2016 23rd International Conference on Pattern Recognition (ICPR)
Cancún Center, Cancún, México, December 4-8, 2016

Building Facade Recognition from Aerial Images using Delaunay Triangulation Induced Feature Perceptual Grouping

Xuebin Qin*, Martin Jagersand*, Xiucheng Yang† and Jun Wang‡
*Department of Computing Science, University of Alberta, Edmonton, Alberta, CA.
Email: {xuebin, mj7}@ualberta.ca
†ICube Laboratory, University of Strasbourg, Strasbourg, France. Email: xiucheng.yang@etu.unistra.fr
‡Institute of Remote Sensing and GIS, Peking University, Beijing, China. Email: wangjun.cau@foxmail.com

***Abstract*—This paper presents a novel feature grouping based framework for building facade recognition from aerial images. A combination of Maximally Stable Extremal Regions (MSERs and steered Determinant-of-Hessian (steered-DoH) are proposed to detect different shapes of blobs from images. Then we employ local parallelogram grouped by these repetitive and evenly distributed blobs to form an point-based regularity measurement. Building facade regions are indicated by these local parallelograms. In our work, we use Delaunay Triangulation (DT) to guide the search of local parallelograms. Our approach can handle images with large range of resolution. Vertical and horizontal assumptions of facades are not required. The experimental results conducted on images with different resolutions and different types of facades demonstrate superior performance on facade recognition both in terms of speed and accuracy ($\mathbf{F_1}-\mathbf{score}$ over 80%) over state-of-the-art methods.**

I. Introduction

With the development of photogrammetry and remote sensing, high resolution aerial images analysis has become a popular yet challenging area of computer vision research. Recognizing buildings is important, and facades are their essential defining features. Applications are in building modeling, navigation, damage assessment and emergency response.

A number of methods for building facades recognition have been proposed. They can be categorized as following three classes:

(1) *Stereo view or three dimensional (3D) point cloud assisted facade extraction.* Zebedin et al. [1] adopt an image-based optimization method to estimate the precise position of vertical facade planes in Digital Surface Model (DSM) reconstructed from aerial images. Meixner and Leberl [2] characterize building facades by mapping 3D facade key corners to 2D aerial images. Zhao et al. [3] and Delmerico et al. [4] extract facades from ground-level images. However, to obtain 3D point cloud from large-scale aerial images is time consuming and may fail on the repetitive structure of facades.

(2) *Edges or straight line segments based facade extraction.* Bansal et al. [5] use satellite images to find contours of building roofs, then map them to corresponding building roofs and their ground boundaries in aerial images by a homography to determine the facade regions. Liu et al. [6] recognize facade areas according to the edge-based vertical and horizontal regularity assumptions. Xiao et al. [7] also use vertical and horizontal lines to extract building facades. However, the assumption can be violated in case of poor data calibration, and for damaged or old buildings. Yang et al. [8] extract facade regions using straight line segments based multi-level feature extraction. This method judges line directions through cluster analysis instead of vertical and horizontal assumptions. However, easy failures on line segments detection makes it fragile.

(3) *Repetitive patterns based facade recognition.* Wendel et al. [9] employ intensity profile descriptors and a voting-based matcher to detect repetitive regions from street level images. However, aerial images have much lower resolution than streetside images, and can not provide discriminative intensity profile descriptors. Schindler et al. [10] use SIFT [11] to extract facade feature points and then introduce a variation of RANSAC-based planar grouping method to detect perspectively distorted lattices of these points. Park et al. [12] propose to substitute SIFT with Kanade Lucas Tomasi corners (KLT) [13], Maximally Stable Extremal Regions (MSER)[14] and Speeded Up Robust Features (SURF) [15] to get more feature points, then cluster and group these points to detect lattices. However, these methods mainly work with those large and highly repetitive building facades. In addition, the state-of-the-art lattice detection algorithm [16] used in [12] and [10] is computationally expensive for aerial images.

In this paper, we also use repetitive patterns to detect building facades, but require neither vertical and horizontal assumptions nor edge and line segment detection. Hence, this method is robust to the variation of building facade condition and image resolution. We substitute the general feature points extraction algorithms with our newly proposed approach, which combines MSERs and steered Determinant of Hessian. In addition, instead of adopting the lattices grouping method developed in [16] we use Delaunay Triangulation (DT) to guide local regularity searching. This grouping method can work with not only large and highly repetitive facades, but any facades with at least four regularly distributed windows. Furthermore, it speed up the search process greatly.

978-1-5090-4846-5/16/$31.00 ©2016 IEEE

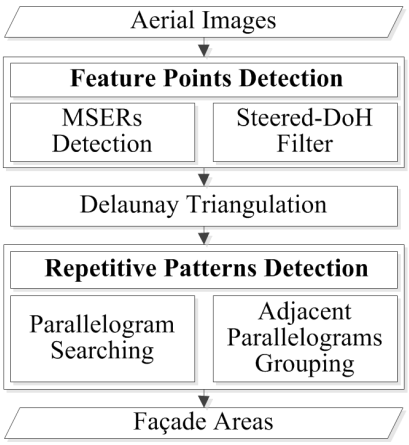


Fig. 1. Workflow of facades recognition

II. Facade Recognition Procedure

Low-level facade features (e.g. windows) are generally distributed in a regular pattern. The human visual system can easily recognize these kinds of patterns. Perceptual grouping refers to the ability to extract salient image patterns and structures from low-level image features without prior knowledge of high-level image content [12]. Our proposed procedure follows this concept, as shown in Fig. 1. Instead of detecting many types of features [16], only MSERs features are detected. Then, a steered Determinant of Hessian is used to find the most salient MSERs features in local areas. Different from iterative procedure of perceptual grouping [12], we introduce Delaunay Triangulation to guide parallelogram perceptual grouping thereby avoiding the heavy computing cost of iteration. We experimentally validate our method on aerial images with different resolutions (0.1-m and 0.3-m) and different facade characteristics.

A. Parameters Tuning of MSERs Detection

MSERs were originally proposed for wide-baseline matching. They are defined by an extremal property of the intensity function in the region and on its outer boundary [14]. Intuitively, they are regions that have either higher or lower intensity than their surroundings [17], and can be different shapes. We use MSERs to detect features since it is capable of detecting arbitrary-shape regions as a whole. This is useful for features perceptual grouping.

In the original formulation, the parameter $q(i) = |Q_{i+\Delta} \backslash Q_{i-\Delta}|/|Q_i|$, which represents the stability, is used to control the MSERs detection [14]. $Q_1, ..., Q_{i-1}, Q_i, Q_{i+1}, ...$ ($Q_i \subset Q_{i+1}$) are a sequence of nested extremal region candidates generated by thresholding image using different intensity level i. $q(i)$ represents the normalized area variation between $Q_{i+\Delta}$ and $Q_{i-\Delta}$ with respect to Q_i. Extremal region Q_{i*} is maximally stable iff $q(i*)$ is a local minimum of $q(i)$. Nevertheless, being maximally stable does not guarantee that an extremal region is an MSER. In general, its area ($|Q_{i*}|$) and variation($q(i*)$) must satisfy certain requirements.

In addition,several parameters of MSERs detector should be determined according to image characteristics. The change rate of i is set to 3 in this work. The area interval threshold $[MinArea, MaxArea]$ that we use for extract more facade features is approximately determined by the ratio $ra = S/R$ of feature size (S) and image resolution(R). In practice, we tend to use a wider range of area interval (i.e. [4,10000]) to detect more features. In addition, high variation ($q(i)$) threshold will produce many redundant features while low one often miss lots of critical features. Therefore, We set variation threshold to a experience value 0.25. Note that the outputs of MSERs detection are ellipses (described as center, axes length (major axis a and minor axis b) and orientation (θ, direction of major axis a with respect to x axis)) generated by fitting MSERs regions.

After detection, redundant nested MSERs are inevitable since their variations are at local minima. Instead of eliminating redundant features by searching global minimum variation, we employ another more efficient and robust filter constructed by steered Determinant of Hessian.

B. Steered Determinant of Hessian

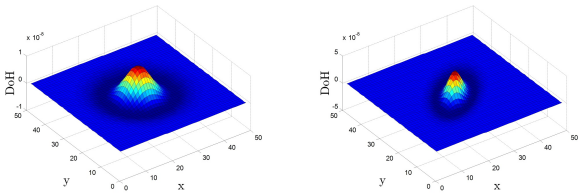
Hence, we propose to use the response of Determinant of Hessian (DoH) to eliminate redundant nested MSERs. The scale-normalized DoH Eq.(1) give strong responses on blobs and ridges [18].

$$\delta^2 H(\mathbf{x}, \delta) = \delta^2 \begin{bmatrix} L_{xx}(\mathbf{x}, \delta) & L_{xy}(\mathbf{x}, \delta) \\ L_{yx}(\mathbf{x}, \delta) & L_{yy}(\mathbf{x}, \delta) \end{bmatrix} \quad (1)$$

where L_{xx}, L_{yy} and L_{xy}, L_{yx} are partial derivatives of $L(\mathbf{x}, \delta) = g(\delta) \oplus I(\mathbf{x})$ where $\mathbf{x} = (x, y)$. $g(\delta)$ is an isotropic, circular Gaussian kernel of scale when $\delta_x = \delta_y = \delta$:

$$g(\delta_x, \delta_y) = \frac{1}{2\pi\delta_x\delta_y} e^{-(\frac{x^2}{2\delta_x^2} + \frac{y^2}{2\delta_y^2})} \quad (2)$$

The problem is when $\delta_x = \delta_y = \delta$ the elongated blobs cannot be detected effectively (Fig. 2 (a)). Therefore, affine Hessian detector was developed in [19]. It uses an iterative method to estimate the affine transformation and local pattern. To simplify this process we use the steered Determinant of Hessian Eq. (3) which is constructed by the second partial derivatives Eq. (5), Eq. (6) and mixed second partial derivatives Eq. (7) of a aisotropic, elliptic Gaussian kernel Eq.(2)



(a) isotropic filter $\delta_x = \delta_y = 40$ (b) aisotropic filter $\delta_x = 20, \delta_y = 40, \theta = \frac{\pi}{4}$

Fig. 2. Isotropic and Aisotropic Filter

($\delta_x \neq \delta_y$) to filter the redundant nested MSERs, as shown in Fig.2 (b).

$$det(steered\_H) \tag{3}$$

$$steered\_H = \begin{bmatrix} \delta_x^2 \frac{\partial^2 g}{\partial x^2} & \delta_x \delta_y \frac{\partial^2 g}{\partial x \partial y} \\ \delta_y \delta_x \frac{\partial^2 g}{\partial y \partial x} & \delta_y^2 \frac{\partial^2 g}{\partial y^2} \end{bmatrix} \tag{4}$$

$$\frac{\partial^2 g}{\partial x^2} = \frac{1}{2\pi\delta_x^3\delta_y}(\frac{u^2}{\delta_x^2} - 1)e^{-(\frac{u^2}{2\delta_x^2} + \frac{v^2}{2\delta_y^2})} \tag{5}$$

$$\frac{\partial^2 g}{\partial y^2} = \frac{1}{2\pi\delta_y^3\delta_x}(\frac{v^2}{\delta_y^2} - 1)e^{-(\frac{u^2}{2\delta_x^2} + \frac{v^2}{2\delta_y^2})} \tag{6}$$

$$\frac{\partial^2 g}{\partial x \partial y} = \frac{\partial^2 g}{\partial y \partial x} = \frac{uv}{2\pi\delta_x^3\delta_y^3}e^{-(\frac{u^2}{2\delta_x^2} + \frac{v^2}{2\delta_y^2})} \tag{7}$$

where

$$\delta_x = a, \delta_y = b,$$
$$u = xcos(\theta) - ysin(\theta),$$
$$v = xsin(\theta) + ycos(\theta).$$

C. Delaunay Triangulation Induced Perceptual Grouping

Due to the large distance between the camera and building facades in aerial images the local perspective deformation can be approximated by an affine transformation which preserves parallelity. Therefore, we propose to use a Delaunay Triangulation (DT) [20] induced parallelogram searching method to group MSERs.

Given a set of MSERs feature points, their spatial neighborhood relations are first represented by a Delaunay Triangulation. Two adjacent triangles $(\mathbf{T_1}, \mathbf{T_2})$ with one common side construct a quadrilateral as illustrated in Fig. 3. Then, similar to [12] three points $\{\mathbf{p_1}; \mathbf{p_2}; \mathbf{p_3}\}$ of a triangle are sampled to form a $(\mathbf{a}, \mathbf{b})$ vector pair given by $\mathbf{p_1} - \mathbf{p_2}$ and $\mathbf{p_1} - \mathbf{p_3}$. If the quadrilateral is a parallelogram, $\mathbf{p_4}$ should equal to $\mathbf{p'}$ which is calculated by $\mathbf{a} + \mathbf{b} + \mathbf{p_1}$. In practice, if the distance d between $\mathbf{p_4}$ and $\mathbf{p'}$ is less than certain threshold (i.e. 5 pixels), the quadrilateral will be taken as a parallelogram. The Delaunay Triangulation induced parallelogram searching and grouping algorithm is shown in Algorithm. 1.

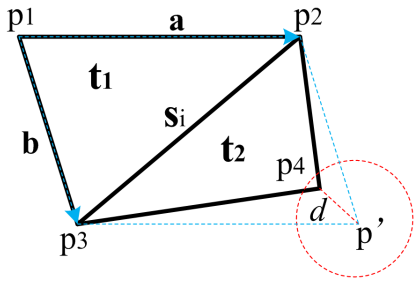


Fig. 3. Parallelogram Grouping

Parallelograms which have common side are grouped as facade areas. However, each group may not represent a whole facade accurately but a part of it.

Algorithm 1 Delaunay Triangulation Induced Feature Grouping

Input: Feature points set P defined by the centers of MSERs features

Output: Facades set F: each element in F is comprised of several triangles

1: Generate Delaunay Triangulation T from feature points P
2: **repeat**
3: Search neighbor triangles (t_m and t_n) which have common side s_i
4: Compute d_i between $\mathbf{p_4}$ and $\mathbf{p'}$ according to Fig. 3
5: If($d_i \leq 5$){mark both triangles as 1}
6: **until** (all sides s in T have been processed)
7: Cluster triangles which are marked as 1 according to their connectivity
8: Return clusters F

D. Noise removal

Although many of the MSERs feature points are randomly distributed in non-facade areas, there still exist a few non-facade points which coincidently satisfy the structure of parallelogram (False Positive). Fortunately, these false positives are usually isolated from each other. Therefore, it is possible to eliminate them by checking the number of parallelograms in each group. In this paper, we eliminate those groups which contain only one parallelogram (two adjacent triangles).

III. Experimental Results

To assess the performance of our method, experiments were conducted on a 8162×5986 pixel and a 7952×3161 aerial image with 0.1-m and 0.3-m resolution respectively. It is worth to note that the second image comes from a post-earthquake area and many buildings are tilted or partially damaged. Therefore vertical or horizontal assumptions of facades are not satisfied. The ground truth of facades in these two images have been manually labeled as red boxes, as shown in Fig. 4.

A. Evaluation and Comparison of Facade Recognition

The qualitative results of our facade recognition are given in Fig. 4. The straight line grouping (SLG) based algorithm [8] has been shown to effectively extract building facade areas without making vertical and horizontal assumptions. To quantify the facade feature extraction results, three frequently used metrics [21] were introduced. $Precision$ (Equation. 8) indicates the extent to which the detected facades correctly correspond to ground truth. $Recall$ (Equation. 9) is a measure of the omission error. $F1-score$ (Equation. 10) is a composite metric that take both correctness and completeness in consideration. The quantitative comparisons between our method and SLG are illustrated in Table I and Table II. The results suggest that the performance of our method and SLG are similar when the image resolution is high (0.1-m). However, our algorithm is much more robust when image resolution decreases. SLG fails when the image resolution is 0.3-m, while our method still achieves $F_1-score$ over 80%. The reason is that straight lines

(a) high resolution (0.1-m)

(b) low resolution (0.3-m)

Fig. 4. Facade recognition results

TABLE I. Comparison of facade recognition (0.1-m)

Method	No. of TP	No. of FP	No. of FN	Precision	Recall	F_1- Score
SLG	112	3	15	97.3%	88.2%	92.5%
Our	115	11	15	91.3%	88.4%%	89.8%

Notes: our method can detect some facades which can not be defined by straight line grouping so that our $N_{TP} + N_{FN}$ is greater than SLG.

TABLE II. Comparison of facade recognition (0.3-m)

Method	No. of TP	No. of FP	No. of FN	Precision	Recall	F_1- Score
SLG	-	-	-	-	-	-
Our	52	7	13	88.1%	80.0%	83.9%

of windows can not be extracted effectively in low resolution images as shown in Fig. 5.

$$precision = \frac{N_{TP}}{N_{TP} + N_{FP}} \tag{8}$$

$$recall = \frac{N_{TP}}{N_{TP} + N_{FN}} \tag{9}$$

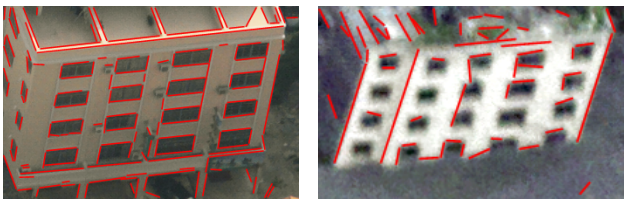
$$F_1 - Score = \frac{2N_{TP}}{2N_{TP} + N_{FP} + N_{FN}} \tag{10}$$

where N_{TP}, N_{FP} and N_{FN} are the number of true positives, false positives and false negatives.

B. Evaluation and Comparison of Feature Detection and Grouping

We validate the effectiveness of the combination of MSERs and steered-DoH proposed as a facade feature detector by showing the results of four typical facades. As shown in Fig. 6, the performance of our method is much better than SURF [15] in extracting facade features, especially for extracting long structures. The quantitative assessment of these four typical facades feature extraction is demonstrated in Table III. The results show that our feature detector extract accurate facade window blobs with less noise. These extracted features enable the success of subsequent perceptual grouping.

Furthermore, we also evauate efficiency and effectiveness in a quantitative comparison with deformed lattice detection [16] on the process of feature grouping on these typical facades. The algorithm of [16] is implemented in hybrid of C++/OpenCV and MATLAB, while ours is implemented in



(a) high resolution(0.1-m) (b) low resolution (0.3-m)

Fig. 5. Straight lines extraction [22]

TABLE III. Quantitative assessment of feature extraction

Method	No. of TP	No. of FP	No. of FN	Precision	Recall	F_1- Score
SURF	151	637	90	19.16%	62.66%	29.35%
Our	229	18	12	92.71%	95.02%	93.85%

TABLE IV. Total time(s) consumption per facade

Method	F1	F2	F3	F4
Lattice	128.9330	62.3260	79.0490	81.7410
Our	6.2001	4.8221	7.4086	6.9834

MATLAB. Both algorithms run on the same machine with a 2.53GHz i5 CPU, 8GB RAM and a Windows 7 64-bit OS. The deformed lattice detection approach has much lower recall than ours, especially on those facades whose windows have different size and shapes (Fig. 6). In addition, our algorithm has a speed advantage, Table IV. The Delaunay Triangulation based grouping finds local parallelograms and then groups them together. Hence parallelograms in one facade do not all have to be the same. That means facades which have different size of window blobs and intervals between blobs can also be grouped correctly. As shown in the fourth row of Fig.6, [16] just groups points with same intervals and fails when the window size changing. Our DT induced group has no problem with that case.

IV. Conclusion

This paper concerns the problem of large-scale facade recognition from aerial images. Our approach leverages properties of repetitive patterns of building facades, namely large number of evenly distributed window blobs. Repetitive windows are used as an indicator of the presence of a facade region. Through the combination of MSERs and steered-Difference-of-Hessian, we have obtained promising facade feature extraction results from images with different resolution. This leads to accurate feature grouping in the second stage. Delaunay Triangulation induced parallelogram searching brings a notable reduction of computation cost and better performance than traditional solutions. Finally, a set of facade regions is represented by successfully grouped triangles. We have compared experimentally and quantitatively with state-of-the-art algorithms. The results show that our approach have superior performance on our challenging datasets. The source code will be distributed later.

Acknowledgments

The authors would like to thank China Survey for providing data. This research is supported by China Scholarship Council.

References

[1] L. Zebedin, A. Klaus, B. Gruber, and K. Karner, "Façade reconstruction from aerial images by multi-view plane sweeping," *Photogrammetrie Fernerkundung Geoinformation*, vol. 2007, no. 1, p. 17, 2007.

[2] P. Meixner and F. Leberl, "Characterizing building facades from vertical aerial images," *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 38, pp. 98–103, 2010.

Fig. 6. Comparison in feature detection and grouping: the first row are the raw images; the second row are features detected by SURF [15]; the third row are feature detection results of our method; the fourth row are lattice grouping results by [16]; the fifth row are the parallelogram grouping results of our Delaunay Triangulation induced grouping method.

[3] P. Zhao, T. Fang, J. Xiao, H. Zhang, Q. Zhao, and L. Quan, "Rectilinear parsing of architecture in urban environment," in *2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2010, pp. 342–349.

[4] J. A. Delmerico, P. David, and J. J. Corso, "Building facade detection, segmentation, and parameter estimation for mobile robot stereo vision," *Image and Vision Computing*, vol. 31, no. 11, pp. 841–852, 2013.

[5] M. Bansal, H. S. Sawhney, H. Cheng, and K. Daniilidis, "Geo-localization of street views with aerial image databases," in *Proceedings of the 19th ACM international conference on Multimedia*. ACM, 2011, pp. 1125–1128.

[6] J. Liu and Y. Liu, "Local regularity-driven city-scale facade detection from aerial images," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 3778–3785.

[7] J. Xiao, M. Gerke, and G. Vosselman, "Building extraction from oblique airborne imagery based on robust façade detection," *ISPRS journal of photogrammetry and remote sensing*, vol. 68, pp. 56–68, 2012.

[8] X. Yang, X. Qin, J. Wang, J. Wang, X. Ye, and Q. Qin, "Building façade recognition using oblique aerial images," *Remote Sensing*, vol. 7, no. 8, pp. 10 562–10 588, 2015.

[9] A. Wendel, M. Donoser, and H. Bischof, "Unsupervised facade segmentation using repetitive patterns," in *Pattern Recognition*. Springer, 2010, pp. 51–60.

[10] G. Schindler, P. Krishnamurthy, R. Lublinerman, Y. Liu, and F. Dellaert, "Detecting and matching repeated patterns for automatic geo-tagging in urban environments," in *IEEE Conference on Computer Vision and Pattern Recognition, 2008. CVPR 2008.* IEEE, 2008, pp. 1–7.

[11] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International journal of computer vision*, vol. 60, no. 2, pp. 91–110, 2004.

[12] M. Park, K. Brocklehurst, R. T. Collins, and Y. Liu, "Translation-symmetry-based perceptual grouping with applications to urban scenes," in *Computer Vision–ACCV 2010*. Springer, 2010, pp. 329–342.

[13] J. Shi and C. Tomasi, "Good features to track," in *Proceedings CVPR'94 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1994.* IEEE, 1994, pp. 593–600.

[14] J. Matas, O. Chum, M. Urban, and T. Pajdla, "Robust wide-baseline stereo from maximally stable extremal regions," *Image and vision computing*, vol. 22, no. 10, pp. 761–767, 2004.

[15] H. Bay, T. Tuytelaars, and L. Van Gool, "Surf: Speeded up robust features," in *Computer vision–ECCV 2006*. Springer, 2006, pp. 404–417.

[16] M. Park, K. Brocklehurst, R. T. Collins, and Y. Liu, "Deformed lattice detection in real-world images using mean-shift belief propagation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 10, pp. 1804–1816, 2009.

[17] P.-E. Forssén and D. G. Lowe, "Shape descriptors for maximally stable extremal regions," in *IEEE 11th International Conference on Computer Vision, 2007. ICCV 2007.* IEEE, 2007, pp. 1–8.

[18] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. Van Gool, "A comparison of affine region detectors," *International journal of computer vision*, vol. 65, no. 1-2, pp. 43–72, 2005.

[19] K. Mikolajczyk and C. Schmid, "Scale & affine invariant interest point detectors," *International journal of computer vision*, vol. 60, no. 1, pp. 63–86, 2004.

[20] F. P. Preparata and M. Shamos, *Computational geometry: an introduction*. Springer Science & Business Media, 2012.

[21] T. Fawcett, "An introduction to roc analysis," *Pattern recognition letters*, vol. 27, no. 8, pp. 861–874, 2006.

[22] C. Akinlar and C. Topal, "Edlines: Real-time line segment detection by edge drawing (ed)," in *2011 18th IEEE International Conference on Image Processing (ICIP)*. IEEE, 2011, pp. 2837–2840.