Heli-Tele: Road Extraction from Helicopter Video

Fatih Porikli, Jie Shao
Mitsubishi Electric Research Labs
201 Broadway, Cambridge, 02139, USA
fatih@merl.com

Hide Maehara
Information Technology R&D Center
5-1-1-Ofuna, Kamakura, Kanagawa, Japan
maeh@isl.melco.co.jp

Abstract

We present a learning based road likelihood computation method that uses aerial imagery and fuses information from several weak features. Our method is automatic, robust, and computationally feasible at the same time. We use road likelihood to align a color helicopter image onto a given street map to find address of buildings visible in the image.

1 Introduction

Alignment of aerial imagery and street maps is major challenge that geographic information systems are facing nowadays. In our setup, we want to determine the address of a chosen location in an image that is captured from a low-flying helicopter. One important application is automatic emergency services e.g. finding the address of a building in fire using aerial imagery. Although GPS information is also available during the flight, it is often noisy with an off-set of 20 meters due to the motion of the helicopter and limited resolution of the GPS data. Therefore, additional refinement is necessary using the image and available street map using the only mutual features, roads, in both image and map as illustrated in Fig.1. Since extraction of roads is a time-consuming and it can not be performed manually for a real-time system, there is a need for automation.

Several approaches extract road candidates and then track roads [1]. One method models the context, such as shadows, cars, tree, etc. to improve the extraction of roads [2]. Learning methods were introduced as alternative automatic method by using grouping of parallel segments [3], detecting ridge-like descriptors using multi-scale methods [4]. Hough transform for the extraction of crossings [5]. Several methods make use of texture features. However, most of the existing approaches are either based on the hard heuristics and very specific to the type of the input data or not robust towards the various road and imaging conditions exist in our application. Note that, it is sufficient to obtain road likelihood for each pixel, but not to precisely extract roads since such likelihood information is all is needed to align the input image to the street map.

To obtain a road likelihood map, we propose an automatic, robust, computationally feasible approach that uses low and high level image features. We selected features that provide most discriminating information by a learning based method, and tested several classifiers to achieve the accuracy and computational simplicity at the same time. In the following sections, we present these features and evaluation of classifiers, and sample results. Our system has already implemented as a part of our commercial aerial image analysis product.

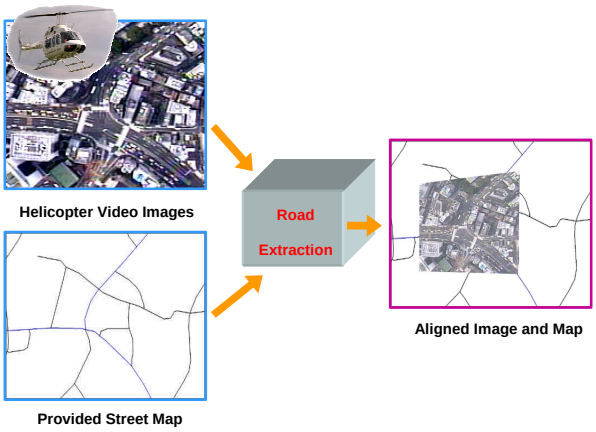

Figure 1: Alignment requires extraction of roads.

2 Road Characteristics

There are several weak cues that indicate roads, however, they are mostly not sufficient by themselves:

- Roads have salient edge features caused by lane dividers and intensity discontinuousness between buildings.
- Edges on the both sides constitute a pipe-line structure.
- Roads usually have homogeneous local orientation distributions.
- Roads are continuous, so are contours.
- Width is almost constant and has an upper bound.
- Local curvature changes in a continuous manner except at cross-sections. Yet most roads are straight locally.
- Density of roads is proportional with the surrounding context.
- Road surface texture is different from buildings.
- Roads have a color range, i.e. they are not green or red.

These cues have advantages and drawbacks. Continuous contours with appropriate curvatures and parallel borders indicate roads, but they may fail when edges are occluded, shadowed, or simply not visible. While texture property may discriminate road regions, depending the type of the road (highway, side-street, etc) texture also changes. Local orientation is relatively uniform when it is compared to non-road areas. However, large open areas, large roof-tops may also present such uniform orientation distributions which are easily confused with road regions.

After all, it is not possible to count on a single cue to classify road and non-road regions accurately. That is why we learn the discriminative features from the data and fuse detection results to achieve robustness.

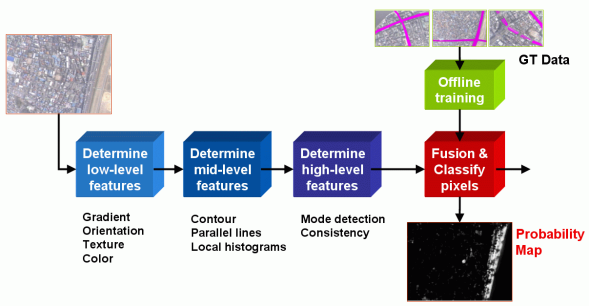

Figure 2: Flow diagram of road likelihood computation

3 Features Selection

To find road likelihood, we compute low-level (gradient, etc), mid-level (contour, etc), and high-level-features as shown in Fig.2. Then, we fuse the feature responses using a classifier that we trained off-line by a set of road and non-road ground truth data that consist of more than 1000 manually marked images. A sample input image is shown in Fig.3. Since the computational complexity is an issue, we compute our high level features for overlapping windows instead of for all pixels. The size of the windows is smaller than the narrowest road.

Orientation information is involved in many features; hence its accuracy is crucial in the entire process. We use a robust estimator: First we apply a pixel-wise adaptive 2-D Gaussian low-pass Wiener filter. The filter uses estimate the gradient mean and standard deviation within local windows. We compute horizontal and vertical gradient magnitudes at each pixel in a block, and we apply the same 2-D Gaussian low-pass filter to the gradients. And, within each block, we aggregate the local orientation. Sample orientation results are given in Fig.4, in which we plot the normal orientation.

We use statistical properties of local orientation distribution to determine more complex features. We have generated several possible statistical features and ratified each feature according to their discriminative power using a set of ground truth images. Some of these features involve a histogram of orientation that is quantized into 12 bins. Some of these features are the maximum of histogram, mean, variance and entropy of the orientation

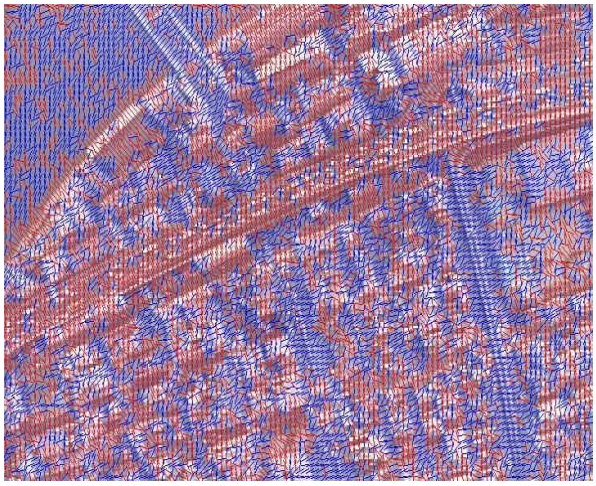

Figure 4: Local orientation map

Figure 3: Yellow lines shows roads to be detected.

distribution within the window, entropy of weighted orientation histogram, convolution of orientation histogram with single Gaussian function and dual (located at opposite angles) Gaussian functions. The maximum value of local orientation histogram reflects whether a principal orientation is existed in local window. A higher value means the distribution of the local orientations is more uniform. Same is true for entropy and variance. Convolution indicates the existence of a dominant orientation direction.

Contour-based features (Fig. 5) complement orientation based features as they capture more global properties, such as continuous edges, parallel lines and relatively straight paths. The contour detector works on line images, which is constructed using a curvilinear structure estimation method [6]. We remove the contours that are in closed curves and shorter in length. It is true that large buildings boundary contours can be easily confused with road contours. One simple solution may is to count vertices of contour segments with polylines. For contours caused by building edges, the number of polygon vertices is usually large, while for contours caused by roads the number of the vertices of polylines tends to stay small. We observed these are the most discriminative contour-based features: length, compactness, number of pixels marked as contour

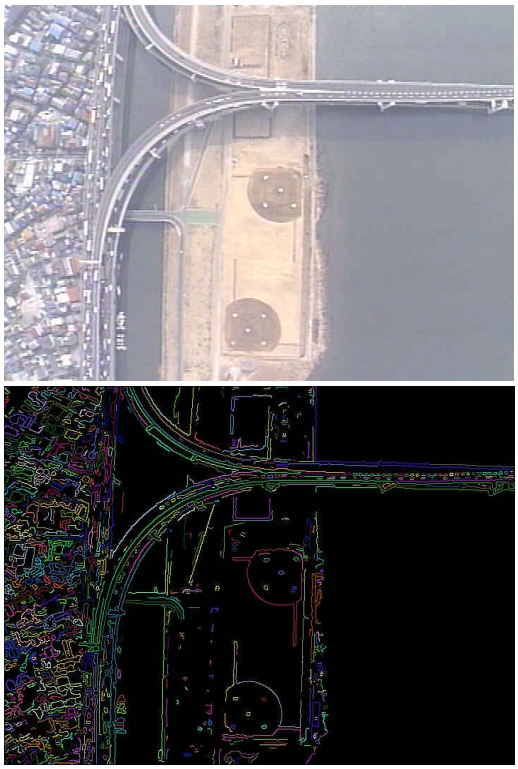

Figure 5: Contours before refinement.

METHOD	Correct Detection ratio	False Alarm ratio
Lane Detectors (no threshold)	0.1472	0.0290
Lane Detectors (threshold)	0.3168	0.0700
15 Features (no threshold)	0.1829	0.0298
15 Features (threshold)	0.3639	0.0864
10 features (no threshold)	0.3438	0.0546
10 features (threshold)	0.5259	0.0937

Table 1. Performance comparison

in a local window, and entropy of contour orientation. Since most of the road contours are elongated lines, road compactness is relatively low compared to contours caused by other edges. Intuitively, contours with slightly changing curvatures have low entropies.

Color is another weak feature; however, intensity is not reliable due to the fact that vehicles and cast shadows cause significant changes in road intensity. Chrominance gives more information and can be used as a filter to eliminate non-road regions such as green fields, red roof-tops, blue water bodies, etc. We use two color histograms to keep the color distribution of roads and non-road regions. Each color channel has 64-bins. Using ground truth images, we train both histograms.

We tested a set of 24 Gabor filters, 3 spatial frequencies and 8 orientations, and used 2^{nd} order complex moments to extract orientation independent texture features. In addition, we computed texture energy. Although texture is an important feature for high quality aerial imagery, in a helicopter setup where the camera is shaking and significant blur exists in the image, the texture is hardly visible and accurate. Hough transforms (parameter space transformations), on the other hand, are popular to detect dominant lines. But especially in rural regions, the excessive amount of edge and line information disturbs the performance of Hough based detectors. Another complication is that roads are not globally straight, thus, Hough should be able to detect curvatures, which is even less accurate considering the computational restrictions of the system.

We compute the above features and convert them into road likelihood maps using nonlinear weighting functions. Until now, we analyzed the response of each feature with respect to others; next, we combine them into a single likelihood map using classifiers to optimize their mutual detection results.

4 Classifiers

We tested two classifiers: linear and nonlinear. Linear classifier is a combination of likelihood maps corresponding to different features. In other words, it finds a weight vector such that the inner product of feature vector of pixels and weights gives the binary labels of pixels provided in ground truth. Basically, we form a very large feature matrix in which each column is a feature vector of a pixel and multiply it with the weight vector to get the label vector. Note that, the feature matrix is already known and the label matrix is given by ground truth, and we only need to take the pseudo-inverse of the feature matrix and multiply it with the label vector to find weight vector. Although this is straightforward, taking pseudo-inverse of a very large matrix is not always feasible, partial MSE and sub-optimal solutions are needed. Besides, such a linear classifier is limited only with the linear combination of features.

We used a multilayer neural network that is trained by back-propagation using 13 separate likelihood maps. We implemented different network structures and the simplest version that has detection performance still as high as the more complicated structures was a three layers implementation; an input layer with 13 nodes, a 20-nodes hidden layer and one node output layer.

5 Experimental Results

We set our local feature collecting window as 31x31 with 5 pixels overlapping on both directions, while the overlapping rate can be changed as system input parameter according to the resolution requirement.

We compared two versions of the presented neural network classifiers with a state-of-art lane detector method using template matching that was not learning based. One version uses 15 features that explained in the previous section. After further optimizing, we refine the feature set to 10 features. Second classifier does not include features such as color, maximum value of histogram, contour length, which may be reliable indicators for other type of imagery.

We trained our classifiers using 20% of the ground truth data and tested with the remaining 80% images, thus there is no overlap between the training and testing sets. We evaluated the performance by correct detection (percentage of accurately detected roan and non-r

oad pixels) and false alarm (percentage of miss classification of roads as non-road and vice-versa) ratios. The difference between the threshold and non-threshold is that when we evaluated the accuracy for non-threshold version, we used all pixels regardless of their spatial energy i.e. a mixture of variance of gradient and texture. The threshold version performs the same evaluation only on the pixels that are salient where the spatial energy is high. We selected pixels that have more than 0.1 energy score (energy is normalized between [0-1]). Table.1 shows the comparison results. We give samples of road likelihood in Fig. 7. As visible, the detection results are very promising, especially when we consider the wide range of road types

As visible, we achieved to improve the correct detection ratio from 31% (lane detector) to 52% while still keeping the false alarms less than 10%. This is 21% increase in correct detection and only 2% reduction in the false alarm rate.

The performance evaluation result also indicates using any number of features (although they may provide acceptable results for some input images) is, in fact, not the best solution since such features tend to inject noise in the classifier.

6 Conclusions

In this project, we implement an automatic road extraction system by a neural network classifier. We explored several statistical features and used a combination of special set of most discriminative features. The presented method almost doubles the detection accuracy, and our throughout experiments prove the effectiveness of the proposed learning based method.

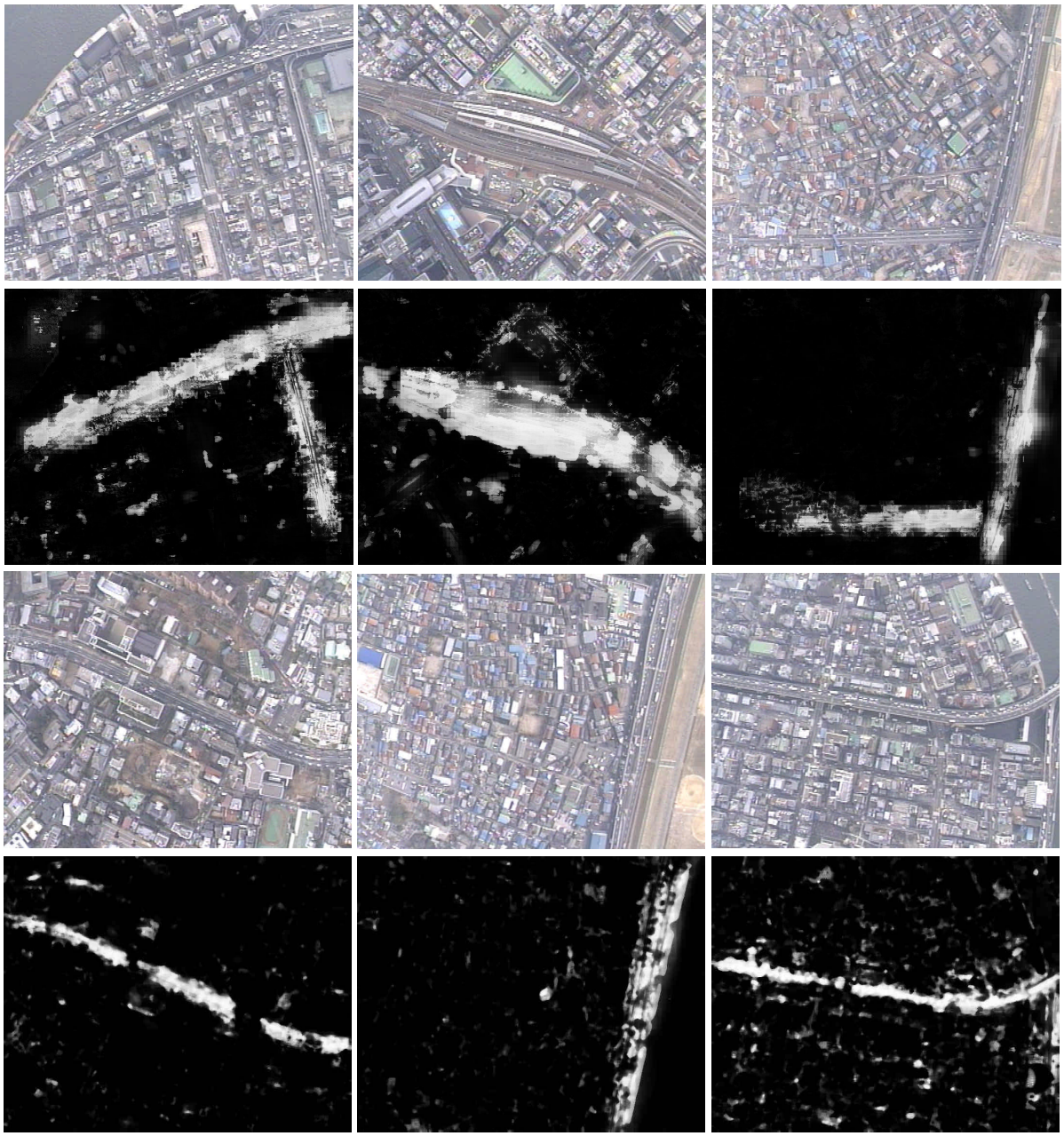

Figure 7: Sample road likelihood maps that will be used in alignment.

One observation from this work is that the blind feature selection and letting the classifier to decide the discriminative features by itself does not always provide the best solution. Unlike most machine learning tasks in computer vision, human expertise plays an important role in application specific definition of features.

References

[1] M. Barzohar, M. Cohen, I. Ziskind, I. D. Cooper, "Fast Robust Tracking of Curvy Partially Occluded Roads in Clutter in Aerial Images," *Aerial and Space Images*, Birkhauser Verlag, pp. 277-286, Basel, Switzerland, 1995.

[2] R. Ruskone, S. Airault, "Toward an Automatic Extraction of the Road Network by Local Interpretation of the Scene, Photogrammetric Week, pp.147-157, 1997.

[3] J. Trinder, Y. Wang, "Knowledge-Based Road Interpretation in Aerial Images", International Archives of Photogrammetric and Remote Sensing, vol. 32, no: 4/1, pp. 635-640, 1988.

[4] S. Pizer, C. Burbeck, J.M. Coggins, D.S. Fritsch, "Object Shape Before Boundary Shape: Scale-space medial axis", *J. of Mathematical Imaging and Vision*, no. 4, pp. 303-313, 1994.

[5] N. Boichis, J. Cocquerez, S. Airault, A Top Down Strategy for Simple Crossroads Extraction", International Archives of Photogrammetr and Remote Sensing, vol.32, pp. 19-26, 1998.

[6] C. Steger, "An Unbiased Detector of Curvilinear Structures", IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 20, no. 2, pp. 113-125, 1998.