

## HIERARCHICAL REPETITION EXTRACTION FOR BUILDING FAÇADE RECONSTRUCTION FROM OBLIQUE AERIAL IMAGES

Jing Zhang<sup>a</sup>, Huang Yao<sup>b, \*</sup>, Wanshou Jiang<sup>a</sup>, Xin Shen<sup>c</sup>

<sup>a</sup> State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, 129 Luoyu Rd., Wuhan 430079, China - (jing.zhang, jws)@whu.edu.cn

<sup>b</sup> School of Educational Information Technology, Central China Normal University, 129 Luoyu Rd., Wuhan 430079, China, - yaohuang@mail.ccnu.edu.cn

<sup>c</sup> Collaborative Innovation Centre of Geospatial Technology, Wuhan University, 129 Luoyu Rd., Wuhan 430079, China - xinshen@whu.edu.cn

WG IV/7 and WG V/4

**KEY WORDS:** Oblique aerial image, Building façade, Partition, Hierarchical repetition, Joint reconstruction

### ABSTRACT:

In this paper we introduce an approach for automatic recognition and reconstruction of building facade structure from oblique aerial images. Contrast to street-view image oblique aerial image has larger field of view but lower resolution, weaker texture and more noise. To overcome these shortcomings, our approach firstly analyses the horizontal distribution density to extract individual facade area from image. Then a hierarchical repetition detection method is employed to partition the facade and recognize structural elements. Finally, the geometry structure of each facade element is reconstructed jointly by all repetitive image tiles. Results show the potentials of the proposed approach.

### 1. INTRODUCTION

Accurate and realistic 3-dimensional models of our urban environment are gaining growing interest for city planning, civil security, navigational aids and entertainment industry. Recently, oblique aerial images become more popular in photogrammetry. These images are taken from aeroplane from tilted angles. Thus they can present the information of both building roofs and façades. Moreover, oblique images are usually captured from multiple directions with specified overlap. Consequently, a certain building could be viewed in multiple images from different perspectives, which means less occlusion and more redundancy for 3D geometry recovery. These unique advantages make oblique aerial images an ideal source for building modelling. Nowadays, a significant progress of stereo matching technique has been achieved (Furukawa and Ponce, 2007; Hirschmuller, 2005). And some commercial software directly utilizes this technique to produce textured surface mesh models for urban scene (Street Factory; Smart3DCapture). The mesh model of building facade has acceptable visual effect but lack of detailed geometric structures and contextual information, which are great valuable in accurate city modelling and scene recognition (Jianxiong, 2012). In this paper, we attempt to develop an automatic pipeline to modelling building facades with abundant details and accurate geometry from oblique aerial images. Although facade reconstruction is a classical problem and deeply studied in Computer Vision (CV) and Computer Graphics (CG) for many years, little research has been carried out on complex facade modelling from oblique images. In the

next section we will give a short review of related research, and describe the special challenge encountered in oblique aerial image.

### 2. RELATED WORKS

Building façade reconstruction is a very hot topic in CV & CG over the last few years (Musialski et al., 2012). In the earlier research, data-driven methods are quite popular (Akbarzadeh et al., 2006; Pollefeys et al., 2008), which directly extract geometrical features from initial data and utilize these features to partition facade and model façade structures in a bottom-up mode. The general drawback of these methods is depending on data quality and relatively sensitive to erroneous or incomplete data. Then grammar-driven methods are introduced to improve reconstructing quality (Parish and Muller, 2001; Ripperda and Brenner, 2009). This kind technique was firstly applied in computer graphics and termed as Procedural Modelling (PM). It employs high-level architectural knowledge to decompose façades deriving a set of structural rules. Subsequently, these rules are expressed by a shape grammar to generate the façade models. Generally, PM needs some man-made grammar schemes to represent specific façade models and mainly used to synthesize realistic looking urban model. Later, Inverse Procedural Modelling (IPM) method is received significant attention (Becker, 2009; Dai et al., 2012; Muller et al., 2007; Simon et al., 2012; Teboul et al., 2010). Different from PM, IPM discover structural rules from input data, then use these rules to reconstruct building models in the PM manner. IPM is

---

\* Corresponding author.

regarded as a better roadmap for façade reconstruction than pure data-driven or grammar-driven methods which can combine observations and domain knowledge efficiently. While the key of IPM is how to derive reasonable rules from raw data. Although numerous methods have been proposed, it is far away from practical application on real-world images due to the facades' prohibitively high complexity and noisy, incomplete measurements.

We aim at a reliable and high-quality façade reconstruction from oblique aerial image in urban area. Contrasting to street view image, oblique aerial image can provide larger view of ground area and higher efficiency of acquisition, this character of oblique images require a method to detect facade regions from patches of roof, ground and tree mixed in one image frame. Then oblique image has lower resolution and weaker texture, which requires a robust method insensitive to erroneous and incomplete measurements. Moreover, building façades in practical urban scene are various and complex. Traditional methods conduct building façade as regular arrangement of a single type window or balcony just like a checkerboard pattern. While real-world façade usually contains various types of windows and balconies which exhibit an amazing variety of irregular mixtures of grid structures. The ubiquity of these facades requires a new method supporting such complex façades modelling.

We propose a novel approach to reconstruct irregular building façade by hierarchical repetition analysis. Assuming building façade is composed of a perpendicular dominant plane and some auxiliary elements like different types of windows, balconies or oriels. The target of façade reconstructing is to recover the elements' geometric structures and align them to proper position. To achieve the general objective, we attempt to tackle three main problems as following:

- (1) To detect individual facade dominant plane from oblique image.
- (2) To discover elements on facade and recover their spatial layout.
- (3) To reconstruct 3D geometry of each element.

In the next section, we apply a series of technique to solve these three problems and then integrate them into an automatic pipeline.

### 3. METHODOLOGY

According to previously mentioned idea, we design a three-stage pipeline to reconstruct irregular façade automatically from oblique images. The detailed implementation flow is described in Figure 1.

The whole flow of façade reconstruction is divided into three main steps: facade detection (grey blocks), facade partitioning (orange blocks) and facade reconstruction (dark green blocks).

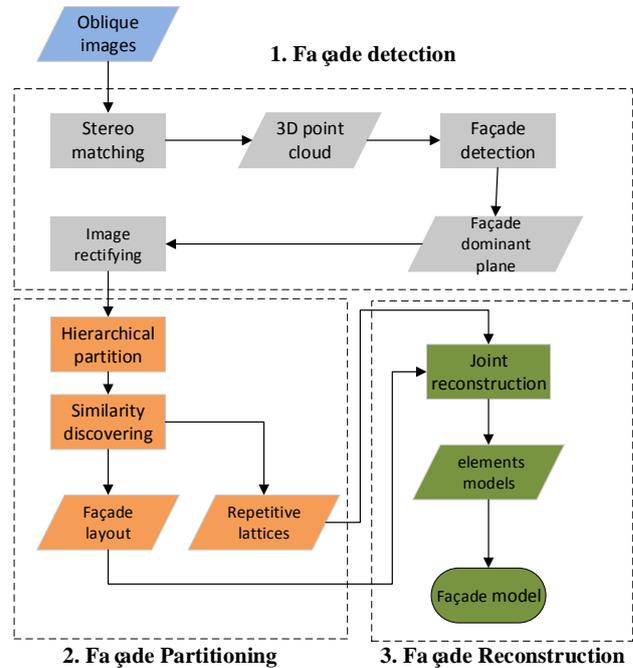


Figure 1 The whole flow of façade reconstruction from oblique images (rectangle represents a processing node, parallelogram represents a data node).

#### 3.1 Facade Detection

With larger field of view than street view image, oblique aerial image covers a wide scope of urban area which contains a variety of land cover such as buildings, roads, vehicles and vegetation. In the first step, we need to detect facade from oblique image automatically. Most existing work on facade identification adopt image classification technique, which relies on a series of well-designed procedure of feature selection and training (Ceylan et al., 2012; Liu and Liu, 2014). Considering that building in urban scene generally juts from the ground and the facade is mostly perpendicular, we design a simple way to identify facade by this 3D priori knowledge.

Firstly, Semi-Global Matching (SGM) method (Hirschmuller, 2005) is used to perform pixel-wise stereo matching to generate 3D point cloud from oblique images. Considering building façades are mostly perpendicular, the obtained 3D points on façades have approximate planar coordinates. We calculate their horizontal distribution density according to formula(1). For each point  $i$ , the horizontal distance between  $i$  and its neighbour points is calculated, then we sum the number of point which horizontal distance is less than a specified distance threshold  $d_{Th}$  as point  $i$  horizontal distribution density  $D_i$ . Then a simple density threshold is used to distinguish points belong to façade areas. In our experiment, the distance threshold is set to 0.3m, and the density threshold is set according to the image GSD and the expected facade minimal height. For example, our oblique image's GSD is 20cm, and we want extract the facade higher than 3 meters, so we set density threshold to 15.

$$D_i = \sum \left( \|p_i - p_j\|_{xoy} \leq d_{Th} \right) \quad (1)$$

After that, the labelled façade points should be divided into individual façade and then the dominant plane for each façade should be extracted.

Assuming facade is generally perpendicular, we project all labelled facade points onto horizontal plane and cast the 3D points segmentation to a 2D straight-line fitting problem. We utilize a robust multiple structures estimation method (J-Linkage) to extract all straight line segments (Toldo and Fusiello, 2008). And then use the resulting lines and the maximum and minimum height of facade points to fit the parameters of 3D dominant planes. Subsequently, the dominant planes are projected to oblique image in order to label the corresponding facade image areas. Then a Transform Invariant Low-Rank Textures (TILT) method (Zhang et al., 2011) is used to ortho-rectify each facade image to the frontal view. Figure 2 shows the flow of facade detection.

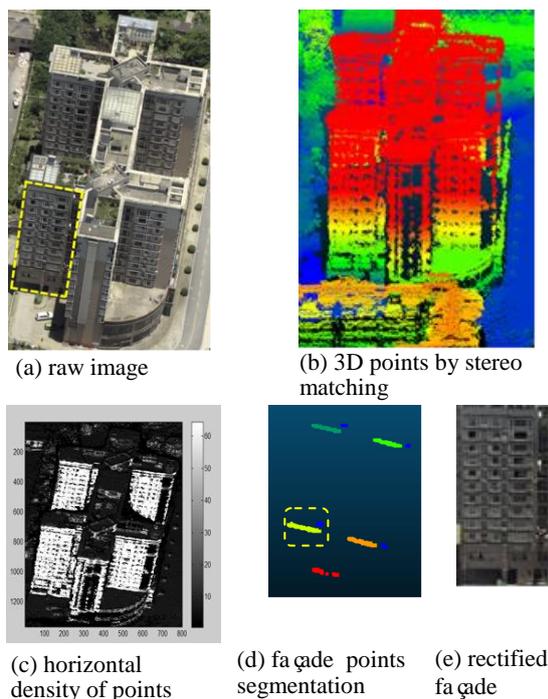


Figure 2 (a) is a part of raw oblique image. (b) is the result 3D points by stereo matching, the points are colored by height ramp from blue to red. (c) shows the points colored by horizontal density. (d) shows the segmentation result of facade points, different color indicates different line segment, blue points are the outliers. (e) is the rectified facade image corresponding to the area marked by a yellow dashed box in (a) and (c).

### 3.2 Facade Partitioning

Considering that building facade in urban scene generally contains a lot of repetitive structures, we intend to utilize such property to partition the facade and eventually identify facade elements and their arrangement. However, this work is challenging due to poor-quality of oblique image and complex layout of facade. We hold the view that if the regular facade can be described by a simple repetitive pattern, the irregular also can be described by a series of simple repetitive patterns. Therefore, we design a hierarchical strategy to split the irregular facade into a set of simple repetitive patterns.

The building facade is expected to be partitioned into a set of horizontal and vertical slices. Different from the traditional method which partitions facade along horizontal and vertical directions simultaneously (Muller et al., 2007), we split facade once along each direction successively. According to prior

knowledge of urban buildings, the horizontal (vertical) splitting places of facade should be placed where vertical (horizontal) edges are rare and horizontal (vertical) edges are dense. After each splitting, a group of image slices can be obtained. Then an image registration method is applied to these slices image to discover the similarity among slices. We use the gradient mutual information to measure the similarity which is proposed robust to noises (Dalal and Triggs, 2005). Then we label the repetitive slices and combine them into a new image. Subsequently, the image splitting and repetition detecting is executed again on the new image along the other direction. The processing is performed iteratively until each slice is small than a certain area. Our partition approach ensures every splitting is based on a certain repetitive pattern determined in last similarity detection. So the whole flow can partition facade hierarchically. By the iterative splitting, the whole facade is partitioned to a group of tiles. Assuming each facade element has a unique repetitive pattern, we classify all the tiles to individual cluster according to their labels marked previously. After that, we utilize similarity repetitive patterns to perform a global optimization toward alignment of repetitive tiles. Firstly, each tile in a cluster is selected as an image template and the template based image match is conduct on the whole facade image. The extreme points of each template matching are extracted and all these extreme points are combined together. Finally, the least-square adjustment is adopt to refine the coordinates of each tile.

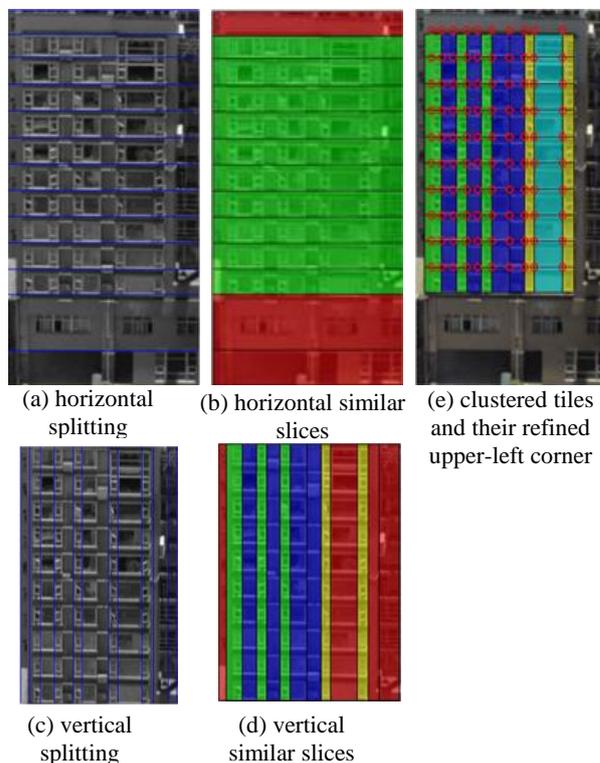


Figure 3 (a) shows the result of facade splitting along horizontal direction. Blue lines indicate the segmentation position. (b) shows the result of similarity detection along horizontal direction. Similar slices are filled with the same color and the slice different from others is colored in red. (c) shows the result of vertical splitting on the similar slices determined in horizontal direction (green area in (b)). (d) shows the similar slices in vertical direction. (e) shows the result of the complete facade partition. Tiles in different cluster are painted in different

color. The red rounds indicated the refined coordinates of each tile.

### 3.3 Facade Reconstruction

The geometry structure of facade element can be represented by a group of line features. In this section, we aim at recovering the wire-frame model for each element. The direct idea is to detect lines from image and group them together. We use a fast line segment detector named LSD to extract line segments (von Gioi et al., 2010).

Due to the low quality of oblique image, it is very common that line detecting in some image tiles cannot achieve desired result. Because repetitive tiles corresponding to an identical element, we integrate all the lines in repetitive tiles to reconstruct the element model. For each tile cluster corresponding to an element, the whole facade structure can be represented by a few element models. The coordinates of repetitive tiles are aligned in the previous step. We can use the relative position among tiles to map all lines in each tile to a unified reference coordinate frame. Then a two-step filtering process is performed to eliminate some non-structural lines. At first step, we suppose that element model is only consisted of horizontal and vertical lines (Manhattan assumption), so we discard non-horizontal and non-vertical lines with a criteria indicating the slope of line, denoted as  $S$ . We abandon those lines with  $S$  less than a predefined threshold  $\tau$  (0.0875 in our experiment, which means any line segment deviate more than  $5^\circ$  away from vertical and horizontal directions will be removed). Then the remainder lines which are nearby to each other (two pixels away on the orthogonal direction in our experiment), will be marked as a group of similar line segments.

$$S = \begin{cases} \frac{|y_{end} - y_{start}|}{|x_{end} - x_{start}|} & |y_{end} - y_{start}| < |x_{end} - x_{start}| \\ \frac{|x_{end} - x_{start}|}{|y_{end} - y_{start}|} & |y_{end} - y_{start}| > |x_{end} - x_{start}| \end{cases} \quad (2)$$

At the second step, a non-maximum suppression process is used to find the correct position and length of line segments. For each group of similar lines, we calculate the image gradient along the orthogonal direction of each line. Then a scan process is started along the direction of similar lines on each possible position, the pixels' gradient value is accumulated within each scan line, the maximum absolute value of accumulation is assigned as the accurate position of the combined line. At last, we seek the point of intersection of all determined lines, and link the intersecting points on horizontal and vertical directions to construct a complete wire-frame model. The element reconstruction results of each processing step is shown in Figure 4. Figure 4(a) shows the result of line segment detection. Different colors indicate lines in different tile clusters, and the red lines do not belong to any cluster. In Figure 4(b), lines belonging to the same cluster are mapping to a unified reference coordinate system and overlap together. Then in Figure 4(c), the overlapping lines are swept by a two-step strategy. Figure 4(d) shows the reconstructed wire-frame models of corresponding 4 tile clusters.

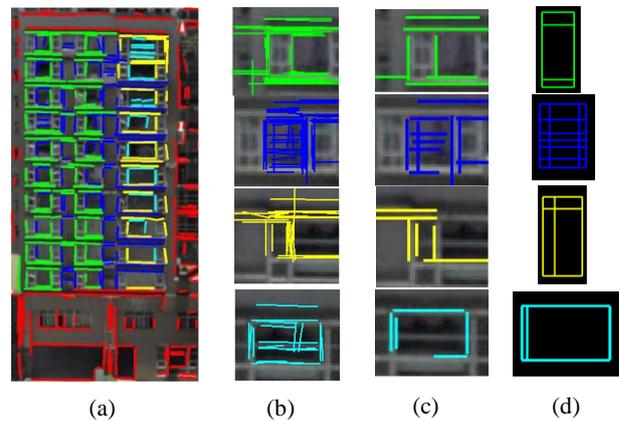


Figure 4 The results of element reconstruction

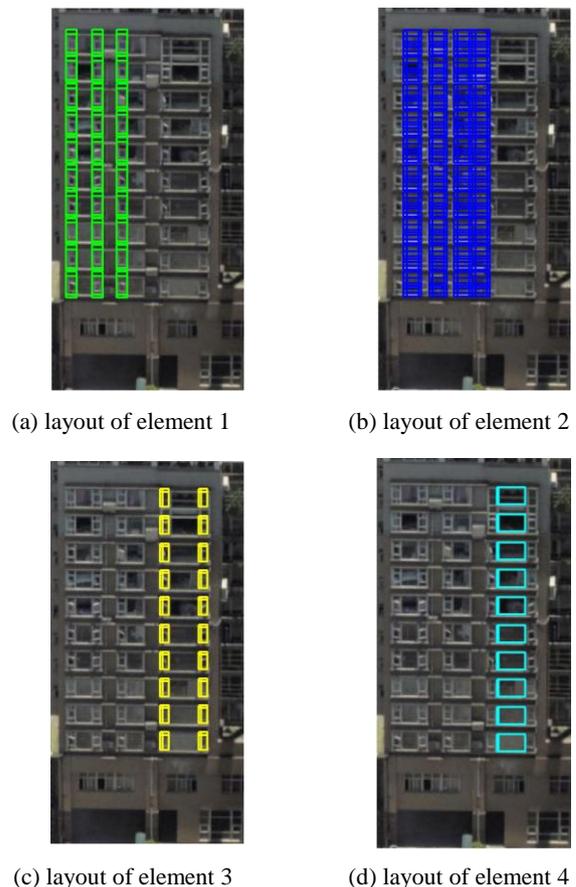


Figure 5 The layout of each element

The wire-frame model of each element is then arranged on facade image according to the coordinates determined in section 3.2 (Figure 5).

## 4. CONCLUSION AND FUTURE WORK

In this paper, we have presented an automatic approach for the recognition and reconstruction of building facade elements from oblique aerial images. Our approach recognizes the facade elements via repetitive structure detection. A hierarchical image splitting and similarity detection method is proposed to partition complex facade. Supposing similar image structures belong to the same facade element, we propose a joint modelling method

to reconstruct all repetitive structures to a uniform geometric element. Our approach is robust to low resolution, weak texture and noise. And the experiment shows that this method is effective to reconstruct complex facade models containing multiple elements.

However, due to our method partitions facade based on image gradient information, it may fail to distinguish wall and facade element if they have similar gradient. Therefore, further enhancement should be taken in future by introducing more high-level knowledge and constraints.

## ACKNOWLEDGEMENT

This research was partially supported by Chinese National Natural Science Foundation Program (NO. 41301367), the National Basic Research Development Program of China (No. 2012CB719904), the Key Laboratory of Geo-informatics of State Bureau of Surveying and Mapping Foundation Program (NO. 201309), and the Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing (No. 13key01).

## REFERENCES

- Akbarzadeh, A., Frahm, J.M., Mordohai, P., Clipp, B., Engels, C., Gallup, D., Merrell, P., Phelps, M., Sinha, S., Talton, B., Wang, L., Yang, Q., Stewenius, H., Yang, R., Welch, G., Towles, H., Nister, D., Pollefeys, M., 2006. *Towards Urban 3D Reconstruction from Video, 3D Data Processing, Visualization, and Transmission, Third International Symposium on*, pp. 1-8.
- Becker, S., 2009. *Generation and application of rules for quality dependent façade reconstruction. ISPRS Journal of Photogrammetry and Remote Sensing* 64, 640-653.
- Ceylan, D., Mitra, N.J., Li, H., Weise, T., Pauly, M., 2012. *Factored facade acquisition using symmetric line arrangements, Computer Graphics Forum. Wiley Online Library*, pp. 671-680.
- Dai, D., Prasad, M., Schmitt, G., Van Gool, L., 2012. *Learning domain knowledge for facade labelling, Computer Vision–ECCV 2012. Springer*, pp. 710-723.
- Dalal, N., Triggs, B., 2005. *Histograms of oriented gradients for human detection, Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, pp. 886-893 vol. 881.
- Furukawa, Y., Ponce, J., 2007. *Accurate, dense, and robust multi-view stereopsis. 2007 Ieee Conference on Computer Vision and Pattern Recognition, Vols 1-8*, 2118-2125.
- Hirschmuller, H., 2005. *Accurate and efficient stereo processing by semi-global matching and mutual information, Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on. IEEE*, pp. 807-814.
- Jianxiong, X., 2012. *3D reconstruction is not just a low-level task retrospect and survey. report*.
- Liu, J., Liu, Y., 2014. *Local Regularity-Driven City-Scale Facade Detection from Aerial Images, Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on. IEEE*, pp. 3778-3785.
- Muller, P., Zeng, G., Wonka, P., Gool, L.V., 2007. *Image-based procedural modeling of facades. ACM Trans. Graph.* 26, 85.
- Musialski, P., Wonka, P., Aliaga, D.G., Wimmer, M., Gool, L.v., Purgathofer, W., 2012. *A Survey of Urban Reconstruction, In EUROGRAPHICS 2012 State of the Art Reports. 2012. Eurographics Association*
- Parish, Y.I.H., Muller, P., 2001. *Procedural modeling of cities, Proceedings of the 28th annual conference on Computer graphics and interactive techniques. ACM*, pp. 301-308.
- Pollefeys, M., Nistér, D., Frahm, J.-M., Akbarzadeh, A., Mordohai, P., Clipp, B., Engels, C., Gallup, D., Kim, S.-J., Merrell, P., 2008. *Detailed real-time urban 3d reconstruction from video. International Journal of Computer Vision* 78, 143-167.
- Ripperda, N., Brenner, C., 2009. *Application of a Formal Grammar to Facade Reconstruction in Semiautomatic and Automatic Environments*.
- Simon, L., Teboul, O., Koutsourakis, P., Van Gool, L., Paragios, N., 2012. *Parameter-free/Pareto-driven procedural 3D reconstruction of buildings from ground-level sequences, Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on. IEEE*, pp. 518-525.
- Smart3DCapture. <http://www.acute3d.com/smart3dcapture/> (10 April 2015)
- Street Factory. <http://www.geo-airbusds.com/streetfactory/> (10 April 2015)
- Teboul, O., Simon, L., Koutsourakis, P., Paragios, N., 2010. *Segmentation of building facades using procedural shape priors. Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, 3105-3112.
- Toldo, R., Fusiello, A., 2008. *Robust multiple structures estimation with j-linkage, Computer Vision–ECCV 2008. Springer*, pp. 537-547.
- von Gioi, R.G., Jakubowicz, J., Morel, J.M., Randall, G., 2010. *LSD: A Fast Line Segment Detector with a False Detection Control. Pattern Analysis and Machine Intelligence, IEEE Transactions on* 32, 722-732.
- Zhang, Z., Liang, X., Ganesh, A., Ma, Y., 2011. *TILT: Transform Invariant Low-Rank Textures, in: Kimmel, R., Klette, R., Sugimoto, A. (Eds.), Computer Vision – ACCV 2010. Springer Berlin Heidelberg*, pp. 314-328.