Proceedings of the
2007 IEEE Intelligent Vehicles Symposium
Istanbul, Turkey, June 13-15, 2007

A Road Detection Algorithm by Boosting Using Feature Combination

SHA Yun, ZHANG Guo-ying, YANG Yong

Abstract: Road detection is one of the most important branches of road following. In this paper we propose a classification-based road detection algorithm by boosting. To fully utilize potential region feature correlations and improve the accuracy of classification, this algorithm introduces the feature combination method into road detection. First, an over-completed feature set is constructed on several linear and non-linear combined functions. Second, a correlation feature set is selected from the over-completed feature set by feature selection algorithm. Then, the boosting, the support vector machine and the random forest classifiers are used to evaluate the correlation feature set and the raw feature set. The results of the experiment shows the performance of boosting classifier based on the correlation feature set provides the best outcome.

I INTRODUCTION

Among the complex and challenging tasks of future road vehicles is automated driving systems. And vision is one of the most important branches in its development. Driving on paved roads is a well-studied problem, and several vehicles have driven successfully over long distances [1]. For the rural road detection, some classification-based algorithm is presented [2][3][4]. The main idea of the classification algorithm is to classify the road image into road or non-road base on the theory of machine learning using the color, texture and coordination as the features. Feature is one of most important ingredients for building a good classification system. The features of each pixel or region in the road image are color, position, luminance etc. A variety of algorithms have been proposed to transform raw features to optimal features, including PCA, SOM, and ISOMAP. Based on these features, successive classification tasks is greatly simplified. As the computation over the whole image is always time consuming, the evaluation speed of learning algorithm is very important.

Manuscript received January 14, 2007. This work was supported in part by the Beijing Municipal Commission of Education under Grant KM200610017007.

SHA Yun, is with Beijing Institue of PetroChemical Technology, Da Xing Square, 102617 China. She is now with the Department of Computer (e-mail: shayun@bipt.edu.cn).

ZHANG Guo-ying, is with Beijing Institue of PetroChemical Technology, Da Xing Square, 102617 China. She is now with the Department of Computer (e-mail:zhangguoying@bipt.edu.cn).

YANG Yong, is with Changchun University of Science and Technology, Weixing Road, Changchun, Jilin Provance, 130022 China He is now with the Computer Department.(e-mail: yy@cust).

Generally, optimal feature set is very hard to be determined manually by prior-knowledge. Some over-completed feature set together with a feature selection algorithm is widely used for this problem. According to Viola [5] and Xiao's [6] experiment, compared with raw feature, these feature combinations carry more class information and tend to be more robust.

In this paper, to exploiting monocular vision for road detection, the feature combination method is introduced into the road detection algorithm, and the linear and non-linear combination functions are employed at the same time. The article is organized as follows: section 2 presents the feature combination and selection, section 3 is the experiment, and then finally the conclusion.

II Feature Combination and Selection

To detection road region in images, the image is segmented by the region growing technique [4]. These regions in road image can be classified into two classes: road and non-road. For each region, there are four kinds of features: the coordinate, the color, the luminance and the size. Usually the correlation of features is not known.

In order to find powerful features for separating the two classes, two steps are used. The first is feature combination and then the feature selection.

A Feature Combination

There are definitely significant correlations between single features. For example gray color regions (color features) in the middle-bottom of the image (coordinate features) are road. Utilizing these correlations may improve the classification performance. This paper is under the assumption that the linear and non-linear functions can express the feature's correlations.

Given $\vec{x}$ is N -by-1 region feature vector $\vec{x} = [x_1, x_2, \ldots, x_N]$ with corresponding label $y \in \{1,-1\}$, where x_i denotes features of a region in the road image, such as the size, color, luminance, coordinates, and etc. In this paper, four feature combination functions are defined as follows.

Definition 1: The first combine function is $\varphi_0(\vec{x}) = \vec{x}$. Function φ_0 is a linear function.

Definition 2: Given $\vec{I}$ is N-by-1 unit vector, then $\varphi_1(\vec{x}) = [x_1\vec{I} + \vec{x}, \ldots, x_i\vec{I} + \vec{x}, \ldots, x_N\vec{I} + \vec{x}]$, and $\varphi_2(\vec{x}) = [x_1\vec{I}$ -

1-4244-1068-1/07/$25.00 ©2007 IEEE.

$\bar{x},...,x_i\bar{I}-\bar{x},...,x_N\bar{I}-\bar{x}]$, where x_i is the i^{th} feature in feature vector.

Definition 3: The forth combine function is $\varphi_3(\bar{x})=[x_1\bar{x}, x_2\bar{x},...,x_i\bar{x},...,x_N\bar{x}]$.

Then the over-completed combined new feature vector is $\hat{\bar{x}}=[\varphi_0(\bar{x}),\varphi_1(\bar{x}),\varphi_2(\bar{x}),\varphi_3(\bar{x})]$. And the dimension of $\hat{\bar{x}}$ is T.

Function φ_0 ensures the raw feature vector as a subset of the over-completed feature set, because may be the raw features themselves are powerful features for classification. Function φ_1 and φ_2 express the linear relationship of features. These two functions indicate the correlation between features. Function φ_3 extends the linear feature correlation.

Each region features are correlated by these four functions and each region's new feature vector $\hat{\bar{x}}$ is composed. Although the combined feature vectors contain the correlation between features, it may enlarge classification model size, slow down training, utilization speed, and defy the course of dimensionality to improve classification of performance. The features in the over-completed feature vector $\hat{\bar{x}}$ should be selected.

B Feature Selection

There are two kinds of algorithms for feature selection. One is feature selected by ranking them; the other is the subset selection method.

Several papers use feature ranking as a baseline method [7] [8] [9]. Many feature selection algorithms include feature ranking as a principal or auxiliary selection mechanism because of its simplicity, scalability, and good empirical success. But the correlation of these features is not considered in this kind of method, and a feature useless by itself may be useful together with others.

There are two kinds of feature subset selection method: they are wrappers, and embedded methods. Wrappers [10] [11] utilize the learning machine of interest as a black box to score subsets of feature step, independently of the chosen predictor. But, the wrappers are often criticized because they seem to be a "brute force" method requiring massive amounts of computation. Embedded methods [12] perform feature selection in the process of training and are usually specific to given learning machines.

In this paper, two feature selection algorithms: boosting and random forest are tried to select the features from the combined feature set, and the results are compared.

Boosting is a general method which attempts to "boost" the accuracy of any given weak learning algorithm. When single feature in combined feature set is used to compose a boosting classifier, it can be regarded as a feature selection process [9]. To explore the error rate of the single-feature-based boosting classifier, the importance of the feature for classification can be indicated. Because of the feature combination, feature selection algorithm on combined feature set by single feature has included the feature correlation information.

Consider we have T dimension features numbered as $1...T$. Each sample $\hat{\bar{x}}$ is denoted by a T-by-1 vector in feature space, $\hat{\bar{x}}=[x_1,x_2,...,x_T]$ where x_i is the i^{th} feature in feature vector. A weak classifier $h_i(x_i,\theta_i)$ depends on only the i^{th} feature and threshold θ_i. If $x_i<\theta_i$ then $h_i(x_i,\theta_i)=1$, otherwise $h_i(x_i,\theta_i)=-1$.

Where θ_i is the threshold which best separates positive and negative samples by the i^{th} feature, and +1 and -1 are the output in case of x_i classifier less than or greater than θ_i, respectively. The output of $h_i(x_i,\theta_i)$ gives a predication of $\hat{\bar{x}}$'s class label (positive or negative). After training there are T models based on single feature.

The error rate of the weak classifier h_i denotes the classification power of i^{th} feature. Then the power of the i^{th} feature for classification defined as:

$$p_i = 1 - \frac{e_i^{trn}+e_i^{tst}}{2} \quad (1)$$

Where the train error of i^{th} model is e_i^{trn}, the test error is e_i^{tst}. Then the features in combined feature set are sorted according to the p_i. Then feature set can be selected from the combined feature set according to order.

In order to evaluate the performance of feature selection algorithm, the random forest feature selection algorithm is also used in this paper. The random forest algorithm is a classification method, but it also provides feature power for classification [10]. Its principle is as follows: A forest contains many decision trees, each of which is constructed by instances with randomly sampled features. The classification is by a majority vote of decision trees. To evaluate feature importance for classification, first we split the training sets to two parts. By training the first and predicting the second we obtain an accuracy value. For the i^{th}-feature in feature vector, we randomly permute its values in the second set and obtain accuracy. The difference between the two numbers can indicate the importance of the i^{th}-feature.

III Experiments

In the experiments, we evaluated the following aspects of our approach. (a) The performance of feature selection algorithm based on boosting with the random forest algorithm as comparison. (b) The performance of boosting algorithm on region classification. (c) The result of the road detection based on boosting. We begin with explanation of the evaluation method, including image sets and our experiments. Then we give and discuss our experimental results.

A Training Data Prepare

We selected various images with shadows, multi-colored paved surfaces, and poorly-defined borders. The images are from several sets of data collected from a camera mounted on vehicle driving on paved, dirt, and complex environment roads. Only the under skyline regions is useful for driving. So each image under skyline is segmented by color, and the raw features of each region in the segmented image are extracted.

For training purposed, we manually select the region on the road in the image which should be labeled as road. This selection is on the color segmented image. Each region feature vector and its label (the road is 1 while non-road is -1) is one instance in training set. We randomly selected 3/4 images to use for training of the classification methods, and the last 1/4 for evaluation.

Then, the features are combined according to the combine functions. There are 13 features in the raw feature vector; after combination, 520 features in the combined feature vector.

B Feature Selection

The Support Vector Machine is used to evaluate the raw feature set and the selected feature set. Its basic idea is to map data into a high dimensional space and find a separating hyper plane with the maximal margin. This paper uses the Lib-SVM software [15], and the Linear SVM is selected.

We compared the SVM classifier [13] performance with manually annotated frames in order to evaluate the accuracy using different feature set. This allowed us to compute the ratio of false positive and false negative. False positives [3] (FP) refer to error rate of actual non-road regions in the image, which were classified by the system as road. False negatives (FN) refer to error rate of actual road areas classified as non-road. In this paper, in order to emphasis the discriminate of the positive error and the negative error, the false positives ratio (FPR) and the false negatives ratio (FNR) are defined as follows:

$$\mathrm{FPR} = \frac{\mathrm{FP}}{\mathrm{NP}} \qquad (2)$$

$$\mathrm{FNR} = \frac{\mathrm{FN}}{\mathrm{NN}} \qquad (3)$$

Here FP and FN refers to the number of the wrong classified regions, NP and NN refers to the number of positive and negative sample in the training set separately. The accuracy is calculated as follows:

$$\mathrm{Accuracy} = 1 \quad \left(\frac{\mathrm{FP} + \mathrm{FN}}{\mathrm{P} + \mathrm{N}}\right) \qquad (4)$$

Fig.1 shows the performance of the SVM classifier on each set of feature selection by boosting and random forest. Feature number refers to the number of feature that SVM classifier model based on, which are M first features in the combined feature set. The feature set selected by boosting is called the B-Feature Set, while the set by the random forest is called the R-Feature Set. The classification accuracy is the SVM classifier based on different feature set.

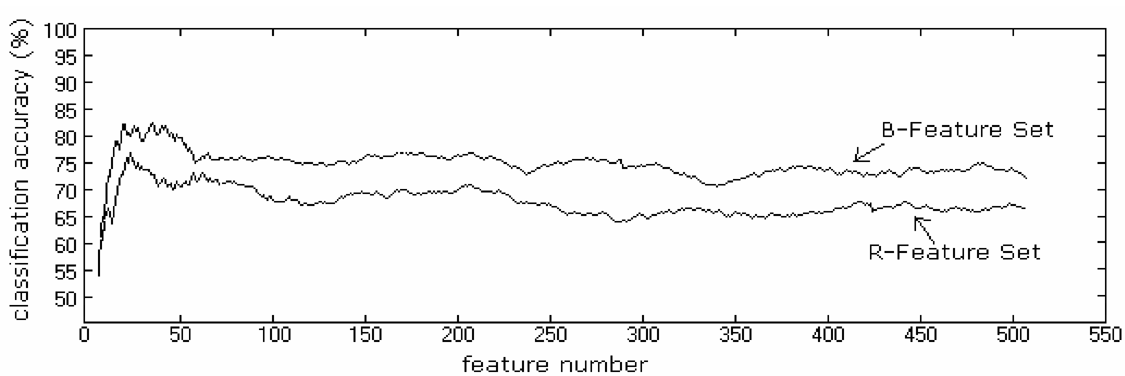

Fig. 1. The classification accuracy of the SVM classifier on different feature set.

From the figure 1, when the number of the feature is about 40, the classifier performance is best, either on the B-Feature Set or on the R-Feature Set. The feature set selection by boosting selection algorithm has comparable performance with that of random forest.

When the feature number is 40, the B-Feature Set is not completely same as the R-Feature Set. Compared with the R-Feature Set, there are more features combined by the function φ_1 and φ_2 in the B-Feature Set. This indicates that the combined features by linear function are more useful for road region classification than the non-linear. The trend of the curves in the figure 1 shows eliminating noisy features can actually improve performance of classifiers.

C The B-Feature Set and the R-Feature Set vs. Raw Feature Set

We evaluate these feature set. As described in section 2.1, we use feature combination between each two features, and the powerful features in over-completed feature sets are selected. We should examine whether selected feature set is valuable for classification. Although it is reasonable in intuition, we need to test the final accuracy by different classifiers. In this paper, three different classifiers are used.

They are Boosting, SVM (Support Vector Machine), and Random Forest.

Therefore, in our experiments we trained the following nine classifier models on three different training feature sets: (a). the SVM model on raw feature set, on B-Feature set and R-Feature set. (b). the random forest model on raw feature set, B-Feature set and the R-Feature set. (c). the boosting model on raw feature set, B-Feature set and the R-Feature set. Table 1 is the result of the nine classification models on different data set.

From the table 1, the performance of the three different classifiers either on the B-Feature Set or the R-Feature Set model performances better than on the raw feature set model. It indicates the combined features can provide more classify information than raw features. The feature combination and selection step is useful for classification.

Table 1 shows the result of the Boosting classifier is better than both SVM and Random Forest. And the FPR is much higher than the FNR. In order to find the reason, the training data set is rechecked.

TABLE I
THE RESULT OF THE CLASSIFICATION BASED ON DIFFERENT FEATURE SET BY DIFFERENT CLASSIFIERS

		Raw	B-Feature	R-Feature
SVM	Accuracy (%)	0.694	0.820	0.726
	FPR (%)	0.542	0.449	0.412
	FPN (%)	0.103	0.058	0.074
Random Forest	Accuracy (%)	0.654	0.807	0.773
	FPR (%)	0.496	0.483	0.404
	FPN (%)	0.091	0.074	0.062
Boosting	Accuracy (%)	0.756	0.904	0.862
	FPR (%)	0.293	0.163	0.121
	FPN (%)	0.189	0.093	0.081

We found there are too few road samples (positive) in the training data set. This is because when the road image is segmented into homogeneous nature color road. The road surface is relatively large regions after segmentation. When manually selected the road region on the segmented road image, the small road regions on the edge of the road often be neglected. So the training set is an extremely imbalanced set. Owing to lack of positive sample, the FPR is relatively high. The boosting classifier result is better than other classifiers, because it's capable of dealing with imbalance data.

In order to examine whether the boosting good performance is only because of the data imbalance, and to improve the performance of the classifier, the training data set is recomposed. For each frame, after segmentation, the road sample and non-road sample is manually selected, while the former training set only select the road region and others are regarded as non-road. These unselected regions in the frame are not recorded in the training data set. In case misclassify the regions on the edge of road are not important for driving. Then we compose a relatively balance training set.

After features combined, the R-Feature Set and the B-Feature Set is re-selected. We also use SVM classifier to analysis the feature selection. When the feature number is 42, the accuracy of the SVM is highest, which is similar to the imbalance training set. While the features in this R-Feature Set and B-Feature Set are slightly different from these Sets which are selected from raw training data set. We selected 42 features in the combined feature set.

Then the nine models are retrained. The result is shown in the Table 2.

TABLE II
THE RESULT OF THE CLASSIFICATION BASED ON NEW FEATURE SET BY DIFFERENT CLASSIFIERS

		Raw	B-Feature	R-Feature
SVM	Accuracy (%)	0.791	0.872	0.839
	FPR (%)	0.244	0.147	0.209
	FPN (%)	0.092	0.105	0.067
Random Forest	Accuracy (%)	0.729	0.892	0.832
	FPR (%)	0.267	0.202	0.164
	FPN (%)	0.197	0.093	0.103
Boosting	Accuracy (%)	0.802	0.971	0.902
	FPR (%)	0.103	0.065	0.074
	FPN (%)	0.141	0.031	0.065

From the Table 2, the FPR is relatively low now. And still the boosting classifier performs better than the other classifiers. And the classifiers on the B-Feature Set and the R-Feature Set performs better than on the raw feature set. After training, we find the models of boosting classifiers are much smaller than the SVM and the random forest.

The boosting classifier is better than SVM and Random Forest, which is only because of the boosting powerful on imbalance data set.

D Road Detection

The boosting algorithm performs better on both the feature selection and classification result, and the training model is smaller than SVM's and random forest's models. So we select boosting algorithm for road detection.

Figure 2 shows the result of it based on the Boosting classifier using model on the B-Feature Set. The blue region is the regions classified into the road.

Fig 2 (a). The cloudy suburban road

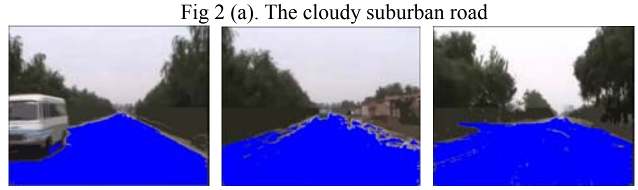

Fig 2 (b). The detection result of cloudy suburban road

Fig 2 (c). The cloudy campus road

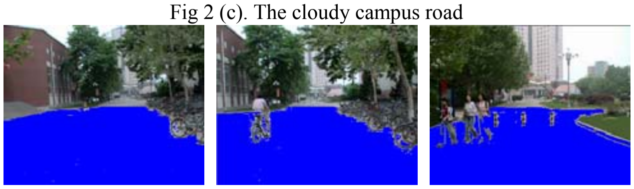

Fig 2 (d). The detection result of cloudy campus road

Fig 2 (e). The in shadow campus road

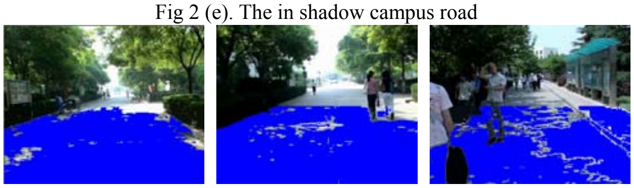

Fig 2 (f). The detection result of the in shadow campus road

Fig 2 (g). The sunny campus road

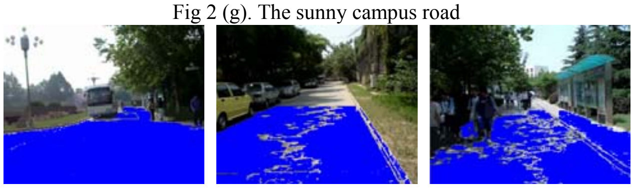

Fig 2 (h). The detection result of sunny campus road

Fig. 2 The road image and the result of the road detection

The detection result of cloudy road is good, because of the homogeneous nature of the road surface. There are miss classified regions on the shadowed and sunny road, especially for the shadow region border on the sunny region. This is because these regions are very small, and the size feature and its combined features are in the B-Feature Set. Furthermore, in the recomposed training set, the manually selected regions size is not small. Therefore, the error rate of the small regions is relatively high.

IV Conclusion

For the classification based road detection algorithm, the features are one of most important ingredients for pattern recognition. To exploit the correlations of the features, four feature combination functions are used in combination. Then the combined feature is selected by feature selection algorithm based on boosting. Then, the boosting classifier is used fore road detection. The result shows that either for imbalance or balance training data set, the boosting classifier on B-Feature set performances provide the best result.

However, for the complex environment or heavy shadow, the error rate is still high. In future work, the adaptive road segmented algorithm, new combination function and other feature selection algorithm will be tried in order to achieve better performance. As work continues, we hope to test this method on a very large set of images, to see if these promising results can deliver a more successful universal system.

References

[1] M. Bertozzi, A. Broggi and A. Fascioli, “Vision-based Intelligent Vehicles: State of the Art and Perspectives,” in Robotics and Autonomous Systems 32, 2000.

[2] Mike Foedisch, Aya Takeuchi, “Adaptive Real-Time Road Detection Using Neural Networks”. Proceedings of the 7th International IEEE Conference on Intelligent Transportation Systems, Washington, DC, October 3-6, 2004.

[3] M. Foedisch, C. Schlenoff, M. Shneier “Towards an Approach for Knowledge-based Road Detection”. (2005)

[4] Ming-Yang Chern and Shi-Chong Cheng. “Finding Road Boundaries from the Unstructured Rural Road Scene”. 16th IPPR Conference on Computer Vision, Graphics and Image Processing(CVGIP 2003).

[5] P. Viola and M.J. Jones, “Robust real-time object detection”, ICCV Workshop on Statistical and Computation Theories of Vision, 2001.

[6] R. Xiao, L. Zhang, and H. J. Zhang. Feature selection on combination for efficient learning from images. http://research.microsoft.com/asia/dload_files/group/vc/2004/accv_rxiao.pdf

[7] R. Bekkerman, R. El-Yaniv, N. Tishby, and Y. Winter. “Distributional word clusters vs. words for text categorization.”JMLR, 3:1183-1208, 2003.

[8] J. Weston, A. Elisseff, B. Schoelkopf, and M. Tipping. “Use of the zero norm with linear models and kernel methods.” JMLR, 3:1439-1461, 2003.

[9] Zhang G. Ying, Sha Y. “Character choosing Based on the Boosting in Pattern Recognition. ” Journal of Beijing Institute of Technology, 47(10): 234-239, June 2004.

[10] Leo Breiman. Random forests. Machine Learning, 45(1):5-32, 2001.http:\\citeseer.nj.nec.com/breiman01random.html.

[11] I.Guyon, J. Weston, S. Barnhill, and V.Vapnik. Gene selection for cancer classification using support vector machines. BIOWulf Technical Report,2000.

[12] E. Amaldi and V.Kann. On the approximation of minimizing non zero variables or unsatisfied relations in linear systems. Theoretical Computer Science, 209:237-260, 1998.

[13] V. Vapnik. Statistical Learning Theory. New York: John Wiley & Sons, 1998

[14] R. E. Schapire, Y. Singer. Improved boosting algorithms using confidence-rated predictions. Machine Learning, 37(3):297-336, December 1999.

[15] Chih-Chung Chang, Chih-Jen Lin. “LIBSVM—A Library for Support Vector Machines”. http://www.csie.ntu.edu.tw/~cjlin/libsvm/