BUILDING ROOF DETECTION FROM A SINGLE HIGH-RESOLUTION SATELLITE IMAGE IN DENSE URBAN AREAS

Zongying Song, Chunhong Pan, Q Yang, Fuxin Li, Wei Li *

The Institute of Automation, Chinese academy of science
No.95,Zhongguancundonglu, Beijing, China
zysong@nlpr.ia.ac.cn

Commission III/3

KEY WORDS: building extraction, high resolution,urban scene,hypothesis-verification, edge verification,Probabilistic model

ABSTRACT:

This work aims at extracting 2D buildings from a single high-resolution satellite image in densely built-up urban areas. The whole algorithm follows the hypothesis-verification strategy. The key contribution of this paper is the edge verification method of the hypothesis verification process, which can greatly improve the accuracy and precision of the building extraction results from complex scenes. To extract more accurate and precise edges, we construct a probabilistic model and an optimization frame. First, a probability being the optimal edge is given to each possible edge, then the constraints of any two possible edges are estimated based on a machine learning method and the image evidences, finally, these constraints and other prior knowledge are integrated into an optimization problem, by solving which these probabilities can be computed and the optimal edge can be selected. At last, we provide some experimental results on large and complex scenes that demonstrate the robustness and accuracy of our algorithm.

Over the past decades, automatic detection and reconstruction of buildings from aerial or satellite images has been an active research topic in both computer vision and digital photogrammetry (M.Fradkin et al., 2001, Suveg and Vosselman, 2004). This is due to an increasing need for building detection and reconstruction in a variety of applications. Especially in cartography, a powerful automatic system can greatly reduce the effort needed to assemble a 2D or 3D digital map. Other applications of this technology include urban planning, GIS update, environmental monitoring and virtual & augmented reality.

Many building extraction approaches have been developed, and most of them are based on feature extraction and feature grouping. In the early stage, building extraction methods attempt to work solely with monocular images (Irving.R and Mckeown.D, 1989, Lin.C and Nevatia.R, 1998, McGlone.J and Shufelt.J.A, n.d.), while recent works have focused on multi-views (Baillard.C and Zisserman.A, n.d., Collins et al., 1998, Fischer et al., 1998, Roux.M and Mckeown.D.M, n.d., S.Noronha and R.Nevatia, 2001, Haala and Anders, 1996) or new data sources (D.Koc and M.Turker, 617–622, Haala et al., 1998, Kim and Nevatia, 2004, Lee.S et al., n.d., San and M.Turker, n.d., S.Noronha and R.Nevatia, 2001, Song et al., n.d.), by which prior knowledge or 3-D information can be obtained.

High resolution satellite images are increasingly available now. That allows extracting more accurate and precise buildings from urban areas. Hence, in this paper, based on this new data source, we propose a building extraction method focusing on 2D rectangular building extraction in densely built-up urban areas of large scenes, by which the accuracy and precision of the results are greatly improved.

The quality of building extraction are closely related to the complexity of the scenes been processed. For high resolution images of densely built-up urban areas, automatic building extraction is very difficult because of some intrinsic challenges. Firstly, there are hundreds of buildings with diverse appearances in the same scenes. They may appear very differently in intensity levels, shapes and structures. Secondly, noises, arising from many kinds of factors such as environments, other man-made objects etc., makes the surroundings and the inside of buildings very complex. Thus, it is very hard to correctly segment buildings from the complex scenes. Thirdly, accompanied with the increase of the image resolution, most buildings' sub-structures are visible. Distinguish buildings with their sub-structures are difficult and intrinsically ambiguous.

To overcome the above difficulties, many approaches have been proposed (Baltsavias et al., 2001, Mayer, 1999, M.Fradkin et al., 2001, N.Paparoditis, 1998, Suveg and Vosselman, 2004). However, until now this is still a partially solved problem, an effective and robust method is still under development(Baltsavias et al., 2001, Mayer, 1999, N.Paparoditis, 1998, San and M.Turker, n.d., Song et al., n.d., Suveg and Vosselman, 2004). In this work, we take use of a global focusing step and the hypothesis-verification strategy. More importantly is that we propose a novel edge verification method, which can effectively discriminates edges from all kinds of noises in complex scenes. In our algorithm, texture feature analysis and region grouping methods are first applied to estimate the locations and cover areas of all buildings, then based on these estimations, lines are selected and grouped to generate building hypotheses, finally, our new hypotheses verification method based on edge verification is applied to obtain the optimal hypothesis.

The rest of this paper are arranged as follows: the related work and an overview of our method would be presented in section 2; Section 3 describes the building hypotheses generation; The hypotheses verification method will be shown in Section 4 ; Finally, some experimental results and the conclusion are given.

1 OVERVIEW

(Song et al., n.d.) describes a typical building extraction approach based on high resolution images. In this method, images are first

*The work was supported by National Natural Science Foundation of China, NSFC/No:60675012.

over-segmented into atomic-regions, and some of them with good performance are selected as seeds. Then, the region grouping method is used to grouping the atomic-regions together if they satisfy some texture and geometry constraints. Thus, an atomic-region group, named as Candidate Building Regions(CBR), is generated for each seed. Next, for each CBR, a rectangular building hypothesis is generated based on the improvement of the minimum enclosed rectangles of each CBR. Finally, some image evidences are collected to delete the false hypotheses.

From the above descriptions, it can be seen that the prior focusing step of paper (Song et al., n.d.) is very simple. By the texture analysis and region grouping methods, it can obtain the optimal estimations of the locations and cover areas of all buildings. Although the hypothesis generation and verification methods of paper (Song et al., n.d.) is very simple, its detection rate is still very good (88.8%). Therefore, in our algorithm, the prior focusing step of (Song et al., n.d.) is still retained, by which we can provide great help for the further hypothesis generation and verification method.

In our algorithm, building hypotheses are generated by the perceptual grouping of lines like many existing methods (Lin.C and Nevatia.R, 1998). The difference is that not all lines are used to generate building hypotheses, only lines selected by CBRs are used. In this way, the combination expansion problem can be avoided. This is very important for the processing of large scenes with hundred of buildings.

The hypothesis verification method is the critical contribution of this paper, in which the traditional whole hypothesis verification is transformed into detailed edge verification, which can be divided into 3 steps: first, based on each generated hypotheses set, four possible edge sets are constructed: left-edge set, right-edge set, top-edge set and bottom-edge set; Then, image evidences are collected to estimate the constraints of any two possible edges; Finally the probability being the optimal edge is computed for each possible edge by the above collected constraints, thus the one with the highest probability can be treated as the optimal edge. Finally, four optimal edges can be generated and integrated to the optimal hypothesis.

Compared with most existing hypothesis verification methods, there are mainly 3 contributions of our method: **1),** The whole hypotheses verification is decomposed into the detailed edge verification. This is because with the complexity increase of the surroundings and the inside of building roofs, the overall performance of hypotheses may be very similar. Hence, it is infeasible to distinguish hypotheses based on the overall performance. **2),** In the edge verification method, the judgement of the relationship of any two possible edges is transformed into a two-class classification problem. In this way, the machine learning methods can be introduced into our frame to help us to make more robust decision. **3),** In the edge verification method, to robust compute the optimal edge for all buildings, a probabilistic model and a Quadratic Programming frame are constructed. All edges are given a probability being the optimal edge, and constraints for any two possible edges and other prior knowledge are integrated into a Quadratic Programming model. Thus, all kinds of information provided by images can be effectively used in a robust decision making frame.

2 BUILDING HYPOTHESES GENERATION

Figure 1 shows the rectangular building roof model used for building hypothesis generation, which is described by $\{\alpha, t, l, b, r\}$. α is the direction of building roof, which has the range of $[0^\circ, 90^\circ)$; (t, l, b, r) are the coordinate values of the four edges, measured in the building coordinate system (axes in red color). The four edges of this model are defined as: **top-edge:** the long edge adjacent to shadow; **bottom-edge:** the long edge opposite top-edge; **left-edge:** the short edge adjacent to shadow; **right-edge:** the short edge opposite left-edge. Based on these definitions, all building hypotheses with different directions can be normalized into a unified frame, even when the scenes have different sunlight directions.

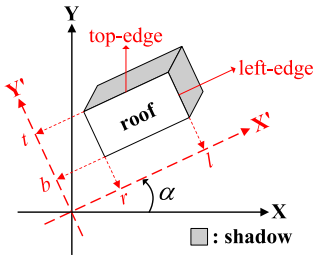

Figure 1: The model of rectangular building roof. α: building roof direction; t: coordinate value of top-edge; l: coordinate value of left-edge; b: coordinate value of bottom-edge; r: coordinate value of right-edge.

Before hypotheses generation, two preparing works need to be finished. First, the prior focusing step of (Song et al., n.d.) should be used to obtain the best estimation of the buildings' location and covering areas. After image over-segmentation and atomic-region grouping, a number of atomic-region groups are generated. These atomic-region groups, also called Candidate Building Regions (CBRs), provide an excellent estimation of the locations and cover areas for all buildings. Secondly, for each CBR, the lines used for the hypothesis generation need to be selected. These lines will be selected if they are inside a CBR or within a certain distance to the CBR's contour. Thus, building hypotheses can be generated by the perceptual grouping of these selected lines.

Because hypotheses generation and verification for all CBRs are in the same manner, in the following we will only consider the processing of one CBR as an example. Let us represent this CBR as V, and the line set as L. For V and L, the building hypotheses are generated by the follow 4 steps (Fig. 2):

Step 1. Line set L is divided into 9 subsets according to the direction of lines: $L_k = \{l | l \in L, \alpha_l \in [(k-1)\cdot 10^\circ, k \cdot 10^\circ)\}, k = 1 \cdots 9$, where α_l is the direction of line l. Perpendicular lines are put into the same subset. For example, lines with direction of 5° and 95° are in the same subset L_1.

Step 2. For each subset L_k, compute a direction α_k, which represents the direction of the whole subset. α_k is determined by the direction of the longest line or the line with the highest intensity in L_k.

Step 3. For each L_k, any four lines which approximately form a rectangle would generate a building hypothesis. Based on the four intersection points of these four lines and the subset direction α_k, a rectangle building hypothesis is generated and recorded by the above five parameters.

Step 4. A simple selection mechanism is performed to eliminate obviously unreasonable hypotheses. For example, too small or too large in size or covering too many shadow regions. This finalizes the generated hypothesis set.

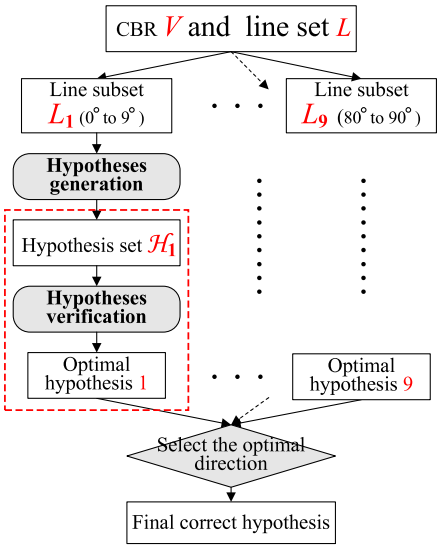

Figure 2: The flowchart of hypotheses generation and verification process of CBR V and line set L. The red frame shows the hypotheses verification process of hypotheses set $\mathcal{H}_1$, which will be detailed explained in Fig 3.

Putting it altogether, 9 hypothesis sets corresponding to each L_k (each one has about 30 hypotheses in our experiments) will be generated for V. Actually, only two or three of them would be nonempty, since most L_ks are already empty.

3 HYPOTHESES VERIFICATION

Fig. 2 describes the whole flow of the hypotheses generation and verification process. After hypotheses generation, 9 hypothesis set are generated for CBR V. Then, the optimal hypothesis for each hypotheses set will be computed by the hypothesis verification method. Finally, one of these 9 optimal hypotheses will be selected as the final correct hypothesis if it has the correct direction. It can be seen that the whole verification process is accomplished through two steps, the first step looks for the best hypothesis of each hypothesis set, which as the key part of this paper will be explained in detail below, and the second step looks for the best direction, which can be easily solved and ignored here.

To compute the optimal hypothesis from one hypotheses set, our new hypothesis verification method decomposes hypothesis verification into the verification of the four edges, as shown in Fig. 3. First, based on each hypotheses set, hypothesis pairs are generated and divided into four subsets. Then the edge-verification method is applied on these subsets to compute the optimal edge for top-edge, bottom-edge, right-edge and left-edge. Finally, these four optimal edges are integrated together to generate the optimal hypothesis for each hypotheses set.

The following of the section is arranged as this: the hypothesis pair generation and classification will be described in subsection 3.1; the critical edge verification method will be presented in subsection 3.2; in subsection 3.3 we will describe the method to obtain the final correct hypothesis; the details of the image evidences used in edge verification will be explained in subsection 3.4.

3.1 Hypothesis pair Generation and Classification

As shown in Fig 2 and Fig 3, here we take the hypothesis set $\mathcal{H}_1$ as an example to show how to generate the four hypothesis

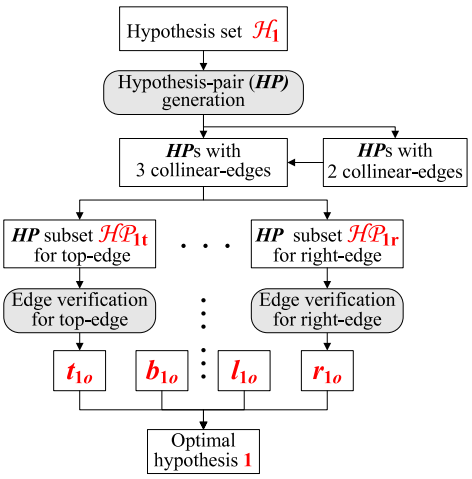

Figure 3: The flow of hypotheses verification for hypothesis set $\mathcal{H}_1$. t_{1o}, b_{1o}, l_{1o} and r_{1o} denote the coordinate values for the optimal top-edge, bottom-edge, left-edge and right-edge. ***HP*** denotes hypothesis pair.

pair subsets. First, any two hypotheses in $\mathcal{H}_1$ form a hypothesis pair, then hypothesis pairs with 2 or 3 collinear edges are selected (Fig. 4 shows some examples about collinear edges). Next, the hypothesis pairs with 2 collinear edges are transformed into two groups of hypothesis pairs with 3 collinear edges (Fig. 4(B)). In this way, a new set which only contains the hypothesis pair with 3 collinear edges is generated. Finally, this new set of hypothesis pairs is divided into four subsets according to the position of the non-collinear edge, which are represented as $\mathcal{HP}_{1r}$, $\mathcal{HP}_{1l}$, $\mathcal{HP}_{1t}$ and $\mathcal{HP}_{1b}$, with r, l, t, b (denoting right, left, top and bottom, respectively) specifying the edge that is not collinear. Fig 4.(A) shows an example belonging to $\mathcal{HP}_{1r}$.

3.2 Edge Verification

Based on the four hypothesis pair subsets generated above, the optimal edges for top-edge, bottom-edge, left-edge and right-edge will be computed by our edge verification method. Here, we still take the hypothesis set $\mathcal{H}_1$ as an example. Since the computation of the four optimal edges are in the same manner, we will only illustrate the computation of the optimal right-edge from $\mathcal{HP}_{1r}$.

Let r_{1o} denote the coordinate values of the optimal right-edge(Fig. 3). Here, we construct a probability model to compute it. First, a possible right-edge coordinate value set $R_1 = \{r_1, r_2, \ldots, r_q\}$ is collected from the hypothesis subset $\mathcal{H}_{1r}$, where q is the number of the distinct coordinate values of right-edges in $\mathcal{HP}_{1r}$ and r_i $(i = 1 \ldots, q)$ are the coordinate values. Then, for each $r_i \in R_1$, we use $p(r_i)$ to denote the probability of r_i being the coordinate value of the optimal right-edge. Thus, to compute the optimal edge, we only need to compute each $p(r_i)$ and take the edge with the largest probability.

As a simple scenario, suppose that there are only three elements r_1, r_2 and r_3 in R_1. For r_1 and r_2, they have the probability being the optimal right-edge as $p(r_1) + p(r_2)$, under this condition the conditional probability that r_1 is optimal becomes $\frac{p(r_1)}{p(r_1)+p(r_2)}$. Suppose $c_{12} = \frac{p(r_1)}{p(r_1)+p(r_2)}$, then we can get a linear constraint $(1 - c_{12})p(r_1) = c_{12}p(r_2)$ about $p(r_1)$ and $p(r_2)$. Similarly, there are linear constraints c_{13} for $p(r_1)$ and $p(r_3)$, c_{23} for $p(r_2)$ and $p(r_3)$. If these three linear constraints c_{12}, c_{13} and c_{23} can all be obtained, then based on $p(r_1) + p(r_2) + p(r_3) = 1$, $p(r_1)$ $p(r_2)$ and $p(r_3)$ can be easily computed.

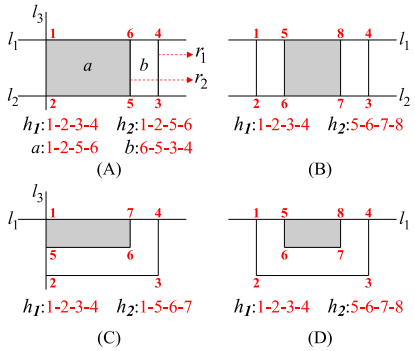

Figure 4: h_1 and h_2 denote two hypotheses forming a hypothesis pair. (**A**): Hypothesis pair with 3 collinear edges; (3-4) is the right-edge of h_1; (5-6) is the right-edge of h_2; r_1 and r_2 denote the coordinate values of the right-edges of h_1 and h_2. Region ***a*** and ***b*** are the share-region and different-region of h_1 and h_2.(**B**): Hypothesis pair with 2 collinear edges. The new generated two groups of hypothesis pairs with 3 collinear edges are:(1-2-7-8,5-6-7-8) and (5-6-3-4,5-6-7-8). (**C**): Another kind of hypothesis pair with 2 collinear edges. (**D**): Hypothesis pair with 1 collinear edge.

The above example is the primitive idea about how to compute the probabilities for all possible edges. The key is that we compute these probabilities by the relationships of any possible edge pairs. Note that it may be very difficult to directly compute these probabilities from image evidences, but the estimation of conditional probabilities from only two possible edges could be much easier. These probabilities can then be chosen to be consistent with the linear constraints generated by the estimated conditionals. In the following paragraphs, we will first estimate the linear constraints c_{ij} from image evidences, and then construct an optimization framework to compute every $p(r_i)$.

3.2.1 The Estimation of Linear Constraints c_{ij}

1. Estimating c_{ij} $c_{ij} = \frac{p(r_i)}{p(r_i)+p(r_j)}$ is a linear constraints $(1-c_{ij})p(r_i) = c_{ij}p(r_j)$ about $p(r_i)$ and $p(r_j)$, to estimate c_{ij}, we need to estimate the relationship of $p(r_i)$ and $p(r_j)$ based on image evidences. Here, a model is proposed to estimate this relationship:

$$c_{ij} = \frac{\text{\# of times } r_i \text{ is better than } r_j}{\text{\# of } \textbf{\textit{HP}}\text{s with right-edges' coordinate value as } r_i \text{ and } r_j}. \quad (1)$$

, where # denotes number, and ***HP*** denotes hypothesis pair. This model is a statistic about the relationship of r_i and r_j under the condition that r_i and r_j appear in the same hypothesis pair, and it can be used as an estimation of c_{ij}.

To compute c_{ij} by equation 1, the hypothesis pairs of $\mathcal{HP}_{1r}$ whose right-edge coordinate values are r_i and r_j are collected. Then the better hypothesis of these collected hypothesis pairs will be selected.

2.Identify the better hypothesis from a hypothesis pair It is very difficult to give a robust identification based on image evidences due to the complexity and diversity of scenes. To effectively deal with this difficulty, the machine learning approach is introduced into our system. We first transform the problem of identifying better hypothesis into a classification problem, and then apply the boosting method to solve this classification problem.

The identification of the better hypothesis is transformed into a two-class classification problem: identify the share-region and different-region of a hypothesis pair is consistent or not (see Fig. 4.(A)). The consistency here means the degree of two regions belong to the same buildings, which not only refers to the similarity of image features but also involves the context constraints of two regions from one same building. Figure 4.(A) shows an example. h_1 and h_2 are two hypothesis forming a hypothesis pair, whose right-edges are not collinear. r_1 and r_2 are the coordinate values of their right-edges. Region ***a*** and ***b*** are the share-region and different-region of h_1 and h_2. If the image evidences of region ***a*** and ***b*** are consistent, these two regions should be merged, which means that h_1 is better than h_2 and r_1 is better than r_2. Otherwise, r_2 is regarded better than r_1. This is the basic idea about how to identify the better hypothesis from a hypothesis pair.

The boosting method is introduced into our system to give a robust identification on the consistency of the share-region and different-region of a hypothesis pair. Boosting is a class of methods of linearly combining weak learners to obtain better classification or regression performances, which can effectively deal with complex classification problems. In this work, we choose the LogitBoost (Friedman.J et al., n.d.) algorithm with 4-level decision trees as the weak learner. Classifiers trained by this algorithm will have complex and accurate decision boundaries. A vector consisting image evidences of hypothesis pairs are used as features in the algorithm. The details of the feature vector will be presented in subsection 3.4. The classifier trained on right-edge will then select the better hypothesis of each hypothesis pair in $\mathcal{HP}_{1r}$.

Since there are only a few images available, the leave-one-out cross-validation protocol is used in the classifier training process. For each run, we single out one image as the test image and use all the other images as training images.

3.2.2 The Computation of $p(r_i)$

After the linear constraints c_{ij} for any $p(r_i)$ and $p(r_j)$ are obtained, a method is needed to compute $p(r_i)$. Here, an optimization problem is constructed, in which all $p(r_i)$ are optimized to be maximally consistent with these linear constraints and other knowledge. Firstly, we minimize the squared residual $((1-c_{ij})p(r_i) - c_{ij}p(r_j))^2$ for every i and j. In this way, p_{r_i} will be consistent with the linear constraints coming from the image evidences. Additionally, $p(r_i)$ should satisfy some other constraints, for example, they need to be nearly proportional with the frequency of coordinate value of r_i appeared in hypothesis set $\mathcal{HP}_1$, which is recorded as $d(r_i)$.

Based on the two optimization objectives above, the optimization of $p(r_i)$ is transformed into quadratic programming:

$$\begin{aligned} \min_{p(r_i)} \quad & \textstyle\sum_{i,j=1}^{q}((1-c_{ij})p(r_i) - c_{ij}p(r_j))^2 \\ & +\lambda \textstyle\sum_{i=1}^{q}(p(r_i) - d(r_i))^2 \\ s.t. \quad & p(r_i) \geq 0, i = 1, ..., q \text{ and } \textstyle\sum_{i=1}^{q} p(r_i) = 1. \end{aligned} \quad (2)$$

in which the constraints like $d(r_i)$ are used as a normalization constraint in the optimization problem, and λ is a weight coefficient. This is a convex quadratic program for which global optimum is guaranteed. Since q is not very large, the optimization problem can be solved easily with a standard solver (LOQO[1]). After all $p(r_i)$ are computed, the coordinate value with the highest probability will be selected as the optimal right-edge's coordinate value (r_{1o}).

3.3 Obtain the Final Correct Hypothesis

Based on subsection 3.2.2, all probabilities $p(r_i)$ are computed for each $r_i \in R_1$, we just need to select the one with the high-

[1] http://www.princeton.edu/ rvdb/loqo/LOQO.html.

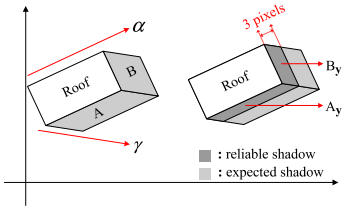

Figure 5: Shadow evidence. Region $\mathbf{A}$ and $\mathbf{B}$ are the expected shadow regions; region $\mathbf{A_y}$ and $\mathbf{B_y}$ are the comparatively reliable shadow regions; α is the building direction; γ is the direction of shadow lines casted by vertical lines.

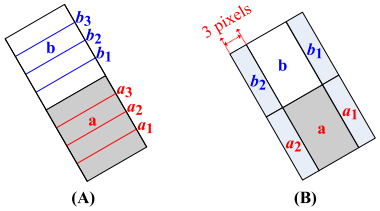

Figure 6: (A): The sample lines for $\boldsymbol{a}$ and $\boldsymbol{b}$. (B) The support regions for $\boldsymbol{a}$ and $\boldsymbol{b}$.

est probability as the optimal right-edge's coordinate value r_{1o}. Following the same manner of subsection 3.1 and 3.2, the coordinate value of the optimal top-edge, bottom-edge and left-edge (t_{1o}, b_{1o}, l_{1o}) can be also computed from the hypothesis pair subset $\mathcal{HP}_{1t}$, $\mathcal{HP}_{1b}$ and $\mathcal{HP}_{1r}$. Thus, the optimal hypothesis 1 of hypothesis set $\mathcal{H}_1$ can be generated (Fig. 3). Similarly, the optimal hypothesis for other hypothesis sets can be also computed, such as the optimal hypothesis 9 for hypothesis set $\mathcal{H}_9$. Finally, for CBR V, 9 optimal hypotheses are computed. Obviously, the one optimal hypothesis with the best direction will be chose as the final correct hypothesis(Fig. 2).

3.4 Image Evidences

Three types of image evidences are extracted to identify the better hypothesis from different aspects, which are shadows, interior similarity and supporting regions. Based on them, a 10-D image evidences vector is computed for each hypothesis pair.

1. Shadows: Shadows are regarded as a robust image feature for building extraction (Lin.C and Nevatia.R, 1998). However, it is very difficult to extract reliable and accurate shadow evidences from complex scenes, since in most cases shadows are casted neither on a flat surface nor on a surface with even intensity distribution. Hence, in this paper, we only extract some comparatively reliable shadow regions as the shadow evidence. Fig 5 shows the 2D shadow model of a building roof, region $\mathbf{A}$ and $\mathbf{B}$ are the expected shadow regions. Regions $\mathbf{A}_y$ and $\mathbf{B}_y$ are the comparative reliable shadow regions used here, which can be computed by the building direction and sunlight direction. According to $\mathbf{A}_y$ and $\mathbf{B}_y$, two shadow percentages $Sa(\mathbf{A}_y)$ and $Sa(\mathbf{B}_y)$are computed : $Sa(\mathbf{A}_y) = N_s(\mathbf{A}_y)/N(\mathbf{A}_y)$, $Sa(\mathbf{B}_y) = N_s(\mathbf{B}_y)/N(\mathbf{B}_y)$, where $N_s(\mathbf{A}_y)$ and $N_s(\mathbf{B}_y)$ are the number of shadow pixels of region $\mathbf{A}_y$ and $\mathbf{B}_y$, and $N(\mathbf{A}_y)$ and $N(\mathbf{B}_y)$ are the number of pixels of region $\mathbf{A}_y$ and $\mathbf{B}_y$.

2. Interior Similarity: Intensity distribution is often used to evaluate the interior similarity of two regions. However, this evaluation has a severe limitation. It can not effectively describe the structural similarity of two regions. But for buildings, the structure information is very important, sometimes two different parts of one building may have different intensity distributions, while

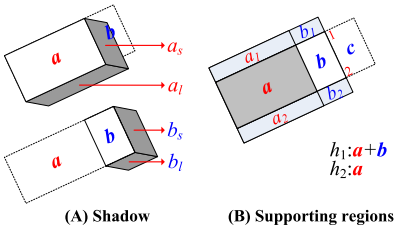

Figure 7: Evidences. h_1 and h_2 are two hypothesis forming a hypothesis pair. $\boldsymbol{a}$ and $\boldsymbol{b}$ are the share-region and different-region of h_1 and h_2. Region $\boldsymbol{c}$ is symmetric to $\boldsymbol{b}$ about edge (1-2); (a_l, a_s, b_l, b_s) are the reliable shadow regions; (a_1, b_1, a_2, b_2) are the supporting regions.

their structures may be similar. Hence, besides intensity distribution, we propose a new measurement method to evaluate the similarity of structures. Figure 6(A) shows the details: $\boldsymbol{a}$ and $\boldsymbol{b}$ are the share-region and different-region of a hypothesis pair; a_1, a_2 and a_3 are three sample lines of region $\boldsymbol{a}$, and b_1, b_2 and b_3 are three sample lines of region $\boldsymbol{b}$. The structure similarity is estimated as follows:

$$\begin{aligned} D_{int}(a,b) &= \frac{1}{n}\sum_{i=1}^{n} Dis(a_i,b_i) \\ Dis(a_i,b_i) &= \frac{1}{m}\sum_{j=1}^{m} |a_i(j) - b_i(j)| \end{aligned} \quad (3)$$

where $D_{int}(a,b)$ is the structural similarity for region $\boldsymbol{a}$ and $\boldsymbol{b}$, a_i and b_i are sample lines, n is the number of sample lines; $Dis(a_i,b_i)$ is the intensity distance of two sampled lines a_i and b_i; $a_i(j)$ and $b_i(j)$ are the intensity values of the j-th pixel on line a_i and b_i; m is the number of pixels on lines a_i and b_i.

The intensity distribution similarity of region $\boldsymbol{a}$ and $\boldsymbol{b}$ are measured by:

$$D_{hs}(a,b) = \frac{1}{2}\{KL(h(a) \mid\mid h(b)) + KL(h(b) \mid\mid h(a))\} \quad (4)$$

where $h(a)$ and $h(b)$ are the normalized intensity distribution histogram of region $\boldsymbol{a}$ and $\boldsymbol{b}$; $KL(p \mid\mid q) = \sum_{x\in\chi} p(x)\log\frac{p(x)}{q(x)}$ is the KL-divergence of two probability distributions p and q.

3. Supporting Regions: Different parts of the same building may have different appearances, so only computing the interior similarity can not completely describe the consistency of two regions. Meanwhile, the surroundings of buildings, *e.g.* road, park or lawn, may be similar in many cases. Hence, these surroundings, which we called supporting regions, can be used as auxiliary information to assess the consistency of two building regions. Figure 6.(B) shows an example about supporting regions, $(\boldsymbol{a}_1, \boldsymbol{b}_1)$ and $(\boldsymbol{a}_2, \boldsymbol{b}_2)$ are the supporting regions of region $\boldsymbol{a}$ and $\boldsymbol{b}$. According to these four supporting regions, two intensity similarities can be computed by equation 4: $D_{hs}(a_1, b_1)$ and $D_{hs}(a_2, b_2)$.

Besides the above four supporting regions, Fig 7.(B) shows another important supporting region $\boldsymbol{c}$, which is a symmetric region of $\boldsymbol{b}$ and used to help us to measure the consistency of $\boldsymbol{a}$ and $\boldsymbol{b}$. For example, if $\boldsymbol{b}$ is more similar to $\boldsymbol{c}$ than $\boldsymbol{a}$, then $\boldsymbol{b}$ is not likely to be a building region, so $\boldsymbol{a}$ and $\boldsymbol{b}$ should be considered as inconsistent. $D_{int}(b,c)$ and $D_{hs}(b,c)$ can be computed by the equation 3 and 4.

4. Image Evidence Vector Based on above three types of image evidences, we can compute an image evidence vector for each

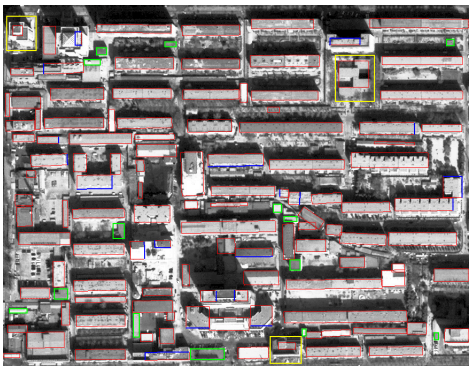

Figure 8: Image **A**. Red rectangles are the results of our algorithm. Green rectangles are lost buildings. Purple rectangles are detected only by our algorithm. Blue lines are the error edges. Yellow rectangles are buildings not used for the statistic of error edges.

hypothesis pair. Fig. 7 shows an example, for hypothesis pair h_1 and h_2, the image evidence vector is: $\{Sa(a_l)$,$Sa(b_l)$,$Sa(a_s)$, $Sa(b_s)$,$D_{int}(a,b)$, $D_{hs}(a,b)$,$D_{int}(b,c)$, $D_{hs}(b,c)$,$D_{hs}(a_1,b_1)$, $D_{hs}(a_2,b_2)\}$. Thus, for each hypothesis pair, a 10-D image evidence vector is computed.

4 EXPERIMENTAL RESULTS AND ANALYSIS

This algorithm has been tested on a number of examples with good results. Two typical results are shown in Fig 8 and Fig 9 and some problems are outlined. The extraction evaluation are shown in Table 2. Three measures are used to evaluate the quality of the extraction result s:

- **DP** $=100 * TP/(TP+TN)$
- **BF** $=100 * FP/(TP+FP)$
- **CEP**$=100 * (TP*4-EE)/(TP*4)$

DP: Detection Percentage; **BF**: Branch Factor; **CEP**: Correct Extract Edge Percentage. TP(True Positive): a building detected by both a person or program; TN(True Negative): a building detected by a person not by a program; FP(False Positive): a building detected by program not by a person; EE: the number of incorrect edges of the buildings detected by both a person or program(TP). The **DP** and **FE** are introduced by paper (Lin.C and Nevatia.R, 1998), which are used to evaluate the building detection results. The **CEP** is proposed to evaluate the accuracy of edge extraction results.

Image	*TP*	*TN*	*FP*	*EE*	**DP**	**BF**	**CEP**
A	111	14	4	**77**	88.8	3.40	**81.6**

Table 1: Result of paper (Song et al.,n.d.)

Image	*TP*	*TN*	*FP*	*EE*	**DP**	**BF**	**CEP**
A	111	14	0	**29**	88.8	0.00	**93.0**
B	100	7	2	22	93.4	1.96	94.4
average	106	11	1	26	91.1	0.98	93.7

Table 2: Result Evaluation.

Fig 8 shows the building extraction result of the test image of paper (Song et al., n.d.). Compared with (Song et al., n.d.), besides the good detection percentage(88.8%), our method has obtained a great improvement on the results accuracy. The quantity of the error edges has decreased from 77 to 29, most error edges of paper (Song et al., n.d.) has been correctly extracted, especially for the independent buildings, nearly no edges are incorrectly extracted. To demonstrate the robustness and feasibility of our algorithm, another test image is shown in Fig 9. Compared with Image A, Image B is more difficult to be deal with, most buildings have been segmented into too small fragments and some of them are combined together. Besides, the shape of most buildings are not strictly rectangular and the direction of most buildings are not strictly horizontal. Method (Song et al., n.d.)can not work well on this image, while our method can still perform very well. The detection percentage of our paper is 93.4%, and the number of error edges is 22, which is still much smaller than 77 of paper (Song et al., n.d.).

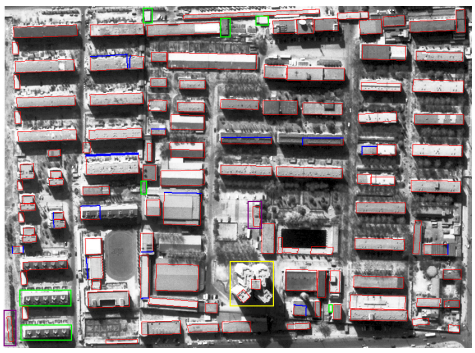

Figure 9: Image **B**.

From table 2, it is clear that our algorithm can effectively deal with the building extraction from complex urban scenes. More importantly, our method can accurately extract building edges regardless of the complex surroundings and substructures of buildings. For the whole about 232 buildings, it detects about 211 buildings, loses about 21 buildings, the average Detection Percentage is about 91.1%. For the 844 edges, about 793 edges are correctly extracted, 51 edges are incorrectly extracted, the average Correct Edge Extract Percentage is about 93.7%.

From the 21 lost buildings, we can observed that most lost buildings have very small size or being too black or too white. Our algorithm can not work well on these kinds of buildings. For the 51 error edges, about 1/3 of them are incorrectly extracted by our program, others are mainly raised by the occlusion problem or the buildings combination problem. Hence, in the future, we intent to enlarge the training set to further improve the robustness and accuracy of our algorithm, and introduce more complex models to deal with the buildings combination problem. Besides, the assistant data source such as DEM or GIS which can avoid the occlusion problem or the small size problem is also taken into our consideration.

5 CONCLUSION

In this paper, a new hypothesis verification algorithm is proposed, which can accurately extract the rectangular buildings from the large and complex urban scenes, and the experimental results about the 250 buildings has demonstrated that our algorithm has a high detection rate and good result quality. Especially, by introducing the boosting method and the optimization frame to the edge verification algorithm, the edge extraction precision has got a great improvement. We believe that our methods will supply great help for the further 3-D building extraction from multiple images or images integrated with other data sources. Besides, this algorithm can also be generalized to other data sources.

The whole flow of this algorithm looks like very complex, also it has multiple levels, but actually the edge verification method as the key part of this algorithm is easy to be understood: image evidences are used to estimate the constraints of any two possible edges, and then these constraints are used to compute the optimal edge. Although the current training set is not large enough, the experimental result is also acceptable. We believe that along with the increase of the training set and the introduction of the complex building model, our algorithm will perform much better.

REFERENCES

Baillard.C and Zisserman.A, n.d. Automatic reconstruction of piecewise planar models from multiple views. CVPR 99,Colorado,USA.

Baltsavias, E., Gruen, A. and Van Gool, L., 2001. Automatic extraction of man-made objects from aerial and satellite images (iii). A.A.Balkema,Lisse.

Collins, R. T., Jaynes, C. O., Cheng, Y.-Q., Wang, X., Stolle, F., Riseman, E. M. and Hanson, A. R., 1998. The ascender systerm:automated site modeling from multiple aeria images. CVIU 72, pp. 143–162.

D.Koc and M.Turker, 617–622. Automatic building detection from high resolution satellite images. RAST 05.

Fischer, A., Kolbe, T. H., Lang, F., Cremers, A. B., Forstner, W., Plumer, L. and Steinhage, V., 1998. Extracting buildings from aerial images using hierarchical aggregation in 2d and 3d. CVIU 72, pp. 185–203.

Friedman.J, Hastie.T and Tibshirani.R, n.d. a statistical view of boosting. Annals of Statistics.

Haala, N. and Anders, K., 1996. Fusion of 2d-gis and image data for 3d building reconstruction. International Archives of Photogrammetry and Remote Sensing 31, pp. 285–290.

Haala, N., Brenner, C. and Statter, C., 1998. An integrated system for urban model generation. In ISPRS Commission II Symposium,Cambridge,England.

Irving.R and Mckeown.D, 1989. Methods for exploiting the relationship between buildings and their shadows in aerial imagery. 19, pp. 1564–1575.

Kim, Z. and Nevatia, R., 2004. Automatic description of complex buildings from multiple images. CVIU 96, pp. 60–95.

Lee.S, Shan.J and Bethel.J.S., n.d. Class-guided building extraction from ikonos imagery. Photogrammetric Engineering and Remote Sensing.

Lin.C and Nevatia.R, 1998. Building detection and description from a single intensity image. CVIU 72, pp. 101–121.

Mayer, H., 1999. Automation object extraction from aerial imageryca survey focusing on buildings. CVIU 74(2), pp. 138–149.

McGlone.J and Shufelt.J.A, n.d. Projective and object space geometry for monocular building extraction. CVPR 94,Seattle,Washington,USA.

M.Fradkin, H.Maitre and M.Roux, 2001. Building detection from multiple aerial images in dense urban areas. CVIU 82, pp. 181–207.

N.Paparoditis, 1998. Building detection and reconstruction from mid-and hing-resolution aerial imagery. CVIU 72, pp. 122–142.

Roux.M and Mckeown.D.M, n.d. Feature matching for building extraction form multiple views. CVPR 94,Seattle,Washington,USA.

San, D. and M.Turker, n.d. Automatic building detection and delineation from high resolution space images using model-based approach. ISPRS Workshop, Topographic Mapping from Space (with Special Emphasis on Small Satellites).

S.Noronha and R.Nevatia, 2001. Detection and modeling of buildings from multiple aerial images. 23, pp. 501–518.

Song, Z., Pan, C. and Yang, Q., n.d. A region-based approach to building detection in densely build-up high resolution satellite image. ICIP 06,Atlanta,GA,USA.

Suveg, I. and Vosselman, G., 2004. Reconstruction of 3d building models from aerial images and maps. ISPRS Journal of Photogrammetry and Remote Sensing 58, pp. 202–224.