

## Color Recognition Based on C4.5 Decision Tree Algorithm

Chao Chen, He-Gen Xu

College of Electronics and Information Engineering

Tongji University

Shanghai, China

E-mail: 1433345@tongji.edu.cn, xuhegen@tongji.edu.cn

**Abstract-**This paper presents an approach to color recognition using C4.5 decision tree algorithm. Before being employed to identify the color, the C4.5 decision tree is trained with sample images under scenes of varied background and illumination. The experimental results show the effectiveness and robustness of the method for color segmentation under the changing environments.

**Keywords-**C4.5 algorithm; color recognition; HSV

### I. INTRODUCTION

Color recognition plays an important role in machine vision. At present, there are various kinds of methods for color recognition. Shi Peiyuan et al. [1] introduced a fuzzy-based color recognition. Yang Xuqiang et al. [2] suggested a color recognition method based on HSI color space, and Zhang Guoquan et al. [3] proposed a similar method, which based on HSV color space. Wang Yunqiong et al. [4] used the support vector machine to identify color. Tang Shengxue et al. [5] analyzed the characteristics of fuzzy neural network and proposed a color recognition base this algorithm. Jiang Yuling et al. [6] used the SOM algorithm to identify color. Pan Hao et al. [7] proposed a color recognition method based on illumination changed compensation. Zhu Mingfeng et al. [8] applied the theory of BP-ANN to identify color. Hu Chuaping et al. [9] introduced a color recognition with spatial pyramid deep learning. Zhao Chao et al. [10], [11] suggested an approach for color recognition based on SVM-Boosting Classifier.

The method of fuzzy-based color recognition has a good effect for image retrieving. Fuzzy-based color recognition defines a set of fuzzy sets of colors firstly, then determines which color every pixel belongs to according to the values in the channels of HSV color space. This method can't identify the color for each pixel accurately, and only an approximating result can be achieved. Especially when used in a complex background, the result of this method is not satisfactory. So it may not suitable for the application which has a complex background or requires an accurate identified result.

As for utilizing HSV space or HIS space for color recognition, the color threshold method is used. The threshold of the specified color could be determined by heuristics. Those methods of color recognizing images determined one pixel whether belongs to a specified color by comparing the value of each channel of the color space. For an image in the same background, the value of each channel of color space will change with the environmental

illumination. Then the result of color recognition would be inaccuracy. So those methods are not suitable for the environment in which light changes frequently.

This paper proposes a color recognition method employing C4.5 decision tree to extract specified color from images under different illumination conditions and backgrounds.

### II. THEORY AND IMPLEMENTATION

#### A. Related Concepts

Decision tree is a tree in which each branch node represents a choice between a number of alternatives, and each leaf node represents a decision [12]. C4.5 rely on the information gain rate to choose the attribute [13], [14], [15], [16], which remedies the deficiency of ID3, for the ID3 uses the information gain to choose the attribute [17], [18].

The assumed training set contains  $m$  categories [15], respectively  $S = \{S_1, S_2, \dots, S_m\}$ . The partition according to  $A$  is  $S' = \{S'_1, S'_2, \dots, S'_v\}$ . Other attribute are similar as  $A$ . Based on the training set, the information entropy of ideal partition is as (1). In equation (1),

$$p(S_i) = \frac{|S_i|}{\sum_{i=1}^m |S_i|}. \text{ Information entropy of partition}$$

according to attribute  $A$  is showed as (2). In equation (2),

$$p(S'_i) = \frac{|S'_i|}{\sum_{i=1}^v |S'_i|}. \text{ Equation (3) is the information gain of}$$

attribution  $A$  to training set and the (4) is the partition entropy of  $A$ . By (3), (4), Information gain ratio of attribute  $A$  is as (5). Similarly we can calculate information gain ratio of other attributes [15].

$$Info(S) = -\sum_{i=1}^m p(s_i) \log_2 p(s_i) \quad (1)$$

$$Info_A(S') = \sum_{i=1}^v p(s'_i) Info(s'_i) \quad (2)$$

$$Gain_A(S') = Info(S) - Info_A(S') \quad (3)$$

$$SplitInfo_A(S') = -\sum_{i=1}^v p(s'_i) \log_2 p(s'_i) \quad (4)$$

$$GainRatio_A(S') = \frac{Gain_A(S')}{SplitInfo_A(S')} \quad (5)$$

### B. Implementation

This method has two steps: first, the training samples are used for learning to create a classifier. Then identifies the specified color in an image based on the classifier. Considering that the components of RGB color space, the  $R$ ,  $G$ ,  $B$  are chosen as the attribute.

In order to construct the decision tree, the training samples are necessary. And to collect the training samples, a color training template is needed. This paper uses a blue training template as shown in Fig. 1 and tries to identify the color of blue as an example. Note that the new method we propose is applicable to all colors in theory.

The color training template is placed in the environment with different light intensities and backgrounds. It means that we have two kinds of image samples: (1) the samples with the same light intensity but different background; (2) the samples with the same background but different light intensity.

For training C4.5 decision tree with the sample images, the images must have different markers to distinguish different colors in the template. In this paper we mark blue pixels as 1 and non-blue pixel as 0 to distinguish the blue pixels and other pixels manually.

Additionally, the sample data should be discretized to reduce computational complexity before being used to train the decision. What's more, pruning method is used before and after the training process of the tree to improve the accuracy and simplify the structure of C4.5 decision tree. The pre-pruning is mainly used for avoiding the following occasion: some pixels have the same value of RGB but different markers [19], [20]. In this case, the number of pixels with different marks should be recorded. Then the kind of pixels which is less in number will be regarded as the noise and should be removed. Post-pruning means removing the noise and isolated points due to abnormal branching after the tree finished [21].

The  $R$ ,  $G$ ,  $B$  values of the image pixels and the marker of each pixel (0 or 1) are used to train the decision tree. The steps of training decision tree are showed as follows:

- 1) Calculate the information gain ratio of each attribute ( $R$ ,  $G$  and  $B$ ).

2) Rank the information gain rate, select the attribute that has the maximum information gain ratio as the root node of the tree.

3) Classify the training samples with the selected attribute.

4) Calculate the information gain rate of the remaining attributes.

5) Rank the information gain rate again, and select the attribute that has maximum information gain ratio as the lower node.

6) Repeat (c) to (f), until meeting the stop condition or indivisible.

7) Achieve the decision tree.

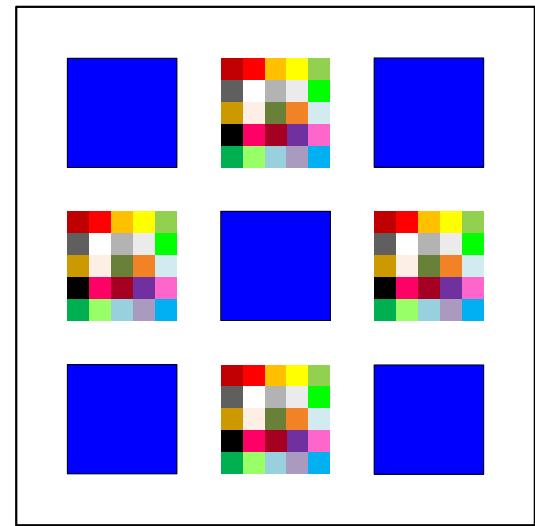


Figure 1. Color training template

The sample images are collected from five scenes, and from 15 to 50 pictures are chosen in each scenes. In each scene the pictures may have different light intensity. Note that the pictures from the same scene may have different light intensity. Randomly select several pictures from each scene for verifying and then 80 to 200 images are selected randomly from the rest images as training samples. Finally, we will get a C4.5 decision tree sketch as Fig. 2. Whether a pixel is blue or not is according to the value of each channel of the RGB color space. The boxes represent the judgment of the property values and the circles represent the classification result. The value of 'YES' represents the pixel is blue and 'NO' represents other colors.

### C. Experimental Results and Analysis.

In this paper, the method of the color recognition based on the HSV color space is used as the comparison group. When using HSV color space to identify color, threshold in each channel should be selected. The threshold range should be selected appropriately to make it be applicable to many scenarios.

In this experiment, cut-and-try method is used to select the threshold. At last, the range of threshold is selected as (6)

$$\begin{cases} 158 < H < 180 \\ S > 100 \end{cases} \quad (6)$$

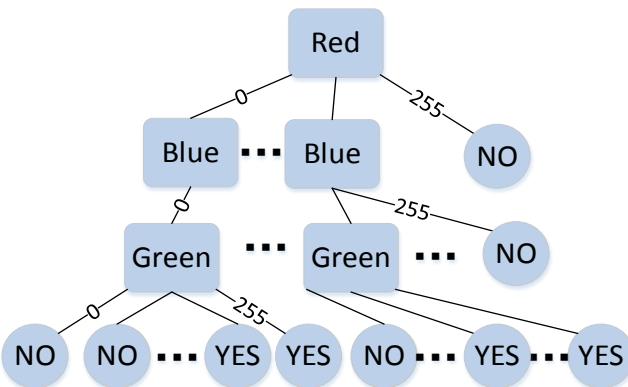


Figure 2. The sketches of C4.5 decision tree.

Figure 3 shows the effects of color recognition with HSV color space, and exhibit a good performance of color recognition in some certain scenes.

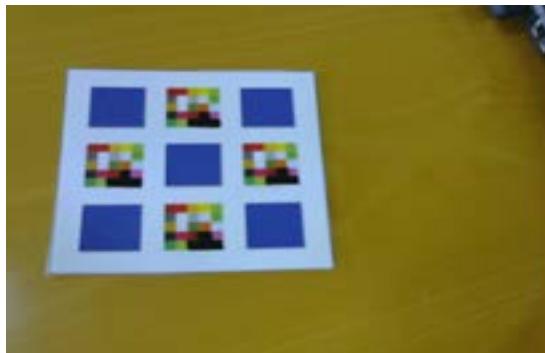
As a comparison, two groups of tests are carried out to show the effectiveness of color recognition base on C4.5 decision tree.

1) *Test 1:* In this experiment, images with the same background but different light intensity are selected for testing.

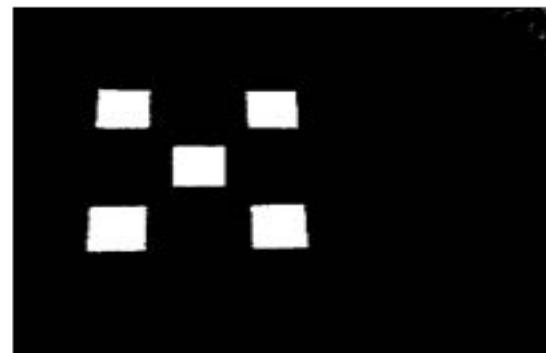
In Table I, The light intensity of the three source images increases one by one. The test shows that when the intensity of light changes, the result of color recognition base HSV color space does not work well enough, but the method based on C4.5 decision tree has a relative good result.

2) *Test 2:* This experiment is used to verify the performance of color recognition for images with the same light intensity but different background.

In Table II, the results of four samples show that the effect of color recognition based on C4.5 decision tree is better than HSV space.



(a) Sample picture



(b) The result of color recognition

Figure 3. The effect of color recognition with HSV color space.



(a) Source image



(b) The result of color recognition

Figure 4. Recognition effect for non-template image.

TABLE I. THE RESULT OF TEST 1

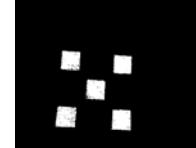
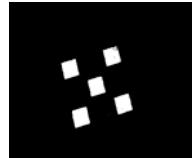
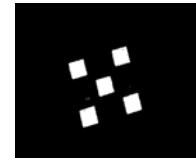
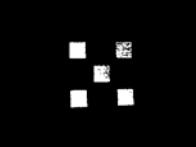
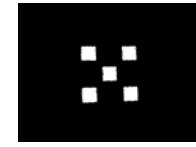
	source image	the result of HSV	the result of C4.5
sample 1: dark light			
sample 2: mild light			
sample 3: strong light			

TABLE II. THE RESULT OF TEST 2

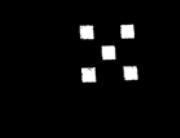
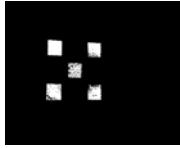
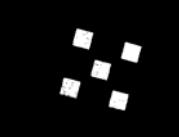
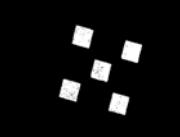
	source image	the result of HSV	the result of C4.5
sample 4: yellow background			
sample 5: dark green background			
sample 6: white background			
sample 7: red background			

TABLE III. THE RECOGNITION RATE OF HSV AND C4.5 IS 150 WHEN THE TRAINING SAMPLE NUMBER IS 150

	p	p1	r1	p2	r2
sample 1	19532	8255	42.3%	18006	92.3%
sample 2	15054	14283	94.8%	14186	94.2%
sample 3	13258	10646	80.3%	11587	87.4%
sample 4	19529	18747	96.0%	18220	93.3%
sample 5	13215	1388	10.5%	11960	90.5%
sample 6	14023	6583	49.8%	11961	85.3%
sample 7	18892	16984	89.9%	17040	90.2%
average	-	-	66.2%	-	90.5%

p: the total blue pixel in the picture, estimated with area.

p1: the number of pixels identified by HSV

r1: the recognition rate of HSV

p2: the number of pixels identified by C4.5

r2: the recognition rate of C4.5

TABLE IV. RECOGNITION RATE OF DIFFERENT TRAINING SAMPLES

Sample number	80	100	130	150	170	190
Recognition rate(%)	80.3	83.4	89.9	91.2	90.1	88.3

Table III shows the comparison of color recognition effect between C4.5 decision tree and HSV color space. The results show that the method of color recognition based on C4.5 decision tree can effectively increase the robustness and accuracy.

Recognition accuracy of C4.5 decision tree is also affected by the number of the training samples, as shown in Table IV. The peak value arrives at the sample number of about 150. The reason for descending of the performance is over-fitting of the tree. Conclusions can be drawn that the recognition rate is not proportional to the number of training samples, appropriate number of training samples shall be chosen to avoid over-fitting.

### III. SUMMARY

This paper proposes a new method for color recognition. First, train the C4.5 decision tree with numbers of template images under scenes with varied background and illumination. Then use the C4.5 decision tree from the first step to identify the color in the images. From the experiments,

it can be concluded that the effect of color recognition has a good robustness and high accuracy to a changing background and illumination. The method of color recognition in this paper can be extremely good for image segmentation in image processing.

The disadvantages of this method include:

- The color recognition based on C4.5 decision tree has a good effect for those scenes which are included in the training samples, but for those are not included, the performance may not be good enough.
- A recognition accuracy of 100% cannot be assured for this method.
- The pre-pruning method before training tree may bring some side-effects. The pruning method is used to remove the noisy, but it also causes some useful information to be eliminated.
- More attributes should be selected. In this paper, only three attributes of R, G and B are used to construct the decision tree, without the use of image's brightness or the light intensity of the

- environment.
- Appropriated number of training samples is difficult to be pre-defined to avoid over-fitting the decision tree.

## REFERENCES

- [1] Shi Peiyuan, Deng Tingquan. "Fuzzy-based color recognition and its application in image retrieval" Computer Engineering and Applications, 2013, 49(18):138-141.
- [2] Yang Xuqiang, Feng Yong, Liu Hongchen. "An object extraction method based on HSI color model" optical technology, 2006, 32 (2): 290-292.
- [3] Zhang Guoquan, Li Zhanming, Li Xiangwei, Wei Weiyi. "Research on color image segmentation in HSV space". Computer Engineering and Application, 2010, 46 (26): 179-181.
- [4] Wang Yunqiong, You Zhisheng, Liu Zhifang. "Vehicle Color Recognition Using Support Vector Machine", Journal of Computer-aided Design & Computer Graphics, 2004, 16(5):701-706.
- [5] Tang Shengxue, Liu Bofeng, Xu Dongfeng. "Color recognition method based on fuzzy neural network", Journal of Transducer Technology, 2003,22(11):57-61.
- [6] Jiang Yuling, Yang Yimin. "Color Recognition of Machine Vision Based on SOM Algorithm", Journal of Guangdong University of Technology, 2011,28(2):40-42.
- [7] Pan Hao, Yang Yimin. "Color Recognition Method Based on Illumination Changed Compensation", Computer and Modernization, 2014,4:201-205.
- [8] Zhu Mingfeng, Du Jianqiang, Zhang Kang, He Yangming, Ding Chenghua. "Research of Tongue Color Recognition Based on BP-ANN with Self-adaptive Network Structure", 2015 7th International Conference on Information Technology in Medicine and Education, 2015:208-211.
- [9] Hu Chuiping, Bai Xiang, Qi Li, Chen Pan, Xue Gengjian, Mei Lin. "Vehicle Color Recognition With Spatial Pyramid Deep Learning", IEEE Transactions on Intelligent Transportation Systems, 2015, 16(5):2925-2934.
- [10] Zhang Chao, Zong Guanghua. "An Approach of Color Recognition Based on SVM-Boosting Classifier", Computer and Modernization, 2013, 9:98-102.
- [11] Wang Anmin, Yang Qi. "A Combination of Color Recognition System of SVM", Machinery & Electronics, 2012,4:15-17.
- [12] John Durkin, Cai Jingfeng, Cai Zixing. "Decision Tree Technique and its Current Research", Control Engineering of China, 2005,12(1):15-19.
- [13] Liu Yaonan. "Analysis and application of C4.5 algorithm". Journal of Dongguan University of Technology, 2012, 19 (5): 47-52.
- [14] Li Xiaowei, Chen Fucai, Li Shaomei. "An improved algorithm of C4.5 decision tree based on classification rules". computer engineering and design, 2013, 34 (12): 4321-4326.
- [15] Miao Yufei, Zhang Xiao Hong. "Improvement and application of C4.5 decision tree algorithm". Computer Engineering and Applications, 2015, 51 (13): 255-259.
- [16] Kusrini, Hartati S, Wardoyo R, Harjoko A. "Efficient C4.5", IEEE Transactions on Knowledge and Data Engineering. 2002, 14(2): 438-444.
- [17] Huang Wen. "The classical decision tree: ID3 and C4.5" Journal of Sichuan University of Arts and Science (NATURAL SCIENCE), 2007, 17 (5): 16-18.
- [18] Thakur D, Markandaiah N, Raj D S. "Re optimization of ID3 and C4.5 decision tree", 2010 International Conference on Computer and Communication Technology (ICCCT 2010), 2010, 448-450.
- [19] Wei Hongning. "Comparison among Methods of Decision Tree Pruning" Journal of Southwest Jiao Tong University 2005, 40 (1):44-48.
- [20] Hae Sook Jeon, Won Don Lee. "Pruning with Majority and Minority Properties", 2014 International Conference on Information Science & Applications (ICISA), 2014.
- [21] Zhang Wenchao, Li Yafen. "A Post-Pruning Decision Tree Algorithm Based on Bayesian", 2013 International Conference on Computational and Information Sciences, 2013, 988-991.