MVA2002 IAPR Workshop on Machine Vision Applications, Dec. 11 - 13, 2002, Nara - ken New Public Hall, Nara, Japan

13—26

Region-based Image Interpretation and Recall

Hui-Yu Huang[1]
Department of Information Management,
Taichung Healthcare and Management University

Yung-Sheng Chen[2]
Department of Electrical Engineering,
Yuan Ze University

Wen-Hsing Hsu[3,4]
Department of Electrical Engineering,
National Tsing Hua University,
Institute of Information Science, Academia Sinica

Abstract

In this paper, we present an approach to primarily achieve the semantic interpretation and the region retrieval for an attentive region in a color image. The main components of the system include image feature extraction, indexing process, as well as linguistic inference rules construction and semantic description. Based on these features, each of attentive regions in an image can be described by a global linguistic meaning. The main procedure consists of two parts: forward and recall processes. The forward process primarily performs the linguistic meaning description of objects for an image, and the recall process reconstructs the region image which is the rough mental image of human memory retrieval. Experiments confirm that our approach is reasonable and feasible.

Keywords: image call, semantic description

1. Introduction

Outdoor image understanding may be composed of the relevant researches involving computer vision and artificial intelligence. Recently, many researchers have considered the viewpoint of human vision in image processing schemes.

When human beings look at an image, some interesting regions are observed at a glance and reflected in our eyes. After performing more complicated physiology operations via human vision cortex, the mechanism can quickly to decide the meaningful objects and further to interpret the linguistic meanings for an image. Hence, based on region-based image interpretation, many methods have been proposed [1, 2]. As for outdoor image understanding, its purpose is usually focused on recognizing and localizing the significant image objects in the scene and that of distinguishing the relative object relationships. This is being still a very important topic.

In addition, according to the fundamental characteristics of fuzzy set theory, the cognitive information of an image can be translated into a fuzzy number (or a membership function) to construct a knowledge base such as human knowledge and experiences [1, 2]. A useful data information may be presented by the fuzzy number, and the data presentation of a fuzzy number can be parameterized to simplify the fuzzy computation. In accordance with the advantages of indexing process and fuzzy set theory, in this paper, we present an approach to perform a human-based image interpretation. This approach combines image features and linguistic database to describe semantic meaning for image region.

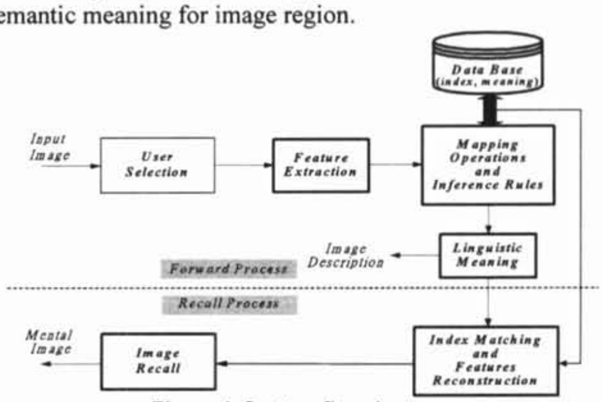


Figure 1. System flowchart.

2. Proposed Method

The major concept in color image interpretation is to describe the semantic meaning of a region based on some features, such as colors, texture, and spatial information. The flowchart of our system is depicted in Fig. 1.

2.1. Region Feature Extraction

Color, texture, and spatial information are three principal features adopted in our method. The color feature contains three components: hue (H), value (V), and chroma (C). Directionality and contrast are used as the texture features. Bounding rectangle of a region represents the spatial information. The corresponding schemes for feature extraction are described as follows.

1 Address: P.O.Box 8-236, Wufeng, Taichung, Taiwan 413, R.O.C. E-mail: **hhy@alumni.ee.nthu.edu.tw**
2 Address: No. 135 Yuan-Tung Rd., Chungli, Taoyuan, Taiwan 320, R.O.C. E-mail: **eeyschen@ee.yzu.edu.tw**
3 Address: No. 101 Sec. 2, Kuang Fu Rd., Hsinchu, Taiwan 300, R.O.C. E-mail: **whhsu@pc05.ee.nthu.edu.tw**
4 Address: No. 128 Academia Rd. Sec. 2, Nankang, Taipei, Taiwan 115, R.O.C.

2.1.1. Color features

Based on HVC color space [3, 4], each of three components in color space can serve as an individual feature and can further use it to compute and analyze the characteristic of each selected region.

First, we compute the color histogram of each selected region within an image to obtain the three color histograms. Let $R(i)$ be one region for an image denoted by a three-element color vector $(h(i),v(i),c(i))$, where i denotes the ith region of an image. Each element represents the corresponding color histogram. To efficiently specify color features, a 6-parameterized fuzzy number (named 6-PFN) is adopted for modeling the histogram.

After performing the 6-PFN process, each element within the three-element color vector is rewritten as the $(a_2(i), a_2(i), A(i), B(i), b_1(i), b_2(i))$ data format. If the image has n segmented regions, we can obtain n color features having 3×6 matrix size. One advantage of using 6-PFN fuzzy numbers is that it can keep the principal information and the properties of objects/regions in an image. In addition, based on the principles of fuzzy operation, the color features will appropriate to the fuzzy computation (e.g., max-min composition).

2.1.2. Texture features

A wide variety of texture features have been described in the machine vision literatures. We chose texture features based on the modified versions of directionality and contrast features proposed in [5]. Owing to the natural images, some objects possess the regular directionality feature, such as "tree" usually presents the feature of vertical orientation. In this paper the directionality characteristic is obtained from the edge's four orientations: 0^0, 45^0, 90^0, and 135^0 indicating the counterclockwise orientation along horizontal axis. Here the edge points are obtained by using the compass gradient operators

Generally, the variance σ^2 about the gray-levels histogram distribution is more preferable as contrast. To perform the polarization of the distribution of block and white on the gray-level histogram, the well known *kurtosis* α_4 is adopted and defined as [5]

$$\alpha_4 = \frac{\mu_4}{\sigma^4}, \tag{1}$$

where μ_4 is the fourth moment about the mean and σ^2 is the variance. This measure is normalized with respect to the range so that it can have the minimum value of one in case of twin peaks. Consequently, combining σ and α_4 for the second-step measure of contrast, it may be expressed as

$$F_{con} = \frac{\sigma}{(\alpha_4)^n}, \tag{2}$$

where n is a positive number. In our experiments, n is set to 1/4.

2.1.3. Spatial relationships

In addition, the spatial location of a region is also used in our system; it can provide the spatial information to construct the indexing process and spatial linguistic meaning definition, such as *left-up*, *left-bottom*, and so on. In order to facilitate spatial constraints, the approach organizes the location information of each region as part of its metadata. The spatial location of each region is represented by two parameters: the region centroid $c_{xy} = (x_c, y_c)$, and the coordinates of its minimum bounding rectangle (x_l, x_r, y_t, y_b), where subscript l, r, t, and b denote (*left, right, top, bottom*), respectively. The minimum bounding rectangle is the smallest vertically aligned box which contains the region completely as shown in Fig. 2(a). Hence, in our experiments, we noticed that the region is more intuitive and effective for catching the spatial information.

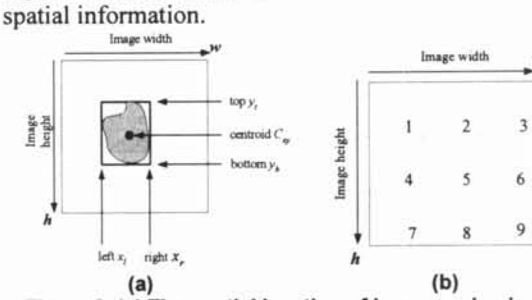
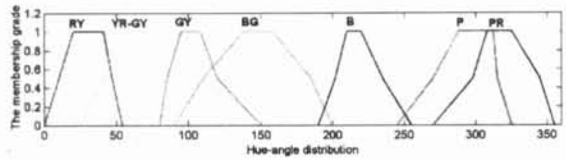


Figure 2. (a) The spatial location of image region is represented by its region centroid and its minimum bounding rectangle. (b) Indexing process of spatial location for an image.

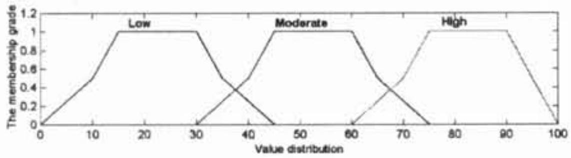
2.2. Indexing process

According to the defined features, number of features is divided into 5 categories containing 6 elements: *hue, value, chroma, edge property, contrast, and spatial position.*

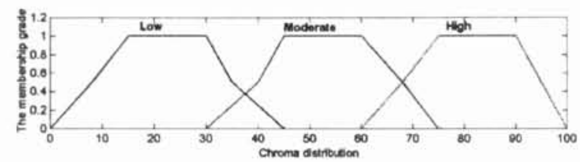
For color features, they have 3 color elements. The color attributes of each region of an image can be presented by the 6-PFN fuzzy number. First, we construct a set of reference 6-PFNs by using training way for all regions in image database. Based on the pre-clustered objects by human experiences, the same objects are further computed by those of color features. We take the average for the relative element of 6 parameters to obtain a set of reference 6-PFNs. Figure 3 shows the reference fuzzy number of h, v, and c. After performing the fuzzy max-min composition between each region and reference fuzzy numbers, the corresponding index value can be obtained. In our experiments, reference h, v, *and* c fuzzy numbers are divided into 7, 3, and 3 classes, respectively. Table 1 shows the membership grade and the relative value for each region for Fig. 4(a).



(a) Hue



(b) Value



(c)Chroma

Figure 3. Reference 6-PFNs for color features.

For texture features, the directionality and contrast are addressed in our system. For directionality feature, the maximum within number of edge points of four orientations for each region is computed and assigned to a relative index value. For example, if the number of edge points of 0^0 is greater than others, then this index value is labeled to 1. Hence, based on such a design, four categories are used to represent the directionality feature. For contrast, we quantize it into 8 levels, and thus it is labelled into 8 indices.

For spatial relationship feature, a standard position index and the corresponding coordinates of its minimum bounding rectangle (x_l, x_r, y_t, y_b) are first defined. Based on Fig. 2(b), we define 36 kinds of spatial positions and spatial linguistic meaning descriptions. After computing the closest spatial distance between each selected region and each of 36 regions in standard definitions, each of selected regions for the image will have its spatial index.

2.3. Semantic Description

Based on the above indexing process for region features, a linguistic database may be constructed by means of artificial intelligence (AI) rules based on human knowledge and experiences. Our AI inference rules are defined as

If *HueIndex=A1*, and *ValueIndex=B1*, and *ChromaIndex=C1* and *EdgeIndex=D1* and *ContrastIndex=E1*, **then** *Output=F1*. ...
If *HueIndex=An*, and *ValueIndex=Bn*, and *ChromaIndex=Cn* and *EdgeIndex=Dn* and *ContrastIndex=En*, **then** *Output=Fn*,

where *Ai~Ei* denote the index values corresponding to the features processed by indexing process. Variable *Fi* is defined as index corresponding to the semantic meaning description of object. For instance, "sky" object is denoted to 1. In our system, seven objects are defined, i.e., we use number 1 to 7 for representing "sky", "building", "flower", "grass", "tree", "ground", and "sunset", respectively.

2.4. Recall Process

According to the forward results, the semantic description and feature definitions will be utilized again and regarded as input data. The recall process can automatically catch the object semantic index and the spatial index corresponding to features in database to represent the mental image. That is, using color features and spatial information, we can obtain the rough mental image based on the semantic description. In our approach, color features have been represented as a set of 6-PFNs, which have also applied for the reference fuzzy numbers of (H,V,C) complements. In order to obtain a set of color values (H,V,C) to specify the object colors, a mean of area method (**MOA**) is used to export the reference fuzzy numbers and is expressed as

$$\begin{aligned} z^* = {} & a_2 + \frac{1}{2} \times w_2 \times (a_1 - a_2) + \frac{1}{2} \times w_2 \times (b_2 - b_1) \\ & + \frac{(w_2 + w_3) \times (A - a_1)}{2} + \frac{(w_2 + w_3) \times (b_1 - B)}{2} \\ & + w_3 \times (B - A), \end{aligned} \tag{3}$$

where $(a_2, a_1, A, B, b_1, b_2)$ are 6 parameters for parameterized fuzzy number, and z^* is an output value. Weight values, w_1, w_2, and w_3, are set to 0.0, 0.5, and 1.0, respectively. After performing MOA process, the exported value of reference fuzzy numbers shown in Fig. 3 with *hue* component is given to (37.25, 96.75, 118.25, 153.5, 220.75, 293.5, 318.0), and the value of the exported reference fuzzy numbers with *value* and *chroma* components is presented to (27.5, 57.5, 86.25) and (28.5, 59.0, 87.25), respectively. Table 2 shows the fixed color values (H,V,C) corresponding the color index.

3. Experiments

The experimental color images are obtained from a stock photo library and WWW. 424 image regions from the image source belonging to seven semantic classes are adopted. We divided the 424 semantic image regions into no overlapping training and test sets as shown in Table 3 for experiments.

Figure 4 shows the testing images. Based on our linguistic rule database, we select one of rules to present a common rule in recall process. The result of MOA for color feature is used to represent the color features of each region, and the spatial information is obtained via the spatial position of each specified region within an image. Table 2 shows the selected rule and fixed color values corresponding to each of the semantic descriptions. Figure 5 displays the recall and the relative linguistic meaning results. The semantic interpretations are shown in Table 4 corresponding to Fig. 4. In Table 4, there exists some regions which were expressed the error linguistic meaning.

4. Conclusions

An approach of image interpretation and recall process has been presented. The major purpose is to interpret the primary semantic description of each specified region of an image, and to represent the rough mental image such as human vision retrieval. In our proposed system, some problems may be produced in forward and recall processes. For forward process, some objects may be interpreted to the fault linguistic meaning because the inference rules in the current database are finite and cannot include all rules, in addition to the training images may be insufficient. For recall process, since only used color features and spatial information to represent the reconstructed mental image, it makes that the results cannot really represent the original object image. To improve these drawbacks, it will be an important future work especially to explore the human visual perception on scene understanding.

5. Acknowledgements

This work was supported in part by the National Science Council of Republic of China under Grant No. NSC 89-2218-E-007-025.

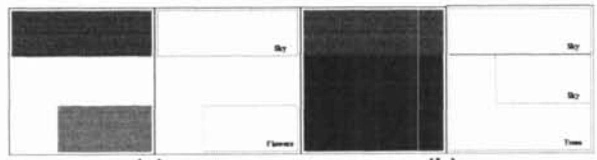
Table 1. The membership grade and relative index value of *H, V, and C* for each region of Fig. 4(a).

Figure 5(a)	Hue categories R (I=1)	YR-GY (I=2)	GY (I=3)	BG (I=4)	B (I=5)	P (I=6)	PR (I=7)	Hue index
Reg. 1	0.00	0.00	0.00	0.086	1.00	0.00	0.00	5
Reg. 2	0.00	0.00	0.00	0.00	0.00	1.00	1.00	6/7

Figure5(a)	Value categories Low (I=1)	Moderate (I=2)	High(I=3)	Value index
Reg. 1	0.00	0.00	1.00	3
Reg. 2	0.60	1.00	0.00	2

Figure5(a)	Chroma categories Low (I=1)	Moderate (I=2)	High(I=3)	Chroma index
Reg. 1	1.00	0.00	0.00	1
Reg. 2	1.00	0.00	0.00	1

(a) (b)
Figure 4. The testing images



(a) (b)
Figure 5. Results of recall and linguistic meaning representation corresponding to Fig. 4, respectively.

Table 2. Representation of the selected rule corresponding to each of the semantic descriptions and MOA result.

Object Semantic Description	Color Index Hue	Value	Chroma	Result of MOA (Hue,Value,Chroma)
sky	5	3	3	(220.75,86.25,87.25)
building	1	3	2	(37.5,86.25,59.0)
flower	6	3	2	(293.5,86.25,59.0)
grass	3	2	2	(118.25,57.5,59.0)
tree	4	1	3	(153.5,27.5,87.25)
ground	2	2	1	(96.75,57.5,28.5)
sunset	7	3	3	(318.0,86.25,87.25)

Table 3. Testing data with 7 semantic objects.

	Sky	Building	Flower	Grass	Tree	Ground	Sunset	Over all
Total # Regions	133	45	42	47	49	48	60	424
#Training regions	20	20	20	20	20	20	20	140
#Test regions	113	25	22	27	29	28	40	284

Table 4. The semantic description for each region corresponding to Fig. 4, respectively. Symbol * presents the error description for this region.

Figure 4		Semantic Description
(a)	Reg. 1	This region is a <u>*Sky*</u> and located in the <u>*Up-Horizontal*</u> of image.
	Reg. 2	This region is a <u>*Flowers*</u> and located in the <u>*Bottom-Biased-Right*</u> of image.
(b)	Reg. 1	This region is a <u>*Sky*</u> and located in the <u>*Up-Horizontal*</u> of image.
	Reg. 2	*This region is a <u>*Sky*</u> and located in the <u>*Center-Biased-Right*</u> of image.
	Reg. 3	This region is a <u>*Trees*</u> and located in the <u>*Biased-Bottom*</u> of image.

References

[1] Y. S. Chen, H.Y. Huang, and B. T. Chen, "An image understanding system designed for feeling the weather state from an image," *Journal of the Chinese Fuzzy Systems Association*, vol. 5, no.2, pp. 89-102, 1999.

[2] H. Y. Huang, Y. S. Chen and W. H. Hsu, "Human-based image interpretation using color and texture and spatial relationships features," *14th IPPR Conf. On Computer Vision, Graphics and Image Processing*, Taiwan, 2001.

[3] R. W. G. Hunt, *Measuring Color*, John Wiley and Sons, England, first edition, 1989.

[4] H. Y. Huang, Y. S. Chen and W. H. Hsu, "Color image segmentation using self-organizing map algorithm," *Journal of Electronic Imaging*, vol.11, no.2, pp.136-148, 2002.

[5] H. Tamura, S. Mori, and T. Yamawaki, "Texture features corresponding to visual perception," *IEEE Trans. Systems, Man, and Cybernetics*, vol.SMC-8, no.6, pp.460-473, 1978.