Evaluation of the Automatic methods for Building Extraction

Julia Åhlén, Stefan Seipel and Fei Liu

***Abstract*—**Recognition of buildings is not a trivial task, yet highly demanded in many applications including augmented reality for mobile phones. Recognition rate can be increased significantly if building façade extraction will take place prior to the recognitionprocess. It is also a challenging task since eachbuilding can be viewed from different angles or under differentlighting conditions. Natural situation outdoor is when buildings are occluded by trees, street signs and other objects. This interferes for successful building façade recognition. In this paper we evaluate the knowledge based approach toautomatically segment out the whole buildingfaçade or major parts of thefaçade. This automatic building detection algorithm is then evaluated against other segmentation methods such as SIFT and vanishing point approach. This work contains two main steps: segmentation of building façades region using two different approaches and evaluation of the methods using database of reference features. Building recognition model (BRM) includes evaluation step that uses Chamfer metrics. BMR is then compared to vanishing points segmentation. In the evaluation mode, comparison of these two different segmentation methods is done using the data from ZuBuD.Reference matching is also done using Scale Invariant Feature Transform. Theresults show that the recognition rate is satisfactory for the BMR model and there is no need to extract the whole building façade for the successful recognition.

***Keywords*—**Building, extraction, recognition, Chamfer metrics, vanishing points, SIFT.

I. Introduction

Modern technology allows for incorporation of computational algorithms and devices into mobile phones and thus production of fast information retrieval,high-quality color displays, high-resolution digitalcameras, and real-time 3D graphics. The fact that the information can betransmitted over data connections and GPS provides for a myriad of applications being crated for mobile devices. Those include many types of services suchas navigation aid, weather reports, or a tool for restaurant guide.For most of these services the geographical location is an essentialpart, but that is not enough, in many applications aiming to be an augmented reality aid we are strongly dependent on the accuracy of the detected object. For example, a person can beinterested in finding information on the object that the devicepointing at, thus the object should be recognized. There is a reason to assume that we might target in several objects situated close to each other and thus sharing the same geographic location. In such situation we need to be more precise in what exact object we would like to subject for augmentation. In this paper wefocus on buildings as a target for recognition. Building recognitioncan be used in various kinds of applications, including surveillance [1], 3-D city reconstruction, real-time mobile device navigation [2], and robot localization [3].

A number of building recognition systems have been proposed in recent years. However, most of them are based on a complex feature extraction process. Buildings are hard to define since no obvious descriptors can be defined. A human observer easily recognizes the differences between a building and a box with drawers. For a computer vision those two objects have similar qualities, rectangular shape, smaller rectangular shapes inside and homogeneity in colors, at least in most of the cases. That makes it a very challenging task to define a set of specific feature descriptors for a building. Generally, most of the existing building recognition systems adopt a complex feature extraction process to represent animage. For example, both global features, shape [4], texture [5], and local features such as SIFT (Scale Invariant Feature Transform) and SURF [6] are integrated to obtain satisfactory performance. Using more features may bring better results [7], however, it also means the feature representation requires more computational cost and is not easy to implement. In light of this, we investigated whether there is a simple way for feature extraction in the building recognition task.

A common approach to segment an object from images is to use a prototype shape, and search for it in the image. This leads to the task of shape matching, which has numerous applications, such as object localization, image retrieval, model registration, and tracking. One way to represent a shape is by a set of feature points, for example edges. In order to match two shapes, point correspondence on the two shapes has to be established.

Generally, there is always some knowledge about the building that can be coded in a set of feature descriptors; however these features can easily produce false matching results in scenes containing similarly shaped and colored objects. Thus we need to set some knowledge about buildings in outdoor science. We suggest creating a set of rules that describes typical surroundings of the building to successfully segment it out. In this paper, we evaluate the performance of a straight forward building recognition model (BRM), where the building of interest is extracted from the rest of the image based on global image characteristics. To compare the recognition rate of the suggested model we compare it to the other building extraction algorithms using a database of

This work was supported in by University of Gävle (Sweden).

reference images, which are vertical edge maps of buildings in question. The fact that only buildings or building pars are present after the first step will significantly improve matching rate, since no other or very few other interfering objects are compared.We will compare the performance of the model with the well-known SIFT method and vanishing point approach.

This paper is organized as follows. In Section II, we review related work on building recognition and shape extraction. In Section III we present the newly proposed model for building recognition (BRM) in detail. In Section IV, we evaluate the performance of BRM. Section V concludes the paper and provides discussion.

II. Related Work

Existing building recognition systems can be roughlydivided into two categories: clustering-based methodsand feature representation-based algorithms. Clustering-basedmethods aim to discover the relationships among different image structures by grouping them into different clusters. Zhang and Koseck´a [8] proposed a building recognition system based on vanishing point detection and localized color histograms. Detected line segments are grouped into dominant vanishing directions and vanishing points are estimated by the expectation maximization (EM) algorithm. After that, image pixels satisfying some certain constraints will be divided into three groups, namely left, right, and vertical and localized color histograms will only be computed on these pixels. Because of the fast indexing step using localized color histograms, this method achieved some improvement in efficiency and has attracted the most attention, however, it is hard to implement when extracting building façade with significant occlusions. Another approach using vanishing points is described in [9] and it is based on the observation that façades are image regions with repetitive patterns containing a large amount of vertical and horizontal line segments. Firstly, scan lines are constructed from vanishing points and center points of image line segments. Hue profiles along these lines are then analyzed and used to decompose the image into rectilinear patches with similar repetitive patterns. Patches are then merged into larger coherent regions and the main building façade is chosen based on the occurrence of horizontal and vertical line segments within each of the merged regions.

Feature representation-based algorithms focus on the process of feature extraction in building recognition.Hutchings and Mayol [10] designed a building recognition system for mobile devices to serve as a tourist guide in the world space. Given a query image, its local features are extracted and described by the Harris corner detector [11] and the SIFT descriptor, respectively.A SIFT keypointis a circular image region with an orientation. It is described by a geometric frameof four parameters: the keypoint center coordinates, its scale (the radius of the region), and its orientation (an angle expressed in radians). The SIFT detector uses as keypoints image structures which resemble “blobs”. By searching for blobs at multiple scales and positions, the SIFT detector is invariant (or, more accurately, covariant) to translation, rotations, and re scaling of the image [12]. For the building detection where we may have images of the same building taken from different location, we may face non-linear changes, which can be impossible to detect by SIFT.In the matching process for mobile applications, a scale can be selected for each query image according to its GPS position. This results in the reduction of search space and the computational cost. However, the system fails in dealing with large viewpoint changes. Another drawback for such method is insufficiency when applied on data with non-linear changes and non-static occlusions, such as moving cars.

Some models for building recognition are simply using local orientation for feature definition [13]. The described model is very simple; however, it offers a modular, computationally efficient, and effective alternative to other building recognition techniques.

Proposed decades ago, Chamfer matching remains to be the preferred method when speed and accuracy are considered. Chamfer matching was first proposed by Barrow et al [14] and improved versions have been used for object recognition and contour alignment. The basic idea is that given two sets of points whereas. $U=u_i$and $V=v_j$are template and query image respectively. Chamfer distance between each point $u_i \in U$ and its closest edge in V as in (1). ‘

$$d_{Ch}(U,V) = \frac{1}{n}\sum_{u_i \in U} \min_{v_j \in V} |u_i - v_j| \quad (1)$$

The template image $U=u_i$is superimposed on the distance image $V=v_j$. An average of the pixel valuesthat the template hits is the measure of correspondence between the edges, called the edge distance. Aperfect fit between the two edges will result in edgedistance zero, as each template point will then hit anedge pixel. The actual matching consists of minimizingthe edge distance. There are many variants ofmatching measure averages, e.g. arithmetic, root meansquare and median.

When using a single template, chamfer matchingcannot handle large shape variations. The chamferdistance is not invariant in regard to translation, rotationor scale. Furthermore, the number of templatesneeded increases with object complexity. Each ofthese cases has to be handled by matching with differenttemplates. In scenes with cluttered building façadesthe chamfer cost function will typically have severallocal minima. In order to make a decision aboutthe object location, orientation and scale, it may be necessary to use a subsequent verification stage [15].Scalable Vocabulary Tree (SVT) algorithms are tested in [16] and a very good performance is presented,however partial occlusion of the building causes thedistribution of features to change, thus affecting theentropy based scoring metric, as well as the SVM training.This fact as well as the requirement of a muchlarger data set for the improved performance will makethis approach not suitable for our study.

III. Building Extraction and Evaluation

This section describes data and presents the algorithm for building extraction.

A. Data Description

Images used to test and develop our method are taken from Zurich Building Image Database (ZuBuD), which is acquired

and prepared by the Department of Information Technology and Electrical Engineering -Computer Vision Laboratory in Zurich, Switzerland.All 1005 images of ZuBuD database were captured by digital cameras of resolution 640x480 without flash. This database contains, for each building, five images were acquired at random arbitrary viewpoints.

Examples of various buildings from ZuBuD used for tests are given in Fig 1. Two different view angles are used for each tested building.

Fig.1 Zubud images with two viewing angles

B. Algorithm Description

The approach suggested in this work emerges from assumption that images of buildings contain quite large areas of sky and in many cases large areas of street pavement or other street coverage below the building façade. Using this assumption the building extraction module processes as described in [17]. Shortly, the approach can be described as a series of the following steps: extraction of sky region, extraction of street coverage, test of the remaining areas to fit a criterion of intensity values and locale position. The last step is an evaluation of the model by matching the found region to an existing database of building features. This last step uses Chamfer metrics. This method istime efficient and robust since there are only few pre-requisites needed in order for this algorithm to work successfully. One is the presence of sky containing any blue color and the other is absence of buildings colored with bluish tones. The evaluation step through Chamfer metrics is adequate since the angle views of tested images are quite similar and thus do not produce very different coefficients when matched to a reference image.

C. Evaluation

Here we suggest evaluating the extraction of building façades by comparison of Chamfer matching explained in the previous section with building detection using SIFT and vanishing point detection described in previous section.All these methods are tested on a set of 50 images from the ZuBuD database. There are 23 buildings with 2 view angles and 2 buildings with 3 view angles. We deliberately avoided situations where the detection is difficult due to severe occlusion caused by trees and cars. For the test database containing 50 images with buildings we created 24 references, where each one represents one particular building. Calculation of vertical edge map is done on all the references. By using one reference for different angles of view of the building in question we put the BRM to a test when the exact match is not possible due to some unknown transforms, which occurs when a userrandomly changes photo shooting position.

The process of evaluation is as follows. We extract buildings using the two algorithms described above and in previous sections and then run recognition process using reference images from database. The test creates 1200 coefficients, which we analyze using basic statistics.

In case of SIFT we calculate keypoints in both the tested image and the reference. We do not automatically extract the building using this method; we only apply keypoints on the references available and the query image, which is already the extracted building region. Thenpercentages of the matching key points are calculated. For instance, if we would calculate matching percentage of an edge map with the reference image containing exactly the same edge map, we would find thatkey pointmatching exhibits 100%. However, such situation is unlikely to occur thus we need to create an allowable fluctuation interval for the matching results. It is reasonable to test with quite low threshold in order to avoid false matching, thus we set it to 30%.This means that in case of match of 70% and higher we consider the image to correspond to a building in the reference edge map. If this value occurs we analyze the result in order to determine if the query image contains the reference.

When we test the recognition rate in the images created by building extraction algorithm that uses vanishing points we calculate the amount of reference image outside the found region. This approach produces match in cases where we have 0% outside the detected area. Because of small differences caused by different viewing locations of the buildings we need to create an interval of percentages that can be considered as a match. We set it equal to 4%. Running the recognition procedure using the references we analyze all the images were the result is calculated to 4% or less. Those images produces match in 68% of the cases, however, only 39% are real matches and 29% are false matches. This means that in 29% of positive results the reference image is present in the query image.

Chamfermatching is used in BRM as the method to decide the recognition rate. As in the other two cases we use low level featuressuch as vertical edge map of the building, which wethen subject to distance transform. We can defend the idea to extract only vertical edges since the distances within windows of the building will not change significantly when taking image of a that building from the convenient distances and locations. In most cases person will choose only few such positions. Some skewnessand rotation will not influence the matching outcomebecause the distance in window frames will still be thesame after those linear transformations. Using this assumption, we match edge points of the detected building with the references stored in the database. This step will create

50 Chamfer distance coefficients for each tested reference image. The Chamfer distance value is calculated using threshold toavoid outliers and in some cases missing edges.Those coefficients are analyzed in order to extract the real matches and the false matches, which can occur when building are similar regarding the defined features.

IV. Results

We use three different approaches to automatically detect buildings in outdoor environments. In all three algorithms both color and local positions are used as feature descriptors. In BRM we segment out the building or parts of the building and then calculate an edge map of the found region. The comparison of the found object is done using a reference edge map. In the Table 1 we can see the results of running the whole database of references on all 50 images containing buildings.

Table 1. Comparison results

Methods	Match	False match	No match
SIFT	25%	13%	62%
Chamfer matching	88%	10%	2%
Comparison to vanishing point extraction method	39%	32%	29%

As we can see in case of Chamfer matching, which is used in BRM, we get a quite high recognition rate, which is 98%, and however, 10% is a false matching. False matching refers to situations where the system recognizes the building although it is not the building defined by the reference. Visual result is Fig. 2 shows the original image, the reference and the Chamfer distance points, which are the correspondences of*i'th* edge point in the template and the detected building façade edge map. In this particular image the sum of Chamfer distance coefficient is calculated to 0.19, which is within the defined interval of allowed values.

Fig. 2 From left to right: Original image, detected part of the façade, Chamfer matching visualization.

As an illustration of false match we show Fig. 3 where an image with a building is searched using the reference representing some other building. Yet the resulting coefficient indicates a match. This situation takes place when the searched building is so inaccurately extracted that the defined features are recognizable in other image locations then the correct ones.

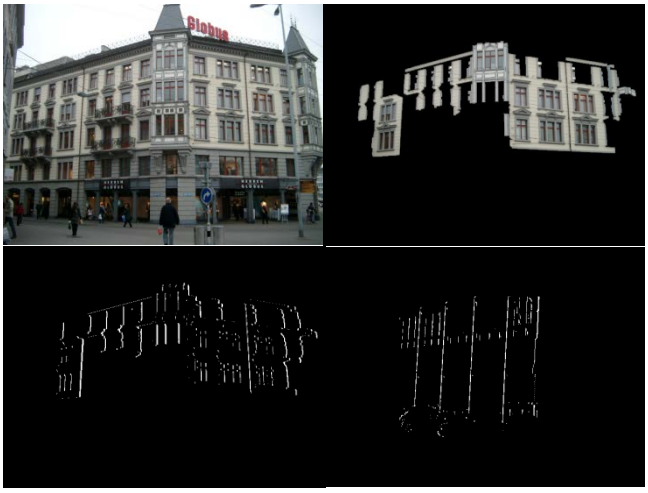

Fig. 3From left to right first row: Original image, detected part of the façade. From left to right second row: edge map of the detected building, reference that produces false match.

In case of SIFT as a matching method, we get a quite unsatisfactory results. We tested all 50 images to the reference images separately, which means that we produced a matrix with a size of 50x24 with the matching percentages calculated by keypoint comparison. The matching percent is calculated to38% and false matching is about 13%. No match situation is about 42%, which mean that the reference edge map could not exhibit the same key points as the found building parts. We can see an example of keypoints measured on the extracted region and the reference edge map with the key points calculated and added, in Fig.4. In most of the false matches quite a big amount of keypoints from reference image are the same position as in query image, which does not contain the same building as the reference. No match number, as can be seen in Table1, is very high. Since we are testing building at two different angles of view and then calculate edge maps, we will face an input that is quite sensitive to a slightest change of the position coordinate, thus leading to a totally different keypoint's localization for the same building at alliterate angle of view. In Fig. 4 we show the image falsely detected by the reference image. Reference image contains a different building that the query image referred to as the Original.

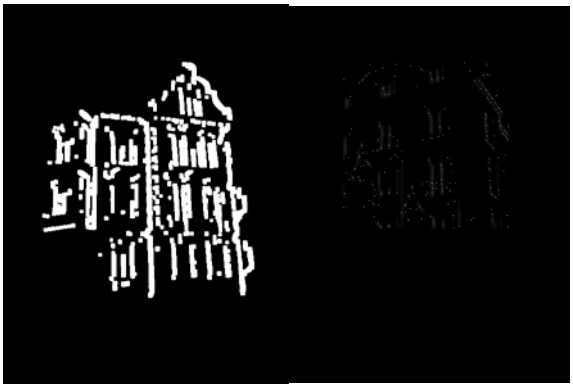

Fig.4 From left to right first row: Original image,façade detected by BRM, Edge map of the reference image. From left to right second row: keypoints of the detected edge map building, keypoints of the reference image.

Testing 50 images with 24 references is very time consuming using SIFT. This is another drawback encountered during the evaluation.

For the vanishing points method of extraction we used the below described approach for determining recognition efficiency. Although there are quite a large areas supposed to represent a building are found in the imageit is not implying that reference map will be fully covered by this region. As a result we calculate percentage of reference edge map that is outside the found area and thus indicates the robustness of building detection algorithm. In perfect match 0% would be a matching result. We analyzed all the images were the result is calculated to 4% or less. Those images produces match in 71% of the cases, however, only 39% are real matches and 32% are false matches. This means that in 32% of positive results the reference image is not present in the query image.One correctmatching example is illustrated in Fig. 5.

Fig.5. From left to right: Original image, region of the facade detected by the vanishing points method, reference image.

In Fig.6 we show en example of false detection of a building extracted by vanishing point method. In the first row we see the building that we search for in our database of 50 images. We even show the extracted part of that building using vanishing points method. In the second row we see the building that is recognised as a buidling from the first row. This result can be interpreted as a false match caused by the insuficient amount of descriptive features or/and incorrectly detected parts of the façade, which is clearly illustrated in the second row of Fig. 6. We see numerouse false mathing for each of the reference edge maps, when they are applied on a vanishing point segmented building regions.

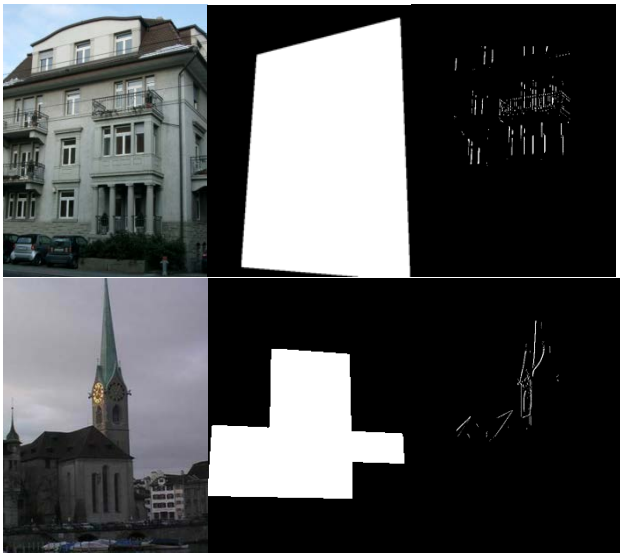

Fig.6 From left to right first row: Original image, reference representing the original image, Edge map of the corresponding reference image. From left to right second row: falsely detected building by the mask from the first row, vanishing point building detection region, reference of the building falsely detected.

In this section we presented evaluation results and showed figures that illustrates the robustness of the two methods of building extraction. BRM on the tested image database shows superior bevavior than the other tested method based on the vanishing points' calculation. Using SIFT as a mathing model gave non-satisfactory results where 62% of the database could not be recognised.

.

V. Conclusion

We have tested two methods for automatic extraction of building facades. Using the results from previous section as a base we can conclude that despite of its simplicity, the Building Recognition Model is more robust than vanishing points approach when tested on the ZuBuD database. We also could show the sensitivity of SIFT method for slightest transform of the objects in images. We tested the performance with reference image representing all available angle views of that building, we could encounter results which reduce the amount of found objects, however, we can defend this choice by the fact that the reference database is significantly smaller thus less calculation time. Studying the results presented in the Related Work section, we see that using references of all thinkable transforms of the searched object does not lead to a successful search in all situations. Our idea of using just one reference image for searching all different views comes from the assumption that in most outdoor situations there are not many places that allows for a good viewing of the building

façade, which should be present as a whole and yet being large enough to discern small details on it. In light of that we can suggest BRM in situations where the following criteria are met:

1) The image of building we wish to extract contains some areas of sky.
2) There is recognizable part of the street below the building façade.
3) Angle of view does not change dramatically, which means no more than 20 degrees of rotation in horizontal path and no scaling

The above described criteria cannot be applied on the SIFT matching method since even a slightest change in object location will lead to a false result or a non-matching result. In this method we would need to search the exact correspondence of the transform in the reference database. This method for matching features is not suitable when automatic extraction of the building façade is done using the described BRM.

The vanishing point method can be the most successful in cases where there is no occlusion or other interfering objects that lie on the same line as building and exhibit the similar histogram profile characteristics as a façade in question. From the results obtained in this study we see that references of edge maps can be found in many images represented by the mask calculated with vanishing point method. The masks sometimes cover quite a big area in an image thus allowing for total overlay of different references. That makes it reasonable to assume that the edge map references are not suitable to use as a feature descriptions for that method. It is also clear that refining the building extraction process will produce a better recognition rate.

Finally, we have shown that for the recognition of the building we do not need to automatically extract the whole region of that object, we just need some part of it so that the features defined in the reference database will correspond to the features in the extracted part of the façade. The process of feature description that is robust and time efficient is a next step of this project.

ACKNOWLEDGMENT

J. Åhlén thanks Computer Science department staff at the University of Gävle for sharing valuable ideas with author.

REFERENCES

[1] D. Bruckner, C. Picus, R. Velik, W. Herzner,and G. Zucker, "Hierarchical semantic processing architecture for smart sensors in surveillance networks",*IEEE Trans. Ind. Inf.*,2012, 8(2),
[2] H. Ali, G. Paar, and L. Paletta, "Semantic indexing for visual recognition of buildings", *in Proc Int. Symp. Mobile Mapping Technol.*, 2007, pp.28-31.
[3] M. M. Ullah, A. Pronobis, B. Caputo, J. Luo,R. Jensfelt, and H. I. Christensen, "Towards robust place recognition for robot localization", *in Proc. IEEE Int. Conf. Robot. Autom.*, 2008, pp.530-537.
[4] A. Jain and A. Vailaya, "Image Retrieval UsingColor and Shape", *Pattern Recogn.*, 29(8),1996, pp. 1233-1244.
[5] W. Robson Schwartz, F. Roberti de Siqueira and H. Pedrini ,"Evaluation of feature descriptors for texture classification", *J. Electron. Imaging*.,21(2), 2012, pp. 023016-023016-17.
[6] C. Valgren, A.J. Lilienthal, "SIFT, SURF & seasons: Appearance-based long-term localization in outdoor environments", *Robotics and Autonomous Systems*, 58(2), 2012, pp. 149-156.
[7] S. Chen, J. Zhang, Y. Li, and J. Zhang, "A hierarchicalmodel incorporating segmented regionsand pixel descriptors for video backgroundsubtraction", *IEEE Trans. Ind. Inf*., 8(1), 2012,pp 118-127.
[8] W. Zhang and J. Kosecka, "Hierarchical buildingrecognition", *Image Vis. Comput*.,2007, 25(5),pp. 704-716.
[9] F. Liu and S.Seipel, "Detection of Façade Regions in Street View Images from Split-and-Merge of Perspective Patches", unpublished.
[10] R. Hutchings and W. Mayol-Cuevas, "Building recognition for mobile devices: Incorporating positional information with visual features", *Comput. Sci., Univ*. Bristol, Bristol, U.K., Tech. Rep., 2005, CSTR-06-017.
[11] C. Harris and M. Stephens, "A combined corner and edge detector", *in Proc. Alvey Vis. Conf*., 1988, pp. 147-151.
[12] David G. Lowe, "Distinctive image features from scale-invariant keypoints,"*International Journal of Computer Vision*,60, 2 ,2004, pp. 91-110.
[13] S. Lee, N. Allinson, "Building RecognitionUsing Local Oriented Features", *IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS*.9, 3, 2013, pp. 1697-1704.
[14] H. G. Barrow, J. M. Tenenbaum, R. C. Bolles,and H. C. Wolf, "Parametric correspondenceand chamfer matching: Two new techniques forimage matching", *In Proc. 5th Int. Joint Conf. Artificial Intelligence.*, 1977, pp. 659-663.
[15] A. Thayananthan, B. Stenger, P.H.S. Torr and R. Cipolla, "Shape Context and Chamfer Matching in Cluttered Scenes", *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2003, 1, pp. 127-133.
[16] V. Chandrasekhar, C.W. Chen, G. Takacs, "Robust Building Identification for Mobile Augmented Reality", *Lecture Notes in Electrical Engineering (LNEE)*, 253, 2013, pp. 923-931.
[17] J.Åhlén and S.Seipel, "Knowledge Based Single Building Extraction and Recognition",*Proc. 8th WSEAS International Conference on Computer Engineering and Applications (CEA '14)*, pp.29-35.