A GRAMMATICAL MODEL FOR ROOFTOP EXTRACTION FROM AERIAL IMAGE

Qiongchen Wang, Zhiguo Jiang

Image Center, BUAA, BEIJING, CHINA, 10076

1. INTRODUCTION

Automatic rooftop detection and localization in aerial images has been an active research topic in past decades in GIS, town planing, etc and recently experienced renewed interest spurred by new applications including object detection and tracking from UAV as well as 3D building reconstruction such as 'CyberCity" and GoogleEarth. In this paper, we focus on two types of rooftops, rectilinear and gable roofs (see Fig. 1), from high resolution chromatic aerial images with near vertical view. Detection of these roofs without utilizing auxiliary information is still a very challenging task for two reasons: 1) the presence of heavy background clutter and the variety of details present on or near the building makes the computation of image primitives unreliable; 2) rooftops have large intra-class variances in both shape and appearance, which makes it impossible to build a simple structural model to account for all appearances. Fig.1 visualizes these challenges. The existing approaches can be separated into three main categories. The first category is contains the *hypothesis and testing* method [1] in which a hierarchical aggregation of 2-D or 3-D image features form the building hypotheses, followed by a hypothesis verification step. The second category is perceptual organization-based building detection, which pose the building detection task as an optimization problem on graph partitioning, integrating geometric and photometric heuristics; The final category is a combination of the above two[2] methods or a combination of one of the two methods with auxiliary information, such as using shadows and walls as rooftop context or a digital elevation model(DEM)as a pre-segmentation of the building.

Fig. 1. Examples of "rectilinear" and "gable" rooftops from aerial images. They have big difference in both shape and appearance, but share three types of quadrangle parts (rectangle, right trapezoid, isosceles trapezoid), as highlighted with colors

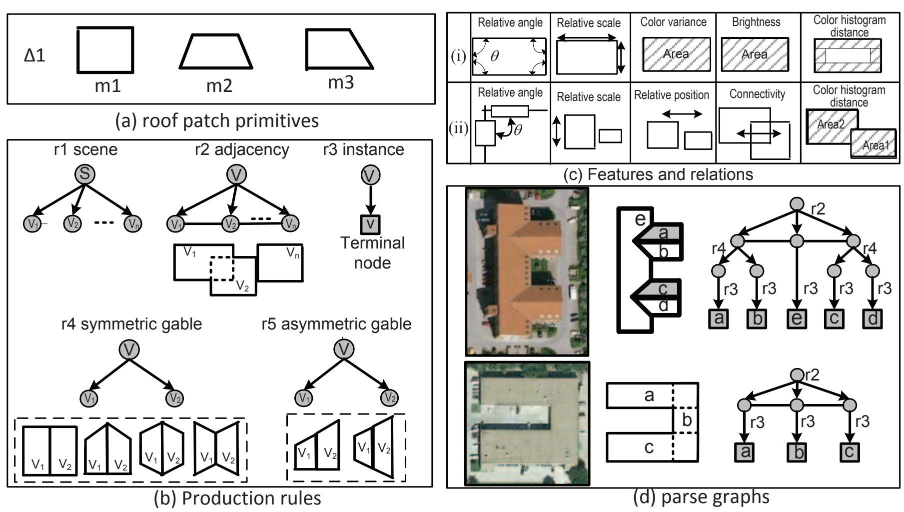

Fig. 2. (a) Roof patch primitives, (b) 5 production rules, (c) Features and relations, (d) two examples of parse graph

2. BRIEF METHOD

In this paper, we present a grammatical model together with an effective bottom-up/top-down inference algorithm for extracting rooftops from a monocular aerial image. We form roofs from a graph grammar [3] consisting of 5 production rules and 3 types

Thanks to XYZ agency for funding.

of commonly shared quadrilateral primitives, as shown in Figure2. Each production rule not only expands an object into its child primitives but also includes a number of equations that constrain the attributes of a parent node and those of its children. We further define a set of horizontal links between nodes at the same level, which account for spatial/appearance constraint between peer nodes. The grammar is used to produce a large number of roof configurations and can be used to represent the wide structural variability of rooftops. In the inference algorithm, the bottom-up step detects a huge number of quadrilateral primitives by hierarchically composing smaller primitives (e.g. L junction, parallel lines) in a variety of ways. The quadrilateral primitives are weighted and sorted according to their probability. They then activate a top-down step to propose neighboring roofs and prune incompatible detections to arrive at the most likely results.

3. EXPERIMENTAL RESULTS

We collect a dataset of 50 free high resolution(2m .5m/pixel) aerial images(including 485 rooftops for total) from GoogleEarth, some of which have very large view (2km*2km) and are stitched together manually from smaller view images. Our data located around (33 54’27.55N,118 22’49.18W) and (52 33’28N 13 33’02E). We select 20 images as a training dataset and use the remaining 30 images as the testing dataset. The rooftops in both the training and testing dataset are labeled with polygons. Since our model is based on line segments, we take the segment as positive sample if it is in the neighborhood of the labeled polygon sides, otherwise, it will be a negative sample. For roof patch primitives and merged rooftop results the area of overlap between the detected results and the ground truth is required to exceed 85%.

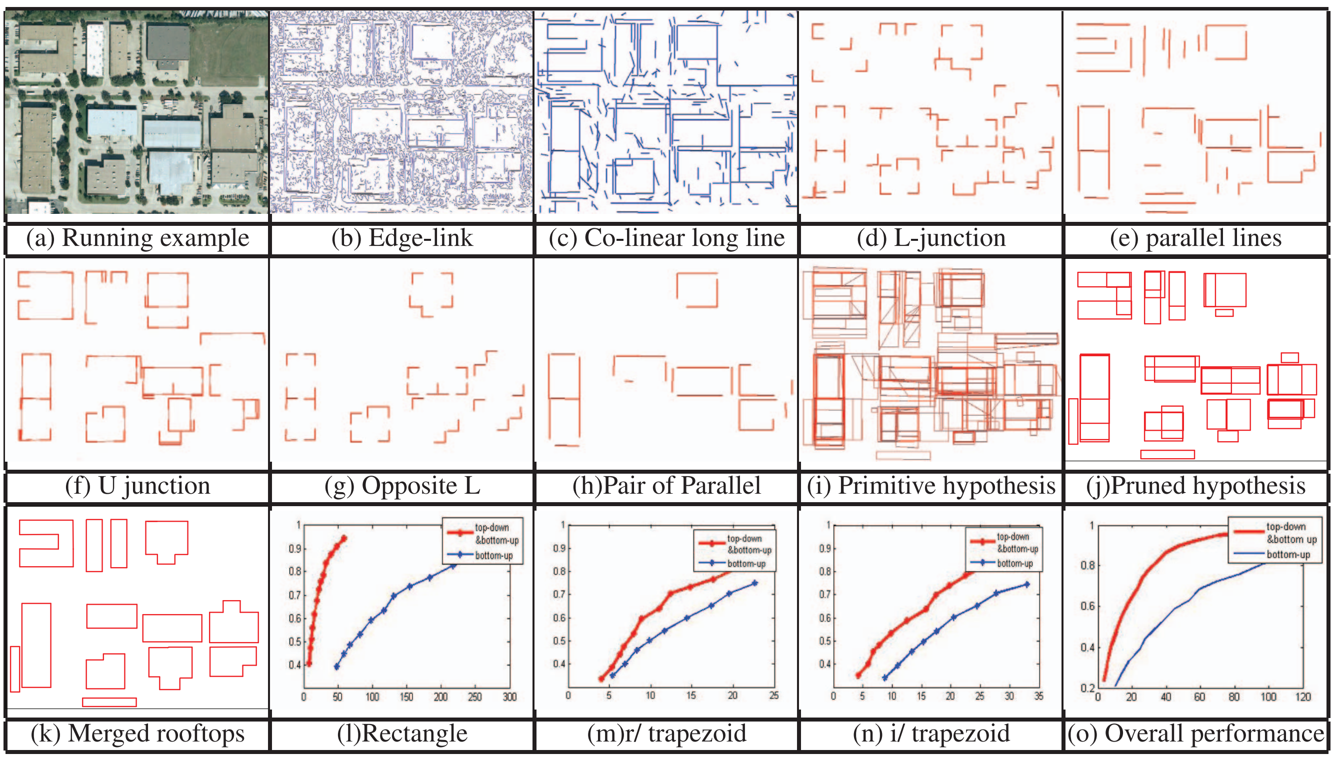

Fig. 3. (a)-(k) A running example, (c)-(h) are the bottom-up proposals for the intermediate parts and (k) is the final detected rectilinear shapes.(l)-(n) are detection performance of three primitives and (o) is detection performance of merged rooftops

4. REFERENCES

[1] C. Lin and R. Nevatia, “Building detection and description from a single intensity image,” *CVIU*, vol. 72, 1998.

[2] A. Katartzis and H. Sahli, “A stochastic framework for the identification of building rooftops using a single remote sensing image,” *IEEE Trans. GRS*, vol. 46, 2008.

[3] F. Han and S. C. Zhu, “Bottom-up/top-down image parsing by attribute graph grammar,” in *ICCV*, 2005.