Fast Classification of Road, Grass, and Trees from Color Outdoor Scenes

Hiroshi WAJIMA* Toru YAMASHITA Takahiro SUGIYAMA Keiichi ABE*

Gifu Technical Institute, Kawasaki Heavy Industries, Ltd.
Faculty of Information, Shizuoka University
Graduate School of Science and Engineering, Shizuoka University

Kawasaki Heavy Industries, Ltd. 1, Kawasaki-Cho, Kakamigahara City, Gifu Pref. 504-0971 Japan
E-mail : wajima_h@khi.co.jp

Abstract

In this paper we propose a method of realizing automatic recognition of vegetation in outdoor scenes rapidly and at low cost, on such a level that a human does at a glance. We intend to utilize this function as a component of guiding of an autonomous vehicle.

In our method a modified hue component is used as the only feature for classification. The input image is divided into subimages, called cells, and the average of square of the modified hue in each cell is calculated. By comparing the averages with two fixed thresholds, the cells are classified into area of soil, grass, or trees.

We examine ten images to get correct rates of classification using this method. The average and the standard deviation of correct rates are about 88% and 4.8%, respectively. Moreover, the method is tested on about two minutes of videotaped data of real out-door scenes. It takes 0.5 second to process a frame (of 640x480 pixels) with an ordinary personal com-puter. The method attains the results roughly equal to human recognition at a glance.

We also apply this method to guidance of an experimental autonomous vehicle. We confirm that it can run both on dirt roads and along painted lines on paved roads.

1 Introduction

This study proposes a method which adopts a simple algorithm using only one feature for fast and rough interpretation of outdoor scenes.

Some studies have been done on color image segmentation of outdoor scenes for various purposes, such as indexing of image databases, guidance for autonomous vehicles, and so on. For example, there are studies[1][2][3] aiming at automatic producing of indexes for image databases. The indexes stand for classification of objects (e.g. cars, roads, trees, etc.) included in an image. Most of these studies focus on interpreting information accurately from a given color image as much as possible. Therefore, they need long processing time and many kinds of features or knowledge, and sometimes need human assistance during the process.

As another field of research, guidance of autonomous vehicles has been studied. For this purpose, the process must be done automatically and rapidly in order to control the vehicle at sufficiently fast speed. In this kind of applications, most of the studies (e.g. BVV2[4] in VaMoRs:Germany, YARF[5] in NavLab:U.S.A., using wide dynamic range vision sensor[6]:Japan, FIVIS/VIP in PVS[7]:Japan, etc.) depend on detecting clean markings in an artificial structure such as a highway lane. There are many systems which have shown good results if they are used under such an environment. Meanwhile, there are only a few systems which can be applied to unpaved roads without lane markings in rural situations, such as ALVINN[8] and SCARF[9]. A famous example is ALVINN system, which has been developed in Carnegie Mellon University. This system is applied to autonomous vehicle series "Navlab", which can run on either dirt roads or paved roads. In this case, however, the output cannot be interpreted as "region classification" understandable for human beings, because this system is composed of an artificial neural network with RGB-data as input and steering commands as output. Another system called SCARF has been also developed in CMU. This system realizes "region classification" by using road and off-road models which are given manually beforehand. It can detect the boundaries and direction of the region of road by Hough transform from the results of region classification. This system needs long processing time because of using Hough transform, however. Furthermore, the wider variety of• road scenes it copes with, the longer time it requires.

In this paper, we propose a method of realizing automatic recognition of vegetation in outdoor scenes rapidly and at low cost, on such a level that a human does at a glance. We intend this method for practical use as an autonomous vehicle, and its algorithm is given in Chapter 2. We show the effectiveness of our method through experiments using real outdoor scene images, real outdoor videotaped images, and an experimental autonomous vehicle.

2 Process of the method

In our method the color of each pixel is transformed into a modified hue component first, and the input image is divided into subimages, called cells here. The average of square of the modified hue components in each cell is calculated as the feature for classifying the cell. Our system can classify cells into area of soil, grass, or trees.

Our method has the following strong points compared to the other approaches.

- Effective classification results are obtained not only in the cases that clear boundaries (e.g., lines on paved roads, boundary between the sky and mountains in a distant view, etc.) are found, but also in the cases of images without distinct boundaries (e.g., dirt roads, vague boundaries between grass and soil in a closer view, etc.) without other supplementary knowledge like geometric road models.
- Each target class of classification can have wide variety of colors and textures.
- Images can be classified regardless of the distance to objects.
- The process can be done fully automatically.
- Processing time is short enough to control a vehicle in real-time, while the system can be implemented at a low cost because it does not need any expensive image processors.

On the other hand, the output of classification is limited to the member of the specified set of classes which is composed of soil, grass, trees, and unknown (near achromatic-colored cell).

2.1 Transformation of Color Space

The key point of this method is to adopt the hue component as the unique feature. In the consequence this method is robust against change of brightness in outdoor scenes. The hue component can be calculated by following equation. [10]

$$\bullet\ hue = \cos^{-1}\left[\frac{1/2\{(R-G)+(R-B)\}}{\{(R-G)^2+(R-B)(G-B)\}^{1/2}}\right]$$

$$(R \neq G, G \neq B,\ \text{ if } B > G \text{ then } hue = 2\pi - hue)$$

We find that it is difficult to classify the vegetation (soil, grass and trees) well by this feature, if hues in an outdoor image are terribly messy. So we modified this original hue value as shown in Fig.1. The rationale for this modification is as follows.

- The origin of the modified hue is put on the midpoint between red and blue, so that the hue of soil around red does not become discontinuous.
- The sensitivity of modified hue from red to green are doubled compared to one of the rest of the hue range, so that hues of grass and trees might be well separated.

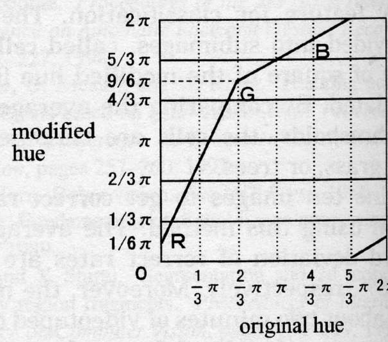

Fig. 1 Modification of hue

An indefinite hue pixel is the one such that R=G=B (e.g., achromatic-colored objects, saturated pixels due to highlights, etc.). The number of indefinite hue pixels is counted in each cell. If the number exceeds a half of the number of pixels in the cell (32x32=1024), then the cell is regarded as achromatic-colored and classified to 'unknown'.

2.2 Selection of the feature

In this study we recognize two major problems in selecting the features to be used. One is how to handle various scales in a image (e.g., changes of resolution in a image between near objects and far ones: lines, boundaries and textures). The other is that real soil, grass, and trees include a lot of colors and sometimes the same colors.

Firstly, the images at outdoor scenes are usually taken in a wide angle with very shallow angle of dip. So such an image has various depth in its different positions and consequently various real-space length per pixel, and outdoor images often do not have clear boundaries between different objects in the images. Furthermore, as the ordinary CCD cameras do not have enough resolution to catch the details of the textures, the ordinary texture analysis cannot be applied adequately. In order to solve these problems, we adopte the average of the square of modified hue as the only feature for extracting soil, grass, and trees portions from ordinary outdoor scene images, regardless of their depth and resolution. This feature is calculated for each square cell in the modified hue image.

Secondly, real outdoor images include a lot of colors. For example, soil area is said to be "brown" in general. But in fact, it consists of many colored pixels which are not only "brown", but also "green", "blue", "red" and a lot of other intermediate colors. In our method, the square of modified hue is selected from many possible features to cope with these problems. The feature is invariable against differences of scales, and it can express the main component color in each cell.

2.3 Setting the threshold

In general, one of difficult problems to realize classification is how to determine a threshold appropriately for an given image. In our method, we use fixed thresholds for all images by adopting the hue-based feature, because hue of an object can be considered as an invariant regardless of the photographic condition if the image is taken under an enough adjustment of white balance.

We use ten sample images to determine the thresholds. Some of them include only soil, grass, or trees, and the others are mixtures of two classes among the three classification objects. The sample images are also prepared in consideration of the following variations.

soil: soil • • gravel
grass: short grass • • tall grass
trees: short distance • • long distance

Each one of these images is divided into square cells of 32x32 pixels. The model result of each square cell is determined by a human observer. In the case that it is difficult to decide the class, multiple choices (e.g., both soil and grass) are allowed. Then, a histogram is made from the model results and the calculated feature of each square cell, as shown in Fig.2.

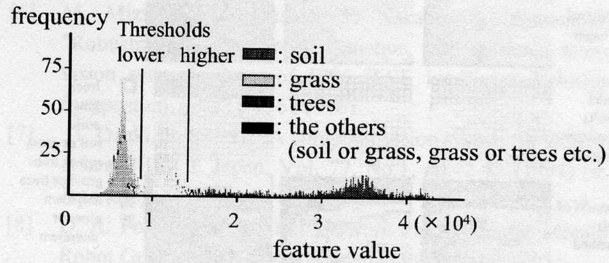

Fig. 2 Histogram of feature values from the sample images and the determined thresholds

From this histogram, the thresholds are determined manually as follows.

Lower threshold between soil and grass : 0.9×10^4

Higher threshold between grass and trees : 1.4×10^4

With these thresholds, the class of each cell of a target image is determined, and the output of this process is obtained as the last step. As mentioned above, this process is simple and uses only the average of the square of modified hue as the feature, so it can be executed in short time.

3 Results of Experiments

In order to confirm the availability of this method, it is applied to classification of outdoor scene images, and guidance of an autonomous vehicle.

3.1 Classification of outdoor scene images

Ten different sample images are analyzed with the fixed thresholds determined as in Chapter 2, and each square cell is classified into one of four classes: trees, grass, soil, or unknown. Some results by this method are shown in Fig.3.

The correct rate is calculated as follows.

• correct rate = (S / T) x 100 (%)

where

S : the number of the square cells which are classified into the same class as in the model results.

T : the total number of square cells in the original image.

The average and the standard deviation of correct rate are about 88% and 4.8%, respectively. The model result of each cell is determined by a human observer. At the case that it is difficult to decide the class, multiple classes (e.g., both soil and grass) are allowed. When the classification result is the same as either one, it is considered as correct.

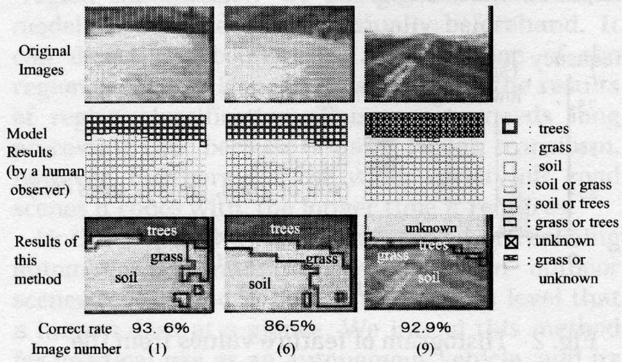

Fig. 3 Results of some experiments

3.2 Comparison of threshold selection

In the previous section, the fixed thresholds in our method are shown good enough for rough interpretation of outdoor scenes. In general, it is said that a method with fixed thresholds does not have sufficient robustness against changes of environment. But in this method, it shows two advantageous points that the fixed threshold is more suitable than a variable threshold through our experiment and considerations. The one point is on robustness. In our experiment, correct rates by a variable threshold vary more widely on various images than those by fixed one, although the average of them is almost the same. We discuss the details later. The other point is that the number of classes must be specified for each image in a variable threshold approach. For example, some images are composed of two classes (soil and grass) and need a single threshold, while others are composed of three classes (soil, grass, and trees) and need two thresholds.

From the viewpoint of robustness, it may be better that the thresholds are adjusted for each image separately. We tried to compare the fixed threshold and a typical method of deciding adequate threshold called Otsu method[11], which is based on discriminant analysis. In deciding the threshold, Otsu method must be given the number of classes. It is given by a human observer on each image in this experiment. On the other hand, the fixed threshold process determines the number of classes automatically. The results of this experiment are shown in Table 1.

Table 1 Statistics of correct rates for fixed threshold and Otsu method

correct rate	fixed threshold	Otsu method
average	88.0%	88.7%
standard deviation	4.8%	8.1%
difference between best and worst	15.8%	26.9%

According to this result, both methods show similar performances as for the average of correct rate, though Otsu method is given the number of classes by human observer. Moreover, the method of fixed thresholds is better than Otsu method concerning the standard deviation and the difference between the best and the worst. Otsu method is based on the discriminant analysis, which in turn assumes Normal distribution for each class. If the distributions of features in one or more classes are quite different from normal, then the resulting threshold would be inadequate. This explains why the correct rates by Otsu method has bigger standard deviation and range.

4 Guidance of an autonomous vehicle

In order to show the effectiveness of this method, it was also implemented on an experimental autonomous vehicle system shown in Fig.4.

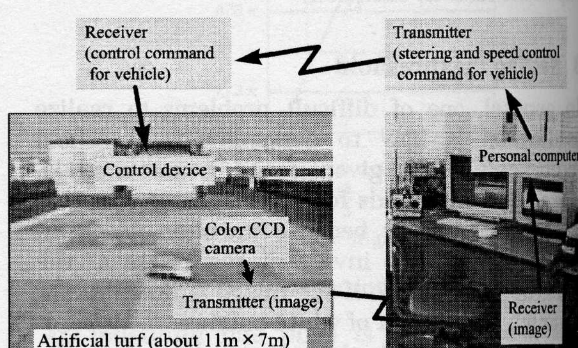

Fig. 4 Experimental autonomous vehicle system

It can run along the mock dirt road (a lap is about 22m long) on artificial turf at the average speed of 0.16m/s. It takes 0.5 second to process each image (640x480 pixels) by a personal computer with Pentium200MHz processor. It is very easy to improve the processing time, by using smaller image size and/or changing to a faster CPU. Even if image size becomes smaller (deterioration in the resolution) to reduce processing time, it does not make the correctness worse so much, because the process is based on invariable feature against differences of scales and does not use edges.

Moreover, the system was tested on about two minutes of videotaped data taken on a real dirt road. The method obtained the results roughly equal to human recognition at a glance. We suppose that the vision system can be used for an real autonomous vehicle in real outdoor scenes.

5 Conclusion and discussions

In this paper we propose a method of realizing automatic recognition of vegetation in outdoor scenes rapidly and at low cost. We show the effectiveness of our method through experiments on real outdoor videotaped images and test course. In our method, it can classify among the area of soil, grass, and trees, which are the main vegetation in natural environments, and unknown. We examine ten images to estimate correct rates using this method. The average and the standard deviation of correct rates are about 88% and 4.8%. Moreover, the method is tested on about two minutes of videotaped data at a real outdoor scene. It takes 0.5 second to process a frame (640x480 pixels) with an ordinary personal computer. The method obtains the results roughly equal to human recognition at a glance.

This feature has been obtained through considerations on the property of human vision and the characteristics of outdoor scenes. On the other hand, our method can be applied only to colored objects. So non-colored objects such as asphalt, concrete and cloud can not be handled. But brightness or saturation could be used as the feature instead of hue for these classes. We will integrate brightness and saturation in our method in order to extend the ability of extraction of vegetation classes more correctly. Furthermore, we are planning to realize an autonomous vehicle which goes around in a factory yard as a guard.

6 References

[1] W. P. J. Mackeown, "A labeled image database and its application to outdoor scene analysis", Doctor thesis, the University of Bristol (1994)

[2] S. Hirata, Y. Shirai, M. Asada, "Scene interpretation using 3-D information extracted from monocular color images", Trans. IEICE Japan, Vol. J75-D-II, No.11, pp.1839-1847 (1992) (in Japanese)

[3] D. K. Panjwani, G. Healey, "Markov random field models for unsupervised segmentation of texture color image", IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 17, No.10, pp.939-954 (1995)

[4] U. Solder, V. Graefe, "Object detection in real time", SPIE, Vol.1388 Mobile Robots V, pp.104-111 (1990)

[5] D. Aubert, "Autonomous navigation of structured city roads", SPIE, Vol.1388 Mobile Robots V, pp.141-151 (1990)

[6] M. Mizuno, K. Yamada, T. Nakano, S. Yamamoto, "Robustness of lane mark detection with dynamic range vision sensor", Proc. of 1996 IEEE Intelligent Vehicles Symposium, pp.171-176 (1996)

[7] T. Ozaki, S. Sasaki, "A computer vision system for vehicle system", IEICE Japan, Vol. 77, No.1, pp.42-45 (1994) (in Japanese)

[8] D. A. Pomerleau, "Neural Network Perception for Mobile Robot Guidance", Kluwer Academic Publishers (1993)

[9] J. D. Crisman, C. E. Thorpe, "Detecting difficult roads and intersections without map knowledge for robot vehicle navigation", SPIE, Vol.1388 Mobile Robots V, pp.152-164 (1990)

[10] M. Yachida, "Robot Vision", Shokodo Co., Ltd. (1990) (in Japanese)

[11] N.Otsu, "A threshold selection method from gray-level histogram", IEEE Trans. on Systems, Man, and Cybernetics, Vol. SMC-9, No.1, pp.62-66 (1979)