Perceptual Grouping of 3-D Features in Aerial Image Using Decision Tree Classifier *

In Kyu Park
School of Electrical Eng.
Seoul National University
Seoul, 151-742, KOREA
pik@sting.snu.ac.kr

Kyoung Mu Lee
Dept. of Electrical and Electronics Eng.
Hong-Ik University
Seoul, 121-791, KOREA
kmlee@wow.hongik.ac.kr

Sang Uk Lee
School of Electrical Eng.
Seoul National University
Seoul, 151-742, KOREA
sanguk@sting.snu.ac.kr

Abstract

We address a new perceptual grouping algorithm for aerial images, which employs a decision tree classifier and hierarchical multilevel grouping strategy in a bottom-up fashion. In our approach, grouping is performed perceptually on 3-D features extracted from 2-D images, in which the Gestalt principles including collinearity, parallelism and L-typed convergence are encoded by the decision tree learning technique. The decision tree is constructed using training samples obtained from the given 3-D reference model. Then, each pair of the extracted 3-D line features of an input image is classified into one of the learned Gestalt primitives. On the other hand, in multilevel grouping procedure, grouping of collated features are performed from lower to higher level, yielding the structured target model. In order to evaluate the proposed algorithm, experiments are carried out on RADIUS model board images. The results show that grouping is performed effectively to extract man-made structures in aerial images.

1 Introduction

In computer vision, object recognition from a complex scene has been an ultimate goal of the research during the past a few decades. Although lots of recognition algorithms have been developed so far, the performances in uncontrolled real situation are far from the satisfaction, compared with the human visual ability. As is well known, grouping of primitive features is one of the most important preprocessing steps for object recognition, since it not only provides higher-level information but also reduces the computational complexity of the process significantly. In this context, grouping features of the same object is, known as perceptual grouping, is more desirable [1, 2, 5].

Over the past several years, computer vision community has noticed various applications of perceptual grouping in many computer vision tasks including stereo matching, model indexing, contour completion, figure-ground segmentation, change detection, and more [1, 2, 5]. The behind notion of perceptual grouping is that image features are not distributed at random, but exhibiting regular pattern, known as the Gestalt principles. In addition, according to the Gestalt psychology, elements in an image are grouped from parts into wholes, based on proximity, parallelism, closure, symmetry, and continuation.

Although perceptual grouping for generic object is not a simple task, owing to the structural regularity of the underlying primitive features, the perceptual grouping approach is considered as a powerful methodology for the problem of extraction and recognition of man-made objects in aerial scenes. However, most existing perceptual grouping algorithms have several inherent drawbacks; they are often subject to heuristics, depending on several parameter settings, and therefore the flexibility is degraded from case to case. Moreover, since they rely on simple geometric properties in 2-D feature space, they would often fail in the noisy images such as aerial scenes.

In this paper, a new robust perceptual grouping algorithm for 3-D features is proposed, which employs a decision tree classifier and a hierarchical multilevel grouping strategy in a bottom-up fashion. The main idea of the proposed algorithm is to utilize not only the Gestalt principles but also the information provided by the given model to improve the robustness and the pruning capability. By employing a decision tree classifier [3, 4] which are trained by the Gestalt principles using the model, first-level grouping is performed di-

*This work was supported by the Agency for Defence Development, Taejon, Korea, and the Automatic Control Research Center in Seoul National University, Seoul, Korea.

0-7803-5467-2/99/ $10.00 © 1999 IEEE

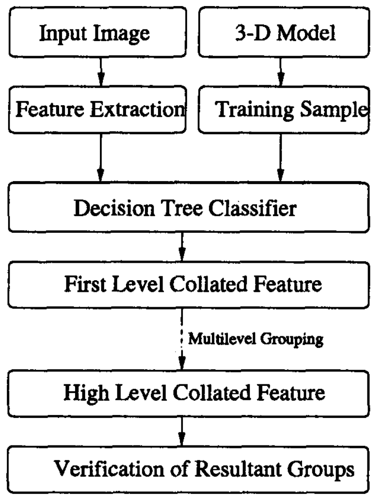


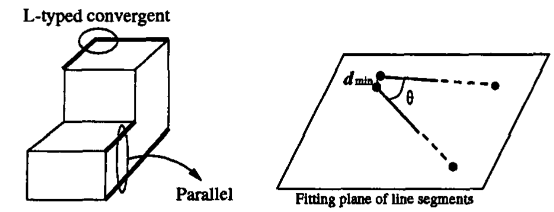
Figure 1: Block diagram of the proposed grouping procedure

rectly, resulting in collated features of collinear, parallel, and L-typed convergent pairs. After the initial grouping, by examining mutual relations between current-level features, subsequent higher level grouping is performed to produce higher-level collated features. Note that since decision tree is known to be very robust to the noise and missing attributes, the proposed approach is also very robust to the input noise. Moreover, unlike conventional algorithms, since the threshold values are determined automatically in decision tree learning, the proposed algorithm is more flexible in real situations.

This paper is organized as follows. Section 2 presents feature extraction and preprocessing technique briefly. In Section 3, decision tree learning and classification algorithm is described in detail. Then, multilevel grouping technique is presented in Section 4. Finally, experimental results and conclusion are provided in Section 5 and 6, respectively.

2 Feature Extraction

In our approach, it is assumed that the model object to be recognized has polyhedral representation, which is common to most man-made objects. Since the line edge drawing can characterize the 3-D model efficiently in 2-D image, line segments are used as the feature. We also assume that the 3-D depth information of the end points of the line segments are provided by other vision techniques such as stereo analysis and range-finding method. Thus, the extracted 3-D line segments are used as the input of the proposed grouping algorithm, which is described in the following sec-



(a) Model and instances (b) An instance and its attributes

Figure 2: Model and training examples

tion.

3 Learning Gestalt Principles Using Decision Tree Classifier

In the first-level grouping procedure, collinear, parallel, and L-typed convergent pairs are extracted by employing decision tree learning technique [3, 4]. Note that they are fundamental elements of the Gestalt primitives. Decision tree classifies instances by sorting them down the tree from the root to some leaf node, which provides the classification of the instance. Each node in the tree specifies a test of some attribute of the instance, and each branch descending from that node corresponds to one of the possible values for this attribute. Decision tree is constructed by learning training examples, which are actually obtained from a given model in our approach. In this procedure, tree learning task is performed by using the well-known ID3 algorithm [3]. Once the decision tree is constructed, then, each pair of input line segment is classified by sorting it through the tree to the appropriate leaf node, then returning the classification associated with this leaf. In this section, the key ingredient of decision tree learning and classification in the proposed algorithm is described in detail.

3.1 3-D Reference Model and Training Samples

As a usual model-based approach, we have 3-D CAD model as shown in shown in Figure 2 (a). Training samples are the set of each pair of 3-D line segments in the model. They are represented by a vector form, in which some geometric relationships are recorded. In real implementation, in order to perform the grouping adaptively and robustly, noise-added models as well as the exact 3-D reference model are used for the training.

3.2 Target Class

In our approach, each instance encodes the relationship of a pair of line segments, and the target classes represent the results of the first-level grouping into

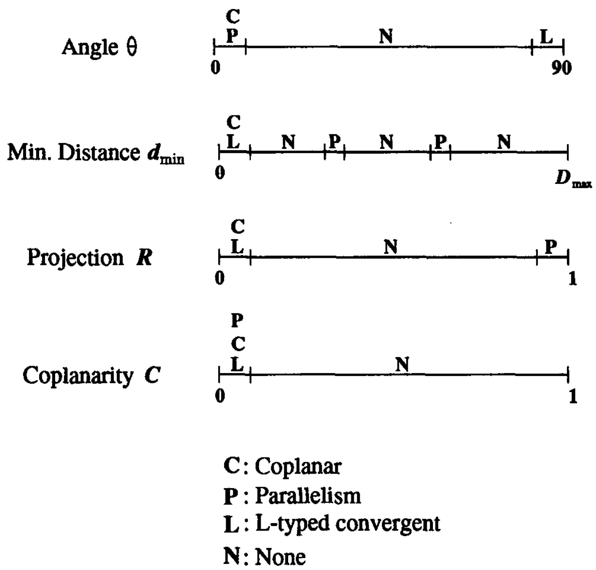


Figure 3: Discretization of continuous attributes

four categories including collinear, parallel, L-typed convergent, and none.

3.3 Node Attributes

Node attributes should provide some measure to distinguish the geometric differences among different target classes. In this research, since 3-D line features are used as the primitives, 3-D geometric properties between a pair of line segments are encoded as the node attributes. In order to distinguish L-typed convergent from collinear and parallels, the between-angle property (θ) is first considered, as shown in Figure 2 (b). Then, minimum distance (d_{min}) is also used to encode proximity. Parallel segments are required to overlap, which means that the normal projection of one on another should be large enough. Thus, the projection ratio (R) is included as an attribute. In addition, coplanarity is another important condition for line segments to be grouped. Since 3-D line segments are used, the coplanarity (C) can be evaluated by calculating the deviation of points from the fitted plane. In summary, an instance is represented by an attribute vector, which consists of these four elements, *i.e.*,

$$I = < \theta, d_{min}, R, C > \quad (1)$$

3.4 Discritization of the Continuous Attributes

The attributes in a learning problem may be categorical or numerical. Since ID3 assumes that all

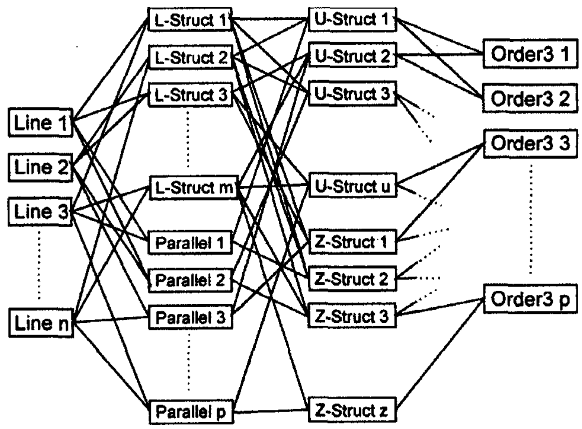


Figure 4: Multilevel grouping

attributes are categorical, the numerical attributes should be discretized prior to tree construction. In our approach, fortunately, the range of each numerical attribute can be divided into a few subintervals without difficulties, as shown in Figure 3. For instance, the between-angles θ of collinear or parallel pairs are near zero, while they are near 90 degrees for L-typed convergent pairs. Thus, the range for angle attribute is divided into three subintervals, denoting that there is three categorical attribute for angle. For other attributes, subintervals are obtained in similar way, by considering geometric characteristics of each target class. Note that, for minimum distance attribute, the range may be divided into several subintervals according to the number of possible parallel pairs. The width of subinterval is determined by the given reference model and a few thresholds. That is, for each **C**, **P**, and **L** instances, we allow small intervals of width δ, then cover the whole range with the intervals. The maximum and minimum of each **C**, **P**, and **L** interval are determined to be the supremum and infimum of the union of the intervals of the consisting instances.

4 Multilevel Grouping

In the proposed algorithm, features are collated from bottom to top, yielding 4-levels of collated features, *i.e.* level-1 (line segment), level-2 (collinear, parallel, and L-typed convergent), level-3 (U-structure), and level-4 (rectangle and closure). Higher-level collated features are obtained by combining current-level features. In order to perform this, each pair of current-level feature is classified as mutually supportive pair or mutually conflictive pair. A pair of current-level feature is considered as mutually supportive, if they share a common lower-level feature.

On the other hand, a pair of current-level feature is regarded as mutually conflictive, if no mutually supportive feature can be found from the possible sub-features of current level. Current-level features can never form higher-level feature, if they are mutually conflictive. Note that higher-level feature is obtained by combining the mutually supportive feature pair. For example, from the parallels and the L-typed convergent pairs, U-structured collated features are obtained, if they share a line segment. Figure 4(b) shows the multilevel grouping network, implying bottom-to-top grouping procedure.

5 Experimental Result

In order to evaluate the performance of the proposed algorithm, experiments are carried out on the RADIUS model board images, in which 3-D coordinates of each vertex of building as well as several 2-D images are provided. In this experiments, we synthesized 3-D graphical model by mapping textures extracted from 2-D images on the 3-D coordinates. In Figure 5, a 3-D reference model is shown, which is used to obtain training samples of the decision tree. Therefore, grouping process is guided to extract Gestalt primitives in the input image, of which the geometric properties are similar to those of the given reference model. It can be done by decision tree learning, which indeed encodes the Gestalt properties of the reference model.

The grouping results are shown in Figure 6 and Figure 7. Figure 6 (a) is the input image which is obtained from top view. The accompanied depth image is shown in Figure 6 (b). In Figure 6 (c), the extracted line features are shown. Beginning with the line feature, Gestalt primitives, *i.e.*, the first level collated features, are extracted as in Figure 6 (d-f), in which collinear, parallel, and L-typed convergent pairs are shown, respectively. Then, second level collated features are constructed as shown in Figure 6 (g). One of the final grouping results is shown in Figure 6 (h), which consists of the features from the model object embedded in the input image. Note that the whole grouping process is carried out in 3-D feature space.

In Figure 7, another experimental result is provided, in which the input image is obtained from an oblique view. The results show that the grouping is performed effectively to extract the model object in the scene.

6 Conclusion

In this paper, we achieve a perceptual grouping technique for 3-D features, which employs decision tree classifier and hierarchical multilevel grouping strategy in a bottom-up fashion. The main idea of the proposed algorithm is to utilize not only the Gestalt principles but also the information provided by the given model to improve the robustness and the pruning capability. Two set of experimental results on aerial images demonstrate that the proposed perceptual grouping algorithm is quite effective even in noisy environment.

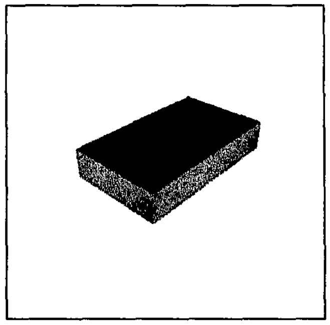


Figure 5: Reference model

Our future work includes grouping-based object recognition and tracking in image sequence. By utilizing the reliable grouping results, the recognition and tracking performance would be much improved.

References

[1] R. Mohan and R. Nevatia, "Using perceptual organization to extract 3-D structures," *IEEE Trans. on Pattern Analysis and Machine Intelligence,* vol. 11, no. 11, pp. 1121-1139, Nov. 1989.

[2] P. Havaldar, G. Medioni, and F. Stein, "Perceptual grouping for generic recognition," *International Journal of Computer Vision,* vol. 20, no 1/2, pp. 59-80, 1996.

[3] J. R. Quinlan, "Induction of decision trees," *Machine Learning,* vol. 1, pp. 81-106, 1986.

[4] U. M. Fayyad and K. B. Irani, "Multi-interval discretization of continuous-valued attributes for classification learning," *International Joint Conference on Artificial Intelligence,* pp. 1022-1027, Chambery, France, September, 1993.

[5] S. Sarkar and K. L. Boyer, *Computing Perceptual Organization in Computer Vision,* World Scientific, 1994.

[6] D. C. Nadadur, X. Zhang, and R. M. Haralick, "Groundtruth outline drawing in model board images," *Technical Report No. ISL-TR-94-01,* University of Washington.

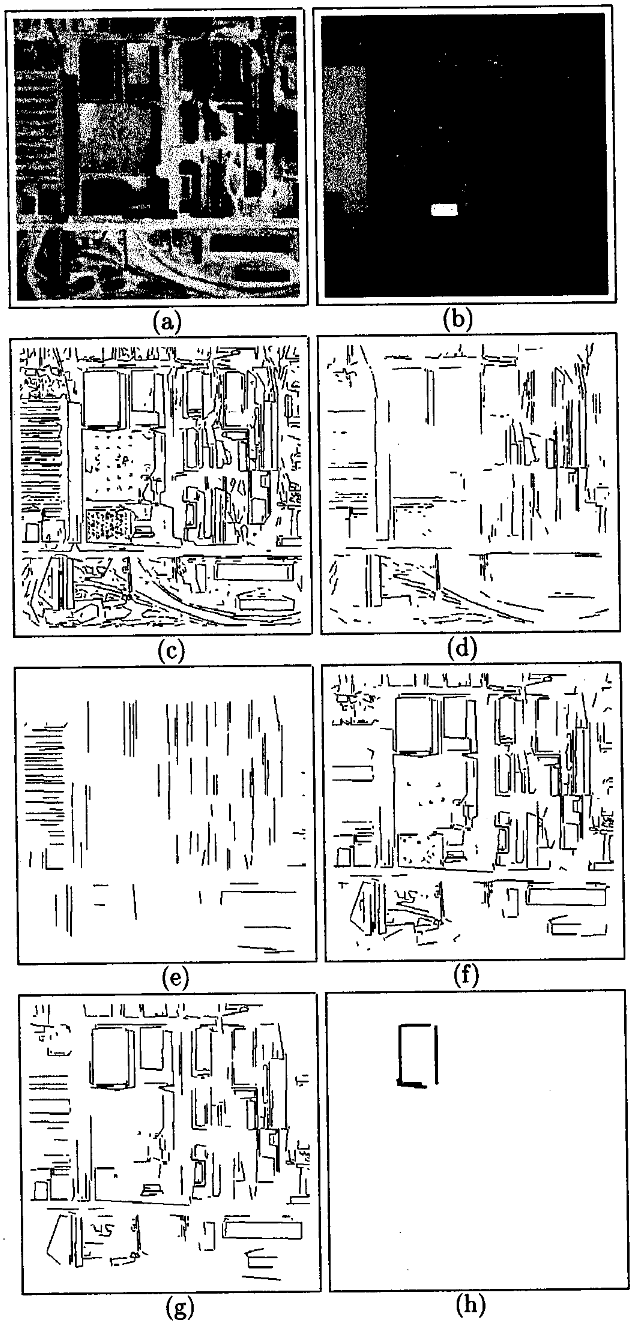


Figure 6: Multilevel grouping result (a) Input scene (from top view) (b) Depth image (c) Extracted line segments (d) Detected collinear pairs (e) Detected parallel pairs (f) Detected L-typed convergent pairs (g) Detected U-structures (h) Grouping result

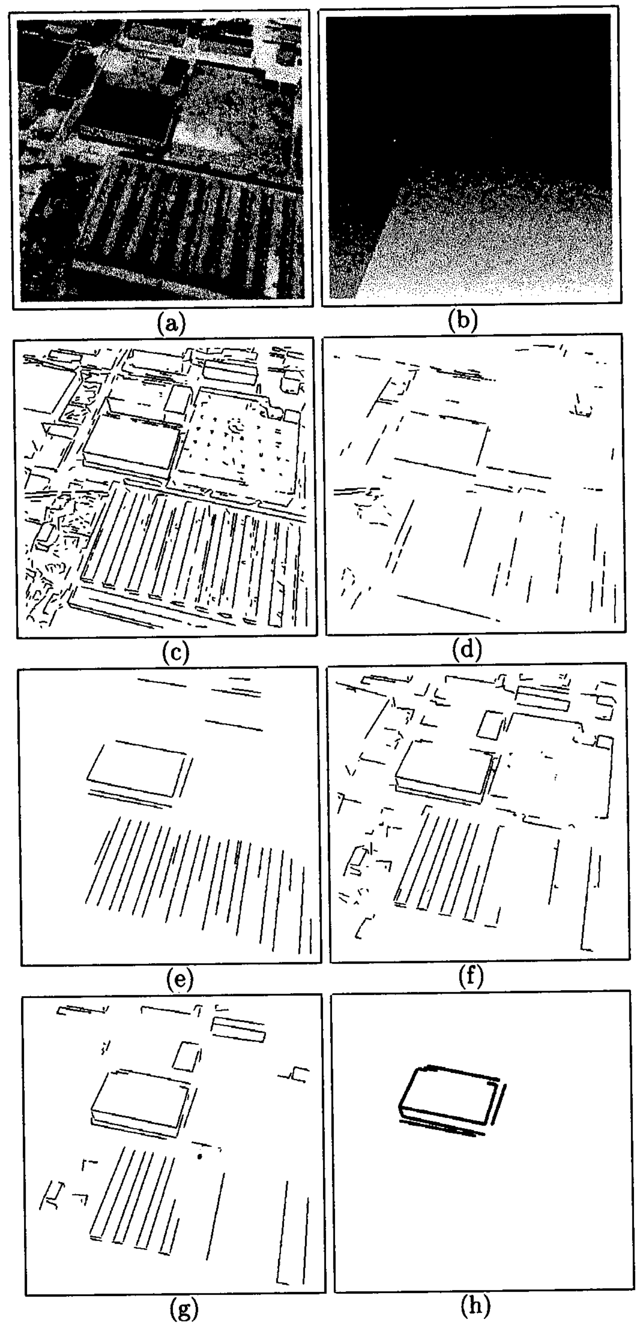


Figure 7: Multilevel grouping result (a) Input scene (from oblique view) (b) Depth image (c) Extracted line segments (d) Detected collinear pairs (e) Detected parallel pairs (f) Detected L-typed convergent pairs (g) Detected U-structures (h) Grouping result