable as the disparity depends on the building height and terrain elevation. The terrain level (in disparity) has to be first subtracted to flatten the area so that building can be automatically detected by thresholding with a given disparity value.

Geo-referencing

Although disparity is sufficient to detect buildings, a more practical solution consists in reconstructing the scene with 3D values representing real coordinates [6]. This necessitates the transformation of the disparity value associated to each (x,y) image position into geographical coordinates (X,Y,Z). The quality of the results highly depends on the precision of the camera and flight parameters. The knowledge of ground points helps to achieve better geographical coordinates.

Correct 3D reconstruction (“Digital Surface Model”) is also necessary if one wants to ortho-rectify the images, to bring them into a geometry corresponding to a top view (nadir) at every pixel and compatible with database coordinate systems. For this, each image point has to be shifted according to its image coordinates, camera position and scene elevation.

3D Building Verification

The knowledge of precise coordinates for every point in the scene is crucial to be able to position the database on the extracted 3D data. 3D Building verification then simply consists in checking that Z values (elevation) are higher in the building polygons. If the database has Z values, they can be numerically compared with some tolerance to account for imprecision and generalisation rules (like the building Z values measured at the gutter level in Belgium).

For better efficiency, it is more appropriate to subtract a DTM (Digital Terrain Model) from the DSM to derive a *local elevation map*. The DTM may be obtained from topo-geographic sources or derived from the DSM by a procedure keeping local low-level values [6]. Building verification is then achieved by the threshold of the local elevation map with a constant Z value (2 or 3m, for instance).

Although the approach is clearly supported by experimental results, some buildings were incorrectly verified, mainly due to their limited size or height and wrong disparity estimation. Most errors concerned garden huts (about 10 m^2, 2.5 m high) and buildings hardly visible due to a lack of contrast resulting in a poor matching for disparity estimation (and hence 3D). Mention that the resolution in Z is in our case on the order of 1 m, so that a threshold at 2.5m generates noise in the local elevation map.

3D Building Detection

The detection of new buildings relies on the threshold of the local elevation map. However, for an automatic solution, new building candidates have to be localised and false targets should be filtered out. In many situations, human supervision is the most appropriate way to deliver a practical 'semi-automatic' solution.

The difficulty with automation is the false alarm rate. Usually numerous trees present size and height similar to buildings and their shape roundness is not an easy criterion on which to base rejection due to the limited geometrical sharpness of the DSM or the variety of building shapes. This fact calls for a multi-modal approach including colour and vegetation index.

2.2. Multi-modal feature extraction

A classical way to overcome the limitation of verification or detection performance with one criterion consists in fusing several modalities [7, 8]. In our case, we dispose of a stereo pair of multi-spectral images. These are used to derive the elevation map, the colour and vegetation index values.

3D

The 3D modality has been presented here above and exploits the local elevation map obtained by the subtraction of the DSM by the DTM. The DTM was derived from the DSM by replacing each DSM value with the percentile 20 of the histogram of DSM values contained in a 50x50m square centred on the considered point. The map DSM-DTM (Fig. 1) clearly highlights elevated objects (building and vegetation) but some points have an incorrect value. Wrong estimations are present in the DSM due to common errors in image matching (occlusion, shadow, repetitive patterns) and in the DTM due to the principle of the percentile 20 which is not valid if there are not enough ground points in the 50x50m square.

Fig. 1. Part of the image DSM-DTM (0.5m) of the Leuven suburb

Colour

The colour modality is handled by considering the "Lab" colour space representation [9] computed from the R, G and B channels. RGB values are indeed highly correlated by intensity. *L*, *a* and *b* represent respectively the intensity, the red-green factor and the yellow-blue factor.

We implemented the following rough approximation of the Lab conversion which is less computationally expensive:

$$L = (r+g+b)/3; \quad a = r-g; \quad b = g-b \tag{1}$$

Fig. 2 displays in RGB about 5% of the area used for test. It is a reference to interpret the other figures.

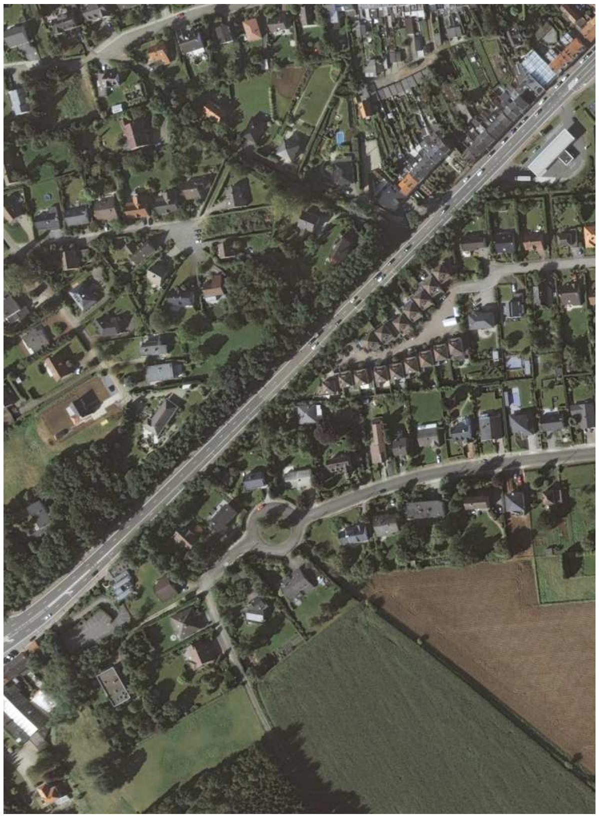

Fig. 2. Part of the ortho-rectified image (0.5m) of the Leuven suburb

Fig. 3. NDVI for the partial image considered (suburb of Leuven)

Vegetation Index

The Normalized Difference Vegetation Index is the last modality included in our multi-modal classification engine. It is computed at each pixel by the ratio:

$$NDVI = (NIR-R)/(NIR+R) \quad (2)$$

with NIR and R being respectively the near infrared and the red image channels. The value is high (bright) for areas with vegetation. As visible in Fig. 3, buildings and roads appear dark.

Geometry

A modality based on geometry was not developed. Buildings generally appear as blobs with quite sharp and linear edges. It should be possible to integrate this information but in practice the size and shape vary considerably and the resolution at 0.5m is still limitative to detect the linear aspect of small building outlines which are precisely the ones to raise difficulties in other modalities.

3. Classification with Histogram

The idea of multi-modality is to combine different sources of information that should collectively improve classification. In the present case, we dispose of examples from the database to be updated that can help learning the modality values encountered in and out building areas. The classification into building or non building will be obtained from a generalisation of the examples presented during the learning phase hence the name ‘supervised’ classification.

Mention that the database may be fully exploited for learning the classifier. For research purposes, one can limit the part of the database used for learning and use the rest for testing the ability of verification. In operational conditions, there is little ground for not using the whole database for learning so that the chance that building types are not represented is at the lowest level.

Histogram Classifier

Our classifier uses histogram to collect statistics about modality values from building and non-building examples. Modality values can represent single or multiple vector components. In our application dealing with 5 dimensions (z, L, a, b, ndvi), two cases were considered: a 5-D histogram and the combination of histograms of lower dimensions. Let us describe how a single histogram classifier works.

The learning phase of our classifier consists in filling the histogram with building and non-building examples of feature vectors. We created a mask with the building polygons of the database. For each pixel in the building mask, we increment the histogram bin corresponding to the feature vector as 'in' if the pixel is in a building or 'out' if not. At the end of the process, we have a number of 'in' and 'out' occurrences for each feature vector value which are then divided respectively by the total number of 'in' and 'out' pixels so that histogram bins now contain the frequency 'inP' for being 'in' and 'outP' for being 'out'. The confidence (in [0..1]) for a feature vector to represent a building pixel is computed as the ratio of 'inP' divided by the sum of frequencies ('inP'+'outP'). Feature values with no occurrence receive a confidence value of -1 to let them be distinguished easily.

The evaluation of the ratio of in and out pixels is more reliable if sufficient points are considered. The original dynamic range for a modality is usually too large and results are less noisy when the number of histogram bins is reduced. However, with too few bins, the discrimination power decreases. We used 32 bins for z and L dimensions and 15 bins for a, b, and NDVI.

The advantages of this histogram implementation for supervised classification lie in the simplicity of code, the reduced number of parameters (number of intervals per modality), the low computational load and the possibility to keep low memory requirements by cutting the problem into several histogram classifiers. However, a more advanced technique like vector quantization could address the problem of optimal values and number of bins, especially if the memory constraint is strong [10].

Combining Classifiers

There are two approaches to realise multi-modal histogram classification.

In the first case, the whole feature space is handled by a single histogram. This leads to a discriminative solution which can take all feature combinations into account. However, the histogram is likely to have sparse areas and the generalisation to unseen cases may be weak. The histogram size might be very large as it results from the product of the number of intervals for each modality.

The second approach considers the separation of the multi-modal problem into sub-classifiers. During development, we tested each modality separately and analysed their relative importance. We designed one 3-D histogram for colour and two 1-D histograms for 3D and NDVI. Then we tested their combination with a fusion of the individual histogram confidences, what was realised by a classical fusion rule (sum, product, min, max) although another histogram classifier could have been used. This 'divide-and-conquer' approach provides more control as individual classifiers can be tailored individually. However, feature vector values may loose their specificity once vector components are considered separately. For instance, red roofs are very specific and give a high confidence for buildings in the 'Lab'-histogram. If a 1-D histogram was used for 'a', the bins populated by red roofs would also hold other colours with the same 'a' value, probably outside buildings, reducing the confidence for building.

Classification by histograms is very fast and easy to implement or analyse. The results of course depend on the quality of the learning samples. These should at best cover the different cases to be encountered. As the solution suggests the intervention of an operator, additional examples may be pointed by him beside the ones collected automatically from the database. This is particularly useful for the histogram areas which were not classified (empty bins).

4. Change Detection Results

In the context of our research, change detection concerns the automatic detection of inconsistencies between the database to be updated and recent imagery. The changes help the project manager to give priorities to the most demanding areas according to the available human and budget resources. Change locations will also guide the operator in focusing on areas were major changes occurred and where operational tasks have to be undertaken (polygon suppression, creation or modification).

According to the type of change, two different situations arise: *building verification* if the database polygons are checked for modification, or *building detection* if new buildings are looked for in the image. Building verification is easier than detection as the database indicates where to search in the image.

Building Confidence Map

The input data for building change detection consist of a pair of aerial images (0.3m resolution) and the building layer of the topo-geographic database.

A Digital Surface Model was extracted from the stereoscopic pair of multispectral images thanks to a proprietary development using several correlation windows of different sizes and a relaxation procedure based on Markov Random Field [6]. With this DSM, the left image was ortho-rectified (at 0.5m) so that the vector database could be correctly superimposed. The vegetation index and Lab-like map were derived from the Near Infrared, R, G and B channels.

The test zone is a 2km x 2km square area around Leuven in Belgium and contains a large variety of objects: motorway, railways, a lake, urban and agricultural zones. It is representative of the type of landscape in Belgium, except the hilly south and dense urban areas.

The confidence map obtained by the 5-D histogram classifier (3D, L, a, b, NDVI) is created in less than 5 seconds for a 4000x4000 image on a 2.33 Ghz computer. A part of the map is displayed in Fig. 4 with the database superimposed as dotted polygons.

Building Verification

We observe in Fig. 4 that the confidence map is consistent with the outlines of the building polygons of the database. Most inconsistencies concern very small buildings (less than 20 m^2, like garden huts) or errors in shadow areas near vegetation.

To estimate the building verification rate, each polygon of the building database was scored with the average of the confidence values in the polygon. Table 1 lists the number of buildings which have a score lower than the threshold given in abscissa (false rejection). Mention that a limit in building area to 20 m2 has been set to avoid most garden huts to be part of the evaluation since their small size and height are the cause of many errors in feature evaluation and classification.

To find a confidence threshold for building acceptance, we looked for errors in the database. In the whole 2x2 km test area, we found 15 polygons which have no corresponding structure in the image. These are called 'Phantoms' in Table 1 which lists their number with a score inferior to the threshold (false acceptance).

Fig. 4: Part of the Building Confidence Map from 5-D classification.

Table 1. Polygon scores for buildings and phantoms (Min area is 20 m2)

Scores*100	**0**	**10**	**15**	**20**	**25**	**30**	**50**	**100**
# Buildings	0	3	10	17	31	42	106	1045
# Phantoms	15	9	5	2	1	0	0	0

In an operational test, one would set the threshold to highlight weakly supported buildings (like the Phantoms) while keeping the number of highlighted buildings low to minimize false alarms. With a value of 0.21, only one phantom would be missed while 20 real buildings (2%) would be incorrectly highlighted. This is valuable as 98% of the polygons would be released from operator verification. However, the reality is more complex since a few buildings are partially changed, what is hardly detectable with the current implementation.

If additional features had to be combined, the histogram memory requirement would become prohibitive. To analyse the influence of handling the classification with several sub-classifiers, we considered a 4-D histogram (3D and Lab-like features) and a 1-D histogram (NDVI). The two sub-classifiers were combined with a product. Results are not as good but consistent. However if we consider the product of the 3 classifiers obtained with one modality left out (Lab+3D, Lab+NDVI, 3D+NDVI), results are similar.

We also tested the change detection ability in the case 90% of the database was randomly chosen for learning. Results were similar, supporting the conclusion that the quality of the results was not fooled by the learning of isolated cases (with poor generalisation) but that the major buildings characteristics are well captured.

Building Detection

Detecting new buildings requires localising candidate zones in the building confidence map. So far, this localisation has not been automated due to the large number of false alarms. A quality evaluation based on observation was undertaken to identify the discrepancies between the database and the confidence map. These consist of several new buildings, very few destroyed buildings and many false alarms arising from small structures or shadow areas. Refer to Fig. 5 for a few annotated examples, where automatic false colouring based on the database and confidence map highlights the different cases. Green areas correspond to confident pixels within a building polygon (no change). Red pixels have a high built-up confidence out of a polygon and should correspond to new (N) or extended (M) buildings. They are sometimes located in shadow areas 'S' (false alarms). Blue regions have low built-up confidence in a polygon and should correspond to a destroyed building (D) or phantom polygon).

False alarms mainly originate from the limited accuracy of the local elevation map and the poor contrast in shadow areas. The DSM has a resolution of about 1m in Z and the match with DTM values is imperfect, so that the local elevation uncertainty may compete with the lowest structures (e.g. 2.5 m). The Normalized Digital Vegetation Index suffers from a wrong estimation in the shadow areas where both infrared and red components have low intensities. We hope to compensate partially for this thanks to the additional bits (8-11) provided by most earth observation sensors.

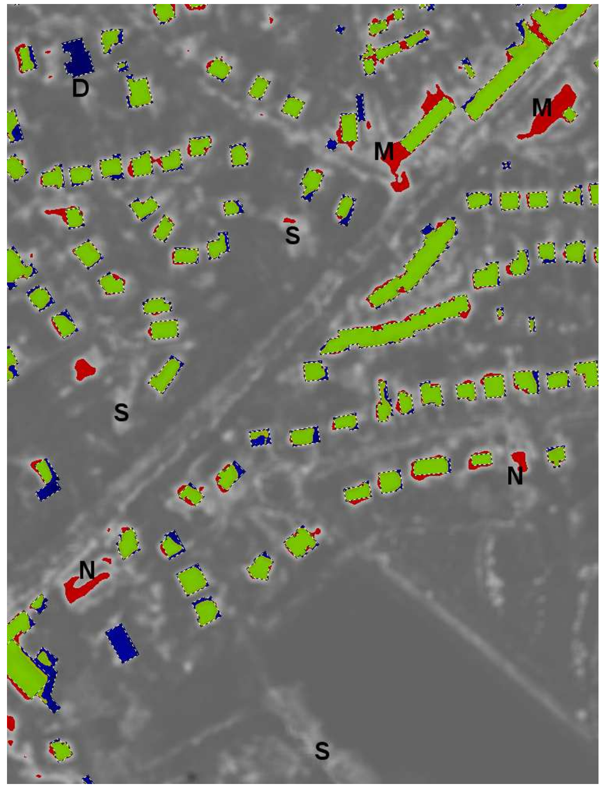

Figure 5: Part of the Building Confidence Map with automatic false coloring: GREEN: confident in DB, RED=confident out of BD, BLUE=not confident in DB. A few examples of New (N), Modified (M) and Disappeared (D) buildings, and Shadow (S) areas.

Feature Relative Importance

When looking at individual feature maps or individual classifier results, we observe that the local elevation brings the highest discrimination power, with a rather good localization precision. Then follows the vegetation index which clearly helps rejecting vegetation zones, except in shadow areas. The colour information is also very useful but its contribution is less predictable. Red roof is a clear supportive example, as this hue is unlikely to be present in other objects, but red containers can be perfect imposters.

State-of-the-art Classifier

In order to evaluate the quality of our feature vector for building classification and to position the histogram approach, we selected Support Vector Machine as a reference classifier. We used a SVM library [10] and we presented the same feature vector with learning thanks to the building map of the database. Using the prediction value (not the binary output) we obtain a confidence map similar to Fig. 4. The quantitative results are slightly worse than Table 1 but as learning and classification are more than 1000 times slower than with our approach, the SVM parameters were not fully tuned.

5. Conclusions

This paper has presented the change detection of built-up areas thanks to multi-modal features composed of elevation, colour and vegetation index values. The information is extracted from a stereo couple of multi-spectral aerial images at 0.3m resolution owned by the Belgian National Mapping Agency (IGN).

The approach consists in learning pixel-wise the building confidence from examples contained in the database. Computing the relative importance of building and non-building pixels for the feature vector composed of the multi-modal information is made by histogram. This way of implementing supervised classification avoids tricky parameter setting, is simple to code, runs very fast and is successful for our application.

The generalisation ability of the classifier necessary to detect new buildings is improved thanks to the reduction of the modality intervals. This represents the only parameters of the histogram classifier but in practice a value between 15 and 30 seem adequate. In the case too many dimensions make the histogram approach unpractical due to memory requirements, the problem can be divided into sub-classifiers but the solution might not to be optimal if we loose the synergy of some modalities.

The execution time is particularly fast: less than 5 seconds for a 4000x4000 image on a 2.33 GHz computer.

We intend to refine change detection by improving the extraction of individual features as suggested above for the local elevation and vegetation index and a special treatment of shadow areas. Concerning histogram classifiers, we would like to setup an automatic procedure for an optimisation of the histogram intervals of the modality value ranges. To a larger extent, we plan to redirect research towards object processing to better capture building modification and propose building vectors.

Acknowledgment

This study has been funded by the Belgian Ministry of Defence. We would like to thanks the Belgian National Mapping Agency for providing the image and vector data gracefully.

References

[1] A. Busch, M. Gerke, D. Grünreich, Ch; Heipke, C.-E. Liedtke, S. Müller, “Automated Verification of a Topographic Reference Dataset: System Design and Practical Results”, In *Int. Archive Photogrammetry and Remote Sensing XXXV, B2*, Istanbul 2004, pp 735-740.

[2] E. Bayers, “Update Topographic Reference Data (Processes, Methods, Needs, Research): Progress Report in the Belgian NGI-IGN”, *ISPRS Archive Vol. XXXVIII, Part 4-8-2-W9, “Core Spatial Databases”*, Haifa, Israel (2010).

[3] E. P. Baltsavias, “Object extraction and revision by image analysis using existing geodata and knowledge: current status and steps towards operational systems”, In *ISPRS Journal of Photogrammetry & Remote Sensing 58 (2004)*, pp 129-151.

[4] C. Beumier, “Building Verification from Geometrical and Photometric Cues”, In *Applications of Digital Image Processing XXX*, Proceedings of SPIE Vol. 6696, San Diego, California (2007).

[5] C. Beumier, “Building Verification from Disparity of Contour Points”, In *Image Processing Theory, Tools & Applications (IPTA’08)*, Sousse, Tunisia, 23-26 Nov 2008, pp. 408-413.

[6] M. Idrissa, C. Beumier, “Quantitative Analysis of 3D Reconstruction Error in the Context of Computational Stereo in Remote Sensing”, In: *GEOgraphic Object-Based Image Analysis*, Ghent, Belgium (2010).

[7] X. Jin, C. Davis, “Automated Building Extraction from High-Resolution Satellite Imagery in Urban Areas Using Structural, Contextual and Spectral Information”, In: *EURASIP Journal on Applied Signal Processing*, pp. 2196-2206 (2005).

[8] S. Müller, D. Zaum, “Robust Building Detection in Aerial Images”, In *CMRT05, IAPRS, Vol XXXVI, Part 3/W24*, pp. 143-148 (2005).

[9] Lab color space, http://en.wikipedia.org/wiki/Lab_color_space

[10] A. Nobel, B. Lugosi, “Histogram Classification Using Vector Quantization”, In *Proceedings of IEEE International Symposium on Information Theory 1994*, p391.

[11] Machine learning Py, https://mlpy.fbk.eu/