Joint 48th IEEE Conference on Decision and Control and
28th Chinese Control Conference
Shanghai, P.R. China, December 16-18, 2009

Color Image Retrieval and Classification Using Fuzzy Similarity Measure and Fuzzy Clustering Method

Xiaojuan Ban, Xiaolong Lv, and Jie Chen

***Abstract*—Color image retrieval and classification are very important in the field of image processing. In this paper, we propose a method which is in token of color characteristic of one image using hue and thereby is used to calculate similarity between two pictures. We also take F-Stat. Measure to find the best threshold so that we can realize fuzzy partition, and finally fuzzy c-means algorithm is used to do image classification and the calculation of subjection degree of one picture in the corresponding class. The corresponding mathematical model is established. Also, we construct a frame in which image comparability is obtained through fuzzy similarity measure based on hue feature vector of image, thus image retrieval is finished, and image classification is realized by fuzzy clustering. The comparison of image retrieval result between RGB feature vector and only-hue feature vector in token of image characteristic is taken. The experiment shows that the idea which takes this fuzzy similarity measure and fuzzy clustering method as the scheme of image retrieval and classification is reasonable and effective.**

Keywords: image retrieval, image classification, fuzzy similarity measure, fuzzy clustering

I. Introduction

As a hotspot in image processing, image retrieval and classification are very important. Image retrieval methods are mainly based on color, texture, shape and semantic-image[1]. Nowadays, the key parts of image retrieval and classification are mainly on three aspects. The first is diversification of similarity degree computing methods including color feature extraction[5,8,9,10] and multi-features combination[6,7], the second figures on mutual retrieval using feed back network[7], and the last is image classification using fuzzy clustering method[9]. Recently, precision and recall rate[2] are often used to evaluate the performance of image retrieval result.

In the field of image retrieval, Nachtegael, M[9] put forward a method to calculate image similarity measure using fuzzy partition of the HSI color space. In this method, when some color doesn't exist in a picture for calculation of similar degree of this color proportion between two pictures, the similarity measure remains 0 with increasing of this color proportion in the other picture; this is apparently not reasonable even though this method seems reasonable as a whole. Furthermore, the hue feature including 360 elements is divided into 8 parts, apparently, the amount is not enough and in the computation of complex image similarity measure, big error will be produced. In the field of image classification, Liu Pengyu[4], etc. propose modified fuzzy c-means algorithm to solve the problem of large-scale image retrieval and classification, the precondition of time saving is considered , but some pictures are always considered one clustering center, as a result, the classification result will not be the best.

In this paper, we propose a color image retrieval and classification method using fuzzy similarity measure and fuzzy clustering. In this method, similar measure of two pictures will be reasonable in the very condition that some color does not exist in a picture. Meanwhile, the selection of the clustering center will be more reasonable than the method above and so the image classification result will be more reasonable too.

II. Mathematical Description of the Problem

A. Mathematical Description of Sample Feature

Hue will remain almost the same in the case of small illumination and saturation alternation. In enormous picture library, we often meet the situation that the same or similar pictures have some differences in illumination and saturation, so it is essential for us to eliminate the impact brought by these two features' changes. So we only use hue histogram to represent image feature and calculate sample similar value. We know that hue value has 360 entries, how ever, a person can't distinguish two close hue value, so it's reasonable to use modified hue feature vector which set the size of hue feature vector 180. Suppose one sample u_i ,then $u_i = \{x_{i1}, x_{i2}, ..., x_{ik}, ..., x_{im}\}$, m is 180. x_{ik} represents the

Manuscript received February 20, 2009. This work is supported by National Natural Science Foundation of P.R.China (No. 50634010 and No. 60503024), Beijing Natural Science Foundation of P.R.China (No. 4092028) and Beijing Key Discipline Development Program of P.R.China(No. XK100080537).

Xiaojuan Ban, professor, is in School of Information Engineering, University of Science and Technology Beijing, Beijing, 100083, P.R.C., and is in Key Laboratory for Advanced Control of Iron and Steel Process (Ministry of Education), University of Science and Technology Beijing, Beijing, 100083, P.R.C. (corresponding phone: 86-10-62334980; fax: 86-10-62332281; e-mail: banxj@ies.ustb.edu.cn). She majors in Artificial Intelligence.

Xiaolong Lv, master, is in School of Information Engineering, University of Science and Technology Beijing, China. He majors in Artificial Intelligence(e-mail: iamlong510@yahoo.com.cn).

Jie Chen, professor, is in School of Automatization, Beijing Institute of Technology, China. He majors in Pattern Recognition and Intelligent System (e-mail: chenjie@bit.edu.cn).

kth element of the feature of sample i (picture i). x_{i1} represents the proportion of $H=1^\circ$ or $H=2^\circ$ in picture i. And x_{i2} represents the proportion of $H=3^\circ$ or $H=4^\circ$ in picture i. In the same way, x_{ik} represents the proportion of $H=(2k-1)^\circ$ or $H=(2k)^\circ$ in picture i.

TABLE I

THE CALCULATION METHOD OF H

H	condition
$60\times\frac{G-B}{R-B}$	$R\ge G\ge B$
$60\times\left(2-\frac{R-B}{G-B}\right)$	$G>R\ge B$
$60\times\left(2+\frac{B-R}{G-R}\right)$	$G\ge B>R$
$60\times\left(4-\frac{G-R}{B-R}\right)$	$B>G>R$
$60\times\left(4+\frac{R-G}{B-G}\right)$	$B>R\ge G$
$60\times\left(6-\frac{B-G}{R-G}\right)$	$R\ge B>G$

It's obvious that $\forall k(1\le k\le m)$, $0\le x_{ik}\le 1$ .i.e. The original range(360) is divided into 180 parts. Every part has two hue values. The corresponding proportion of pixel in which the hue is equal to this part's two hue values is one element of this picture's feature. So, m elements are combined to form the feature vector of the picture. It's not necessary to do data normalization since every element in the feature vector is in [0, 1]. To some extent, the efficiency is improved. H is used to replace hue, it's calculation method is shown in Table 1.

B. Mathematical Description of Sample Similarity Value

Definition I. The similar value between picture u_i and picture u_j is r_{ij}

$$r_{ij}=1-\frac{\left(p\times\sum_{k=1}^{m}\left|x_{ik}-x_{jk}\right|\right)}{\max\limits_{1\le e,f\le n,e\ne f}\sum_{k=1}^{m}\left|x_{ek}-x_{fk}\right|} \tag{1}$$

Note: $0<p<1$, p is fuzzy similarity measure calculation constant. (p plays the role that doesn't let the sample similarity value of the most different two pictures be 0).

Obviously, when $x_{ik}=0$, if x_{jk} increases, then r_{ij} decrease. Videlicet, the unreasonable case that when the number of some hue value pixel in one picture is 0, with the increasing of the pixel number of the hue value in another picture, the similarity value does not change will not take place.

C. Mathematical Description of F-Stat. Measure

Note: Given threshold λ, image classification need λ and transfer close-wrap $t(R)$, class number is c, and n_i the number of pictures in calss i. All pictures in class i are $u_1^i,u_2^i,...,u_{ni}^i$.

Note: $\bar{x}_k^i=\frac{1}{n_i}\sum_{j=1}^{n_i}x_{jk}(k=1,2,...,m)$, It's the average value of the kth feature element in class i. Suppose that the clustering center vector of class i $\bar{u}^i=\left(\bar{x}_1^i,\bar{x}_2^i,...,\bar{x}_m^i\right)$, and the clustering center vector of all pictures is $\bar{u}=\left(\bar{x}_1,\bar{x}_2,...,\bar{x}_m\right)$, where $\bar{x}_k=\sum_{j=1}^{n}x_{jk}\Big/n,(k=1,2,...,m)$.

We use F to represent F-Stat. Measure, then

$$F=\sum_{i=1}^{c}\frac{n_i\,\|\bar{u}^i-\bar{u}\|^2}{(c-1)}\Bigg/\sum_{i=1}^{c}\sum_{j=1}^{n_i}\frac{\|u_j^i-\bar{u}^i\|^2}{(n-c)} \tag{2}$$

Note: The difference between $\bar{u}^i$ and $\bar{u}$ is calculated by the formula: $\|\bar{u}^i-\bar{u}\|=\sqrt{\sum_{k=1}^{m}\left(\bar{x}_k^i-\bar{x}_k\right)^2}$, Similarly, The formula $\|u_j^i-\bar{u}^i\|$ represents the difference between u_j^i in class i and the clustering center $\bar{u}^i$.

Apparently, the classification result will be more reasonable when all pictures in one class are closer to each other and all class centers are farther from each other.

D. Mathematical Description of Fuzzy-c Partition Matrix

Suppose that sample fuzzy-c partition matrix $A=(a_{ij})_{c\times n}$.

Note: c is the number of classes when the threshold is λ (not including those classes having only one sample). n is total number of all pictures. a_{ij} represents the subjection degree of u_j in class i, $\forall i,j(1\le i\le c,1\le j\le n)$, $0\le a_{ij}\le 1$, $\forall j(1\le j\le n),\sum_{i=1}^{c}a_{ij}=1$, it denotes that the summation of all subjection degrees of sample u_j in all classes is 1.

E. Mathematical Model

Based on these descriptions of sample feature, similarity measure, F-Stat. Measure, sample fuzzy-c partition matrix, we describe the image retrieval problem as:

Precondition: u_1 is the picture needed to be retrieved, and $u_2\sim u_n$ is the search library of pictures.

Requirement 1: using modified hue feature to represent sample Stat. index, calculating the similar degree between two pictures using formula 2, and recompose $u_2\sim u_n$ in terms of big first, small second rule.

Requirement 2: using fuzzy-c means method to finish fuzzy classification, using F-Stat. Measure to find the best threshold, then calculating the best fuzzy partition matrix A_{best}, and list the classification result. Finally, $\forall i$, the subjection degree of u_i is obtained.

III. Resolvent of the Problem

A. Image Retrieval Method

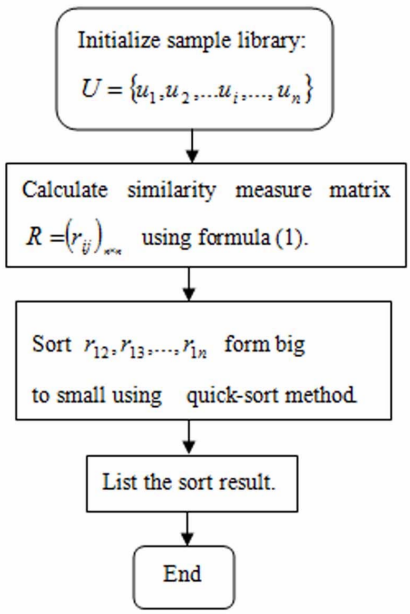


Fig.1. Image retrieval process

The similarity measure plays a key role in the image retrieval result. We finish image retrieval using modified hue feature vector to characterize the image to calculate the similarity value in advance. Thereby, image retrieval is finished. The realization process is shown in Fig.1.

B. Image Classification Method and Realization

1) Calculation of Transfer Close-wrap of Similarity Measure Matrix

Firstly, we calculate $R^2 = R \circ R$, if $R^2 = R$, then $t(R) = R$, else calculate $R^4 = R^2 \circ R^2$, if $R^4 = R^2$, then $t(R) = R^2$, similarly, if $R^{2(k+1)} = R^{2k}$, then $t(R) = R^{2k}$. Finally, we obtain transfer close-wrap which is also fuzzy equivalence matrix.

2) Calculation of Initial Fuzzy-c Partition Matrix

Calculation process of initial fuzzy-c partition matrix $A^{(0)}$ is: the kind of class including only one sample is not in the consideration and those including 2 samples at least is in the fuzzy-c partition matrix. If u_j belongs to class i, then $a_{ij} = h$, h is a constant, $0.5 < h < 1$, $\forall k(1 \le k \le c, k!=i)\ a_{kj} = (1-h)/(c-1)$. The goal that initial value is evaluated in this way consists with the idea that making the matrix have assorted characteristic and thereby reducing the convergence time of fuzzy-c means algorithm latter, finally improve the efficiency of the algorithm. Given $c \in \{2,3,...,n-1\}$.

3) Calculation of Best Fuzzy c-partition Matrix

Clustering analysis after confirming the best classification could make the result more intuitionistic. Also, the calculation process can be considered as an uncertain system[3], and we could find better result at full steam, it will be better if the end condition ε is smaller to some extent. The process of fuzzy clustering is as follows:

a) Initial fuzzy c-partition matrix $A^{(0)} = (a_{ij})_{c\times n}$

Note: c is the classes number under the best threshold λ_{best} (not including the kind of class having only one sample). Some calculation details are mentioned in section 3.2.2, it will not be mentioned again. $\forall i, j(1 \le i \le c, 1 \le j \le n)$, $0 \le a_{ij} \le 1$, $\forall j(1 \le j \le n), \sum_{i=1}^{c} a_{ij} = 1$, it shows that the summation of every sample u_j in c fuzzy subclasses is 1.

b) When the iteration number is $d(d = 0,1,2...)$, calculate clustering center vector $v_i^{(d)} = \sum_{j=1}^{n} (a_{ij}^{(d)})^r u_j \Big/ \sum_{j=1}^{n} (a_{ij}^{(d)})^r, 1 \le i \le c$.

c) Calculate $A^{(d+1)}$

$$a_{ij}^{(d+1)} = \frac{1}{\sum_{k=1}^{c} \left(\frac{\| u_j - v_i^{(d)} \|}{\| u_j - v_k^{(d)} \|} \right)^{\frac{2}{r-1}}}, 1 \le i \le c, 1 \le j \le n$$

d) Calculate the difference between fuzzy c-partition matrix $A^{(d+1)}$ and $A^{(d)}$ using

$$\| A^{(d+1)} - A^{(d)} \| = \sqrt{\sum_{i=1}^{c} \sum_{j=1}^{n} \left(a_{ij}^{(d+1)} - a_{ij}^{(d)} \right)^2},$$

if $\| A^{(d+1)} - A^{(d)} \| < \varepsilon$ ($0 < \varepsilon < 1$, ε is the end condition, a very small constant.), then the algorithm is over, else $d = d + 1$, return to step 2.

e) Finally, fuzzy c-partition matrix $A^{(last)} = (a_{ij}^{(last)})_{c\times n}$, marked as A_{best}. $\forall u_j \in U$, if $a_{ij}^{(last)} = \max_{1\le k\le c} a_{kj}^{(last)}$, then put u_j in class i in the matrix $A^{(last)}$, and $a_{ij}^{(last)}$ is the subjection degree of sample j belongs to class i. The subjection degree calculation of the kind of class having only one sample is not necessary.

4) Process of Image Classification Algorithm

The process of image classification algorithm is shown in Fig 2.

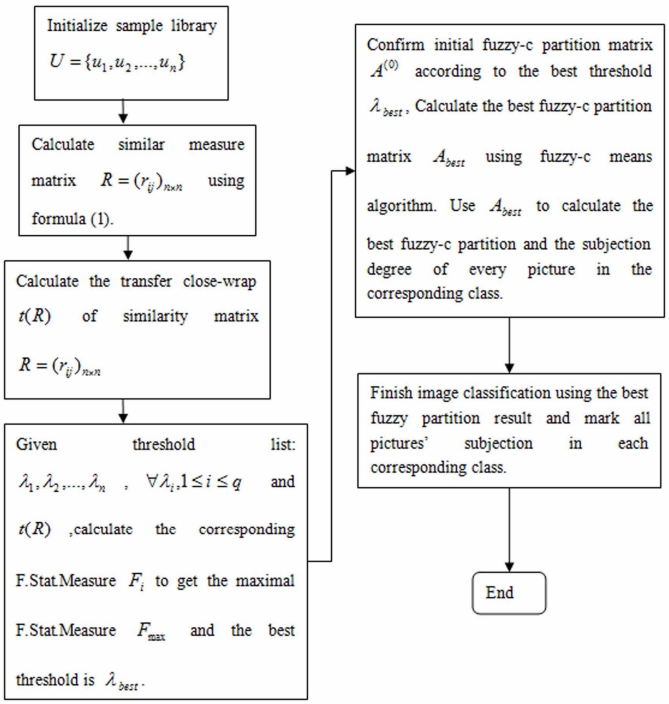


Fig. 2. Image classification process

IV. Experiment and Analysis

Precondition: On the assumption that the total number of all picture samples(including the picture needed to be retrieved) $n=21$, fuzzy similarity value calculation constant $p=0.96$, given the threshold list in the best threshold computing process is $\{0.60,0.61,0.62,\ldots,0.85\}$, The constant confirming initial partition is $h=0.8$,the iteration end standard $\varepsilon=0.001$.

The picture needed to be retrieved and given picture gallery is shown in Fig 3.

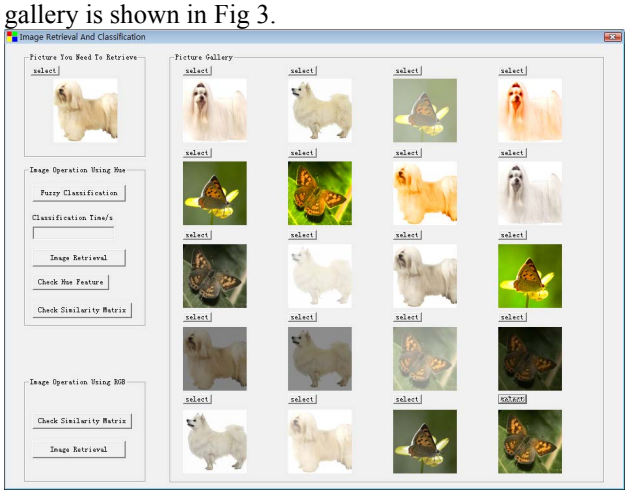


Fig. 3. All picture samples

In Fig 3, the leftmost picture (The serial number is 0) is needed to be retrieved. All other pictures are given in picture gallery (The serial number is 1,2,3,…,20 counting from the left right firstly, and from the top down secondly). Saliently, several types exist. All pictures' dimension is 100×100.

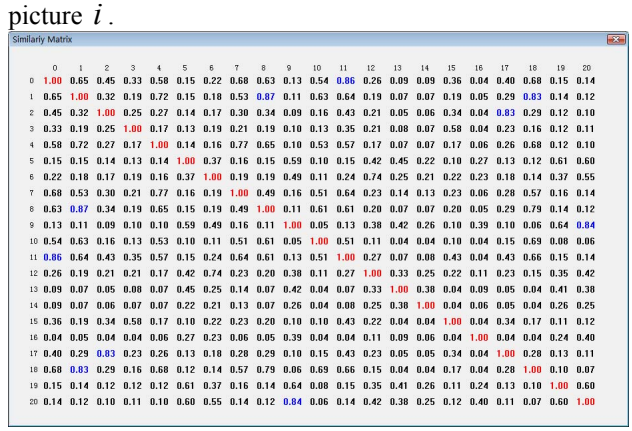
A. Comparison and Analysis of Image Retrieval Result

1) Image Similarity Measure Computing using RGB Index

Three Stat. histograms represent one pictures' feature. The first is color attribute RED histogram, the second is color attribute GREEN histogram, and the last is color attribute BLUE histogram. The vector sizes are all reduced to 128(original: 256). Note that The Stat. index is $u_i=\{v_{ir},v_{ig},v_{ib}\}$, and

$$v_{ir}=\{x_{ir1},x_{ir2},x_{ir3},\ldots,x_{ir128}\}$$
$$v_{ig}=\{x_{ig1},x_{ig2},x_{ig3},\ldots,x_{ig128}\}$$
$$v_{ib}=\{x_{ib1},x_{ib2},x_{ib3},\ldots,x_{ib128}\}$$

Note: x_{irk} is the proportion of pixels in which the RED attribute is $2(k-1)$ or $2k-1$ in picture i. Similarly, x_{igk} is the proportion of pixels in which the GREEN attribute is $2(k-1)$ or $2k-1$ in picture i, x_{ibk} is the proportion of pixels in which the BLUE attribute is $2(k-1)$ or $2k-1$ in picture i.



Similarly Matrix

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
0	1.00	0.65	0.45	0.33	0.58	0.15	0.22	0.68	0.63	0.13	0.54	0.86	0.26	0.09	0.09	0.36	0.04	0.40	0.68	0.15	0.14
1	0.65	1.00	0.32	0.19	0.72	0.15	0.18	0.53	0.87	0.11	0.63	0.64	0.19	0.07	0.07	0.19	0.05	0.29	0.83	0.14	0.12
2	0.45	0.32	1.00	0.25	0.27	0.14	0.17	0.30	0.34	0.09	0.16	0.43	0.21	0.05	0.06	0.34	0.04	0.83	0.29	0.12	0.10
3	0.33	0.19	0.25	1.00	0.17	0.13	0.19	0.21	0.19	0.10	0.13	0.35	0.21	0.08	0.07	0.58	0.04	0.23	0.16	0.12	0.11
4	0.58	0.72	0.27	0.17	1.00	0.14	0.16	0.77	0.65	0.10	0.53	0.57	0.17	0.07	0.07	0.17	0.06	0.26	0.68	0.12	0.10
5	0.15	0.15	0.14	0.13	0.14	1.00	0.37	0.16	0.15	0.59	0.10	0.15	0.42	0.45	0.22	0.10	0.27	0.13	0.12	0.61	0.60
6	0.22	0.18	0.17	0.19	0.16	0.37	1.00	0.19	0.19	0.49	0.11	0.24	0.74	0.25	0.21	0.22	0.23	0.18	0.14	0.37	0.55
7	0.68	0.53	0.30	0.21	0.77	0.16	0.19	1.00	0.49	0.16	0.51	0.64	0.23	0.14	0.13	0.23	0.06	0.28	0.57	0.16	0.14
8	0.63	0.87	0.34	0.19	0.65	0.15	0.19	0.49	1.00	0.11	0.61	0.61	0.20	0.07	0.07	0.20	0.05	0.29	0.79	0.14	0.12
9	0.13	0.11	0.09	0.10	0.10	0.59	0.49	0.16	0.11	1.00	0.05	0.13	0.38	0.42	0.26	0.10	0.39	0.10	0.06	0.64	0.84
10	0.54	0.63	0.16	0.13	0.53	0.10	0.11	0.51	0.61	0.05	1.00	0.51	0.11	0.04	0.04	0.10	0.04	0.15	0.69	0.08	0.06
11	0.86	0.64	0.43	0.35	0.57	0.15	0.24	0.64	0.61	0.13	0.51	1.00	0.27	0.07	0.08	0.43	0.04	0.43	0.66	0.15	0.14
12	0.26	0.19	0.21	0.21	0.17	0.42	0.74	0.23	0.20	0.38	0.11	0.27	1.00	0.33	0.25	0.22	0.11	0.23	0.15	0.35	0.42
13	0.09	0.07	0.05	0.08	0.07	0.45	0.25	0.14	0.07	0.42	0.04	0.07	0.33	1.00	0.38	0.04	0.09	0.05	0.04	0.41	0.38
14	0.09	0.07	0.06	0.07	0.07	0.22	0.21	0.13	0.07	0.26	0.04	0.08	0.25	0.38	1.00	0.04	0.06	0.05	0.04	0.26	0.25
15	0.36	0.19	0.34	0.58	0.17	0.10	0.22	0.23	0.20	0.10	0.10	0.43	0.22	0.04	0.04	1.00	0.04	0.34	0.17	0.11	0.12
16	0.04	0.05	0.04	0.04	0.06	0.27	0.23	0.06	0.05	0.39	0.04	0.04	0.11	0.09	0.06	0.04	1.00	0.04	0.04	0.24	0.40
17	0.40	0.29	0.83	0.23	0.26	0.13	0.18	0.28	0.29	0.10	0.15	0.43	0.23	0.05	0.05	0.34	0.04	1.00	0.28	0.13	0.11
18	0.68	0.83	0.29	0.16	0.68	0.12	0.14	0.57	0.79	0.06	0.69	0.66	0.15	0.04	0.04	0.17	0.04	0.28	1.00	0.10	0.07
19	0.15	0.14	0.12	0.12	0.12	0.61	0.37	0.16	0.14	0.64	0.08	0.15	0.35	0.41	0.26	0.11	0.24	0.13	0.10	1.00	0.60
20	0.14	0.12	0.10	0.11	0.10	0.60	0.55	0.14	0.12	0.84	0.06	0.14	0.42	0.38	0.25	0.12	0.40	0.11	0.07	0.60	1.00

Fig. 4. The similarity measure matrix using RGB vector to characterize the picture

The difference between picture i and picture j under the RED attribute is $DEF_{rij}=\sum_{k=1}^{128}\left|x_{irk}-x_{jrk}\right|$. Similarly, the difference between picture i and picture j under the GREEN attribute is $DEF_{gij}=\sum_{k=1}^{128}\left|x_{igk}-x_{jgk}\right|$ and the difference between picture i and picture j under the BLUE attribute is $DEF_{bij}=\sum_{k=1}^{128}\left|x_{ibk}-x_{jbk}\right|$. The whole difference between picture i and picture j is $DEF_{ij}=(DEF_{rij}+DEF_{gij}+DEF_{bij})/3$.

The final similarity measure of these two pictures is $r_{ij} = 1 - p \times (DEF_{ij} \big/ \max_{1 \le e,f \le n, e \ne f} DEF_{ef})$.

The similarity measure matrix result is shown in Fig.4. The retrieval result is shown in Fig.5.

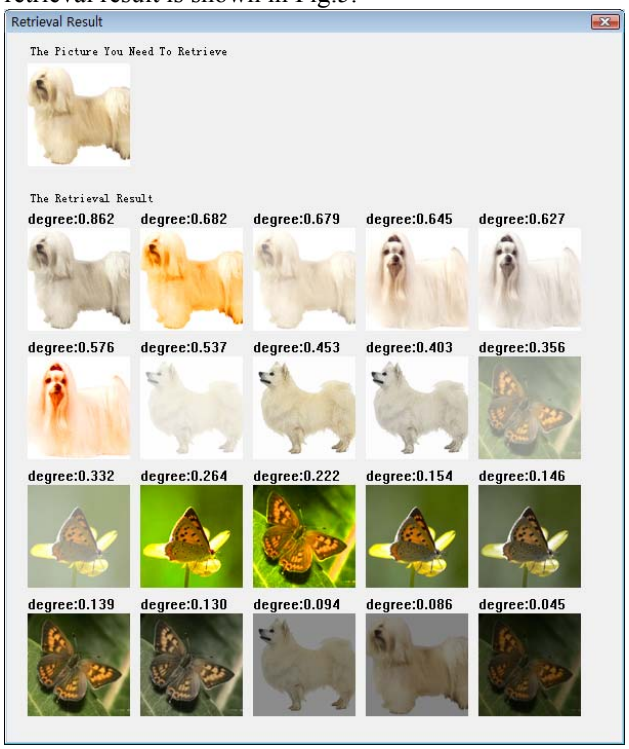


Fig. 5. Image retrieval result using RGB vector to characterize the picture

In Fig.5, the decimal fraction above the picture is the corresponding similar value.

From the similarity measure matrix and retrieval result, we know that it's very sensitive to the change of illumination and saturation using RGB vector to characterize the picture. This problem leads to the low similar value of two similar pictures such as the picture needed to be retrieved and picture 13 in the picture gallery. It indicates that the retrieval result is not good using RGB feature vector.

2) Image Similarity Measure Computing using Hue Index

The hue histograms of all pictures are shown in Fig.6.

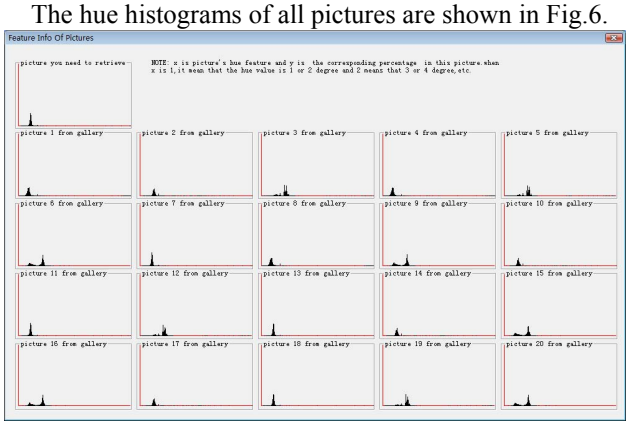
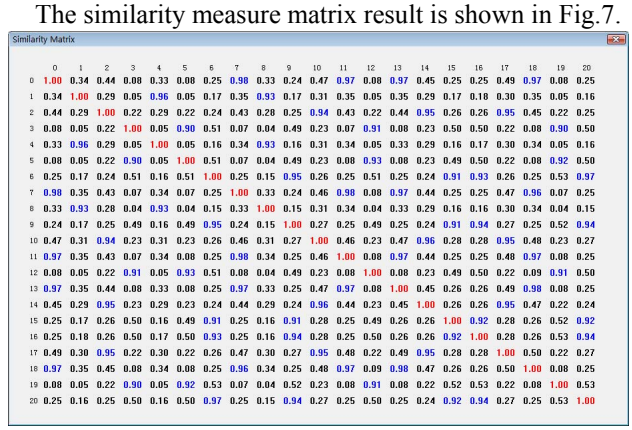


Fig. 6. Hue histograms of image

In Fig.6, the topmost is the hue histogram of the picture needed to be retrieved. The others come from the picture gallery. Obviously, picture 0(needed to be retrieved), 7,11,13, 18 are subject to one type, picture 1,4,8 are subject to another type. In every histogram, abscissa x is an integer, and $1 \le x \le 180$, $y \in [0,1]$, the point (x, y) in the histogram of picture i represents the proportion of pixels in which the hue value is $H = 2x - 1$ or $H = 2x$. It's easy to find that these 2 results are the same as the intuitionistic result from Fig.3. So do other types. So, it is obvious that in the calculation of similarity measure, this feature selection is effective. Of course, in the situation that two picture's hue histograms are close to each other while the corresponding pixels' positions are very different from each other or the size of the pictures is very large, we can divide every picture into 2 parts, 4 parts,…, 2^k parts etc equally ($k \ge 1$). And then we calculate the hue feature vector of every part in the same place similarly and add all parts' differences to get the total difference of two pictures. Thereby, we can get the total similarity measure result. (The value of k is decided in frondose case). This paper refers to fuzzy similarity measure and fuzzy partition mainly, we can change some details in some special cases just like picture dividing, we will not give the unnecessary details below.

The similarity measure matrix result is shown in Fig.7.



	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
0	1.00	0.34	0.44	0.08	0.33	0.08	0.25	0.98	0.33	0.24	0.47	0.97	0.08	0.97	0.45	0.25	0.25	0.49	0.97	0.08	0.25
1	0.34	1.00	0.29	0.05	0.96	0.05	0.17	0.35	0.93	0.17	0.31	0.35	0.05	0.35	0.29	0.17	0.18	0.30	0.35	0.05	0.16
2	0.44	0.29	1.00	0.22	0.29	0.22	0.24	0.43	0.28	0.25	0.94	0.43	0.22	0.44	0.95	0.26	0.26	0.95	0.45	0.22	0.25
3	0.08	0.05	0.22	1.00	0.05	0.90	0.51	0.07	0.04	0.49	0.23	0.07	0.91	0.08	0.23	0.50	0.50	0.22	0.08	0.90	0.50
4	0.33	0.96	0.29	0.05	1.00	0.05	0.16	0.34	0.93	0.16	0.31	0.34	0.05	0.33	0.29	0.16	0.17	0.30	0.34	0.05	0.16
5	0.08	0.05	0.22	0.90	0.05	1.00	0.51	0.07	0.04	0.49	0.23	0.08	0.93	0.08	0.23	0.49	0.50	0.22	0.08	0.92	0.50
6	0.25	0.17	0.24	0.51	0.16	0.51	1.00	0.25	0.15	0.95	0.26	0.25	0.51	0.25	0.24	0.91	0.93	0.26	0.25	0.53	0.97
7	0.98	0.35	0.43	0.07	0.34	0.07	0.25	1.00	0.33	0.24	0.46	0.98	0.08	0.97	0.44	0.25	0.25	0.47	0.96	0.07	0.25
8	0.33	0.93	0.28	0.04	0.93	0.04	0.15	0.33	1.00	0.15	0.31	0.34	0.04	0.33	0.29	0.16	0.16	0.30	0.34	0.04	0.15
9	0.24	0.17	0.25	0.49	0.16	0.49	0.95	0.24	0.15	1.00	0.27	0.25	0.49	0.25	0.24	0.91	0.94	0.27	0.25	0.52	0.94
10	0.47	0.31	0.94	0.23	0.31	0.23	0.26	0.46	0.31	0.27	1.00	0.46	0.23	0.47	0.96	0.28	0.28	0.95	0.48	0.23	0.27
11	0.97	0.35	0.43	0.07	0.34	0.08	0.25	0.98	0.34	0.25	0.46	1.00	0.08	0.97	0.44	0.25	0.25	0.48	0.97	0.08	0.25
12	0.08	0.05	0.22	0.91	0.05	0.93	0.51	0.08	0.04	0.49	0.23	0.08	1.00	0.08	0.23	0.49	0.50	0.22	0.09	0.91	0.50
13	0.97	0.35	0.44	0.08	0.33	0.08	0.25	0.97	0.33	0.25	0.47	0.97	0.08	1.00	0.45	0.26	0.26	0.49	0.98	0.08	0.25
14	0.45	0.29	0.95	0.23	0.29	0.23	0.24	0.44	0.29	0.24	0.96	0.44	0.23	0.45	1.00	0.26	0.26	0.95	0.47	0.22	0.24
15	0.25	0.17	0.26	0.50	0.16	0.49	0.91	0.25	0.16	0.91	0.28	0.25	0.49	0.26	0.26	1.00	0.92	0.28	0.26	0.52	0.92
16	0.25	0.18	0.26	0.50	0.17	0.50	0.93	0.25	0.16	0.94	0.28	0.25	0.50	0.26	0.26	0.92	1.00	0.28	0.26	0.53	0.94
17	0.49	0.30	0.95	0.22	0.30	0.22	0.26	0.47	0.30	0.27	0.95	0.48	0.22	0.49	0.95	0.28	0.28	1.00	0.50	0.22	0.27
18	0.97	0.35	0.45	0.08	0.34	0.08	0.25	0.96	0.34	0.25	0.48	0.97	0.09	0.98	0.47	0.26	0.26	0.50	1.00	0.08	0.25
19	0.08	0.05	0.22	0.90	0.05	0.92	0.53	0.07	0.04	0.52	0.23	0.08	0.91	0.08	0.22	0.52	0.53	0.22	0.08	1.00	0.53
20	0.25	0.16	0.25	0.50	0.16	0.50	0.97	0.25	0.15	0.94	0.27	0.25	0.50	0.25	0.24	0.92	0.94	0.27	0.25	0.53	1.00

Fig. 7. The similarity measure matrix using hue vector to characterize the picture

In Fig.8, we can easily find that picture 7,11,13,18 all have a similar value larger than 9.0 with picture 0, also we will find that picture 1,2,4,8,10,14,17 all have a similar value larger than 0.3 with picture 0 (they are close to each other to some extent)even though they are not greatly same as picture 0. This result is the same as the intuitionistic sense. It illustrates that using hue vector to characterize picture's feature is reasonable.

The retrieval result is shown in Fig.8, In Fig.8, the pictures are sorted from high to low.

From the sort result, we know that it's not sensitive to the change of illumination and saturation using hue vector to characterize the picture. This finally brings a reasonable result which consists with person's intuitionistic sense.

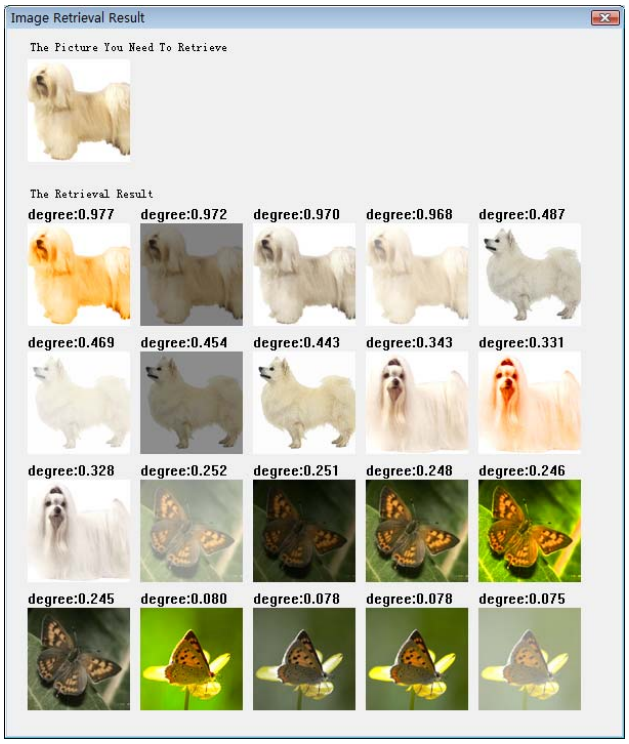


Fig. 8. Image retrieval result using hue vector to characterize the picture

3) Comparison of Image Retrieval Result using RGB Index and Hue Index to Calculate the Similarity Measure

In the experiment, it is obvious that the image retrieval result is bad using RGB feature vector in case of illumination and saturation change while it does not lose the validity using hue feature vector in the same situation. It illustrates that the similarity measure method using hue feature vector is more effective and reasonable. Of course, image classification will not be reasonable using RGB feature vector since the similarity measure is not reasonable using RGB feature vector.

B. *Image Classification Result and Analysis*

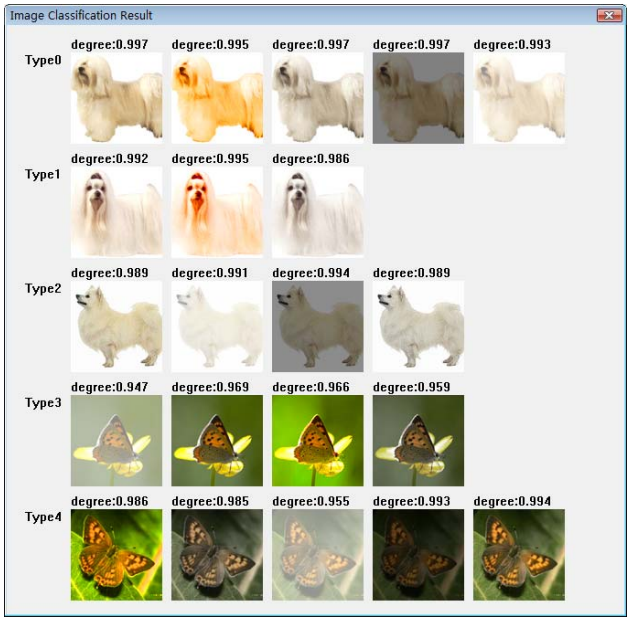


Fig. 9. Image classification result

In Fig.9, there are 5 types. Note that the figure above the picture is the subjection degree where the picture belongs to the corresponding type. Apparently, classification result and subjection degree are identical with the person's direct sense. This shows that image classification using fuzzy c-means algorithm to finish fuzzy clustering is effective.

Obviously, we can guess that the classification result using hue index will be better than using RGB index because in the preliminary step of classification, the retrieval result using the first is better than the second.

V. Conclusion

In this paper, image retrieval and classification problem are studied, an image retrieval and classification method using fuzzy similarity measure and fuzzy c-means algorithm to finish fuzzy clustering is proposed. In this method, similarity measure is calculated by hue feature vector which is not sensitive to illumination and saturation change and we reduce the size of hue vector on the premise that the retrieval result is not affected. We also use F-Stat. Measure to find the best threshold, use fuzzy-c means algorithm to carry out the analysis of fuzzy clustering, and finally realize image retrieval and classification. The experiment result illustrates this method's efficiency and feasibility. It can be extended to selection of video content and recognition of dynamic object.

References

[1] Yu Xiaohong, Xu Jinhua, "The Related Techniques of Content-Based Image Retrieval", Computer Science and Computational Technology, 2008. ISCSCT '08. International Symposium on, vol.1, pp.154-158.

[2] Shirahatti, N.V.Barnard, K, "Computer Vision and Pattern Recognition", 2005. CVPR 2005. IEEE Computer Society Conference on,vol.1, pp.955-961

[3] Yuanqing Xia, Jie Chen, P. Shi, G.P. Liu, D. Rees, "Guaranteed cost and positive control of uncertain systems via static output feedback", International Journal of Innovative Computing, Information and Control, vol. 4, No.6, 2008.

[4] Liu Pengyu, Jia Kebin, Zhang Peizhen, "AN EFFECTIVE METHOD OF IMAHE RETRIEVAL BASED ON MODIFIED FUZZY C-MEANS CLUSTERING SCHEME", Signal Processing, 2006 8th International Conference on, vol.3.

[5] Gwangwon Kang, Junguk Beak, Jongan Park, "Features Defined by Median Filtering on RGB Segments for Image Retrieval", Computer Modeling and Simulation, 2008, EMS '08. Second UKSIM European Symposium on, pp.436-440.

[6] Jeong-Yo Ha, Gye-Young Kim, Hyung-Il Choi, "The Content-Based Image Retrieval Method Using Multiple Features", Networked Computing and Advanced Information Management, 2008. NCM '08. Fourth International Conference on, vol.1, pp.652-657.

[7] Nagashima, K; Nakada, M; Osana, Y, "Similarity-based Image Retrieval by Self-Organizing Map with Refractoriness", Neural Networks, 2007, IJCNN 2007 International Joint Conference on, pp.2647-2652.

[8] Sabeti, L; Wu, Q.M.J, "New similarity measure for illumination invariant content-based image retrieval", Automation and Logistics, 2008. ICAL 2008. IEEE International Conference on, pp.279-283.

[9] Nachtegael, M; Van der Weken, D; De Witte, V; Schulte, S; Melange, T; Kerre, E.E, "Color Image Retrieval using Fuzzy Similarity Measures and Fuzzy Partitions", Image Processing, 2007. ICIP 2007. IEEE International Conference on,vol.6. pp.VI-13-VI-16.

[10] Wang, S.L; Liew, A.W.C, "Information-Based Color Feature Representation for Image Classification", Image Processing, 2007. ICIP 2007. IEEE International Conference on,vol.6. pp.VI-353-VI-356.