# DERIVING HIGH-LEVEL CONCEPTS USING FUZZY-ID3 DECISION TREE FOR IMAGE RETRIEVAL

*Ying Liu*[1], *Dengsheng Zhang*[1], *Guojun Lu*[1] , *Wei-Ying Ma*[2]

[1]Gippsland School of Computing and Information Technology,
Monash University, Vic, 3842, Australia
{ying.liu, dengsheng.zhang, guojun.lu}@infotech.monash.edu.au
[2]Microsoft Research Asia, No. 49 ZhiChun Road, Beijing, 100080, China
wyma@microsoft.com

**ABSTRACT**

To improve the retrieval accuracy of content-based image retrieval, an important task is to reduce the 'semantic gap' between low-level image features and the richness of human semantics. In this paper, we present a region-based image retrieval system with high-level semantic concepts used. The contribution of the paper is two-fold. First, salient low-level features are extracted from arbitrary-shaped regions. Second, a fuzzy-ID3 decision tree learning method is proposed to derive association rules which map low-level image features to high-level concepts. Experimental results prove that by reducing the 'semantic gap', the proposed system not only improves the retrieval accuracy, but also supports users in query-by-keyword.

## 1. INTRODUCTION

Content-based image retrieval (CBIR) was introduced in the early 1990's for large image database retrieval. CBIR systems index images by their own visual contents, such as color, shape, texture. Although many sophisticated algorithms have been designed for color, shape, and texture description, the performance of conventional CBIR systems is still far from satisfaction. This is due to the 'semantic gap' between the limited descriptive power of low-level image features and the richness of user semantics [1][2][3].

Literature survey shows that the methods developed for narrowing down the 'semantic gap' to improve retrieval accuracy include: 1) Using machine learning or data mining techniques to associate low-level image features with high-level concepts [2][3]. 2) Introducing Relevance feedback (RF) into the retrieval loop for continuous learning through on-line interaction with users [1]. As users are often interested in specific regions rather than the entire image [1], most current systems are region-based.

In this paper, we develop a region-based image retrieval (RBIR) system with high-level concepts used. Most work in this area focus on high-level semantics without much effort made in low-level region features. This paper makes the following contributions. First, salient low-level features are extracted from arbitrary-shaped regions. Second, we propose a fuzzy-ID3 decision tree method based on the work in [4] to derive a set of association rules which maps low-level image features to high-level concepts. Experimental results confirm the substantial performance of the system.

The remaining of the paper is organized as follows. Section 2 includes the related work and the research focus. Section 3 explains the proposed system in details. Experimental results are given in Section 4. Finally, Section 5 concludes this paper.

## 2. RELATED WORK AND RESEARCH FOCUS

In our system, each region is described by its color, texture and spatial location. We do not consider shape feature as it is not important for regions in natural scenes.

Instead of using traditional color features such as color moments or color histograms [1][5], we define a perceptual color for each region by its dominant color in HSV space which is more natural in vision.

Texture features commonly used in CBIR include spectral features such as Gabor features[6], statistical features such as coarseness, contrast, directionality [7]. Statistical texture features, though work well in classifying Brodatz textures, are less effective when applied to natural scene images [8]. Our experimental results also prove this to be true. In addition, there are two problems to be considered. 1) In many systems, texture features are obtained during segmentation from pixels or small blocks[5][8]. Such features may not well represent the property of an entire region. In some other systems, segmentation does not produce texture features. Hence, it is necessary to study texture feature extraction from the whole region after segmentation. 2) Transforms such as Gabor filtering require the input image to be rectangle. An instinctive way is to obtain an inner rectangle (IR) from a region on which filtering can be performed. This works when the size of the filtering mask is much smaller than the size of the IR. But many regions in RBIR systems are small, and the coefficients obtained can't well describe the region. To solve these problems, we present a extended rectangle(ER) texture feature extraction algorithm. By initial padding, this algorithm extends an arbitrary-shaped region into a larger rectangle onto which Gabor filtering is applied. Then a set of coefficients best describing the region is obtained, from which texture features can be extracted.

Besides color and texture features, we define spatial location of a region as 'bottom, middle, top' in an image.

Another focus in this paper is to derive high-level concepts from low-level features. Supervised learning such as Support Vector Machine (SVM)[1], Bayesian classifier[2] can be used for this purpose. The problem with such learning algorithms is that a large amount of labelled training data is required, but it is tedious and error-prone to provide such data. Another problem with SVM which is good at binary-decision making is that its performance degrades when applied to multiple concepts classification. Decision tree learning methods are mathematically less complex and have been used to deduce association rules associating low-level image features with high-level concepts [3]. In this paper, we present a 'fuzzy'-ID3 decision tree method.

## 3. SYSTEM DESCRIPTION

In this work, we use 'JSEG' [9] to segment each database image into regions homogeneous in color and texture.

### 3.1 Low-level features

#### *3.1.1 Color feature*

HSV is the most natural color space in visual. As the regions are color homogeneous, it is possible to use the average HSV value of all pixels as its perceptual color ('Ave-cl'). However, in some cases, due to the inaccuracy in image segmentation, pixels not belonging to the interested region might be included. Hence, we use the dominant color as the perceptual color of a region. For this, we first calculate the HSV space color histogram (10*4*4 bins) of a region and select the bin with maximum size. The average HSV value of all the pixels in the selected bin is used as the dominant color and referred to as 'Dm-cl'. Figure 1 gives two examples.

1(a) 1(b) 1(c) 2(a) 2(b) 2(c)

**Figure 1. Average color and dominant color**
(a) Original region (b) Ave-cl (c) Dm-cl

#### *3.1.2 Texture feature*

Our texture feature extraction algorithm includes two parts. Firstly, an arbitrary-shaped region of *M* pixels is extended into a rectangle (ER) of *N* ( $N \geq M$ ) pixels by padding some values outside its boundary. We use Zero padding due to its simplicity in implementation. Then a band of Gabor filters with 4 scales and 6 orientations are applied to the ER and we select the *M* largest coefficients from the *N* coefficients in each subband. From these selected coefficients, mean and variance are calculated as texture features. This is the simplified version of our algorithm in [10] in which an iterative loop is used to find the set of coefficients best describing the original region. We observed slight performance degrade using the simplified version.

To compare the performance of different texture features, we implemented texture feature extraction from IR as well. Due to the various possible region shapes available, it is difficult to obtain maximum size IR. We design a simple method which can find a reasonable large size IR, though may not necessarily the maximum. For each point *P* along the region boundary and within the region, we find the largest IR with its left-top corner being *P*. Among all the IRs obtained, we chose the one with maximum size. Figure 2 gives a few examples.

Note that for fast Gabor filtering, the height and width of ER and IR must be power of two.

#### *3.1.3 Spatial location*

Besides color and texture features, spatial location is also useful in region classification. For example, *sky* and *sea* could have similar color and texture features, but their spatial locations are different with *sky* usually appears at the top of an image and *sea* at the bottom.

We define 3 simple spatial locations as 'top, middle, bottom'. First, we define the spatial center $S(X,Y)$ of a region as in (1). *N* is the number of pixels in a region and $p_m(x,y)$ represents the x, y coordinate of a pixel. Then the spatial location of a region is defined as 'top' (middle, bottom) if $S(X,Y)$ is located in the top (middle, bottom) 1/3 of an image.

$$S(X,Y) = \frac{1}{N}\sum_{m=1}^{N} p_m(x,y) \qquad (1)$$

Finally, each region is described by a 3-d color feature, 48-d texture feature, and its spatial location. Each dimension of the color and texture features is normalized to [0,1].

### 3.2. Decision tree

ID3 is a decision tree method based on Shannon's information theory. Given a sample data set described by a set of attributes and an outcome, ID3 produces a decision tree which can classify the outcome value based on the values of the given attributes. Based on our image data, we define 10 classes (concepts): *grass, forest, sky, sea, sand, firework, sunset, flower, tiger, fur*. Here, we only consider black ape fur as included in our image data set. In our case, the attributes are the low-level image features and the outcome is one of the 10 concepts. We collect 400 sample data with 40 for each class.

At each level of the ID3 decision tree, the attribute with smallest entropy is selected from those attributes not yet used as the most significant for decision making. Spatial location is obviously less significant than color and texture features. We need to determine between color and texture which is more significant. We define the entropy of color feature $I_c$ as below.

1) For each class, we calculate the mean color feature of all the 40 regions as its *representative color feature* $C_i^R = \{c_i^h, c_i^s, c_i^v\}, i = 1,\ldots,10$, and we initialize the array counting the number of correct classification as Correct_Num[i]=0.

2) For the $j_{th}$ sample region in class *i*, we calculate its Euclidean distance to the representative color feature of each class $d_{(j,i)}^m, m = 1,\ldots,10$, and find the minimum value $d_{(j,i)}^{m_0}$. If $m_0 = i$, then we consider this region correctly classified based on color feature, and Correct_Num[i]++;

3) Repeat 2) for all the sample date. Thus, the probability of correct classification is P[i]=Correct_Num[i]/40, i=1,…,10.

4) The entropy of color feature is calculated as

$$I_c = \frac{1}{10}\sum_{i=1}^{10} -\log_2 P[i] \qquad (2)$$

In our results, $I_c = 0.134$. Similarly, we obtain the entropy of texture feature as $I_T = 0.192$. This tells that color feature is more significant and should be used at the first level of the tree to split the data into subsets. The original ID3 method requires the real region features to be discretized as input. In our method, we compare region feature to the *representative feature* of each class to obtain the corresponding class type of each region. In this way, the complex feature discretization procedure is avoided. We refer to our method as 'fuzzy'-ID3. We consider the classification successful if P[i]>0.7.

Finally, we obtain a decision tree as shown in Figure 3. For example, if the dominant color of a region is most close to the representative color of either concept *grass* or *forest* and its spatial position is 'top', then we classify it as *forest*. Results show that color feature can successfully distinguish *sunset, sand,* and *fur* from others. *Grass* and *forest* are classified into same class as their colors are similar, the same to *sky* and *sea*. Texture features can successfully recognize *flower, tiger, firework* regions. To separate *sky* from *sea*, *grass* from *forest*, spatial position is used. Statistics from our sample date set show that no *forest* region appears at the bottom of an image while 90% of the *grass* regions are. No *sky* region is at the bottom while 92% of the *sea* regions are.

Using testing sets of 250, 320, 400 data to test the performance of the decision tree, we obtain the average correct classification probability of each class as: *0.85, 0.8, 0.76, 0.72, 0.86, 0.81, 0.93, 0.72, 0.80, 0.77*, an overall average of 0.802.

### 3.3 Retrieval with high-level concepts

The user can either submit a keyword (one of the 10 concepts) as query, or specify an interested region from a query image. In this work, we specify an interested region from the query image and provide the corresponding keyword. The system first finds all database images containing region(s) of same concept. Then, all these candidate images are further ranked according to their EMD[11] distances to the query image. We refer to this process as '*retrieval with concepts*'.

## 4. EXPERIMENTAL RESULTS

We use 5,000 Corel images (pre-classified into different categories each containing 100 images) as test data. 'JSEG' segmentation produces 29187 regions (5.84 regions per image on average) with size no less than 3% of the original image. We ignore very small regions considering that regions should be large enough for us to study their texture features. Precision (Pr) and Recall (Re) are used to measure retrieval performance. Let $K$ be the number of images retrieved, with $K$=10,20,…100, we calculate the average Pr and Re of all the queries performed, and obtain the Pr~Re curve.

### 4.1 Different texture features

We extracted texture features from ER, IR. We also have the three most significant Tamura features-coarseness, contrast, directionality. To compare their performance, we performed 30 queries. For each query image submitted, the system ranks the database images according to their EMD distances to the query. Both color and texture features are used in distance calculation. This is referred to as '*retrieval without concepts*'. The results in Figure 4 show that ER outperforms the other two.

### 4.2 Retrieval with concepts

We selected 30 query images from the database, containing interested regions of different concepts. A few query images and the specified regions are given in Figure 5. We compare the performances of *retrieval with/without high-level concepts*. Figure 6 displays that using high-level concepts, the retrieval accuracy is improved. The retrieval results for query 1 and 2 (with keywords *sea*, *flower*) are given in Figure 7 as examples. Table 1 lists the number of relevant images retrieved with different total number of retrieved images ($K$) for query 1, 2, 3.

**Table 1. Number of relevant images retrieved**

| Q\K | 10 | 20 | 50 | 80 | 100 |
|---|---|---|---|---|---|
| 1 | 7 | 9 | 20 | 25 | 29 |
| 2 | 10 | 18 | 38 | 45 | 48 |
| 3 | 8 | 12 | 18 | 24 | 28 |

## 5. CONCLUSIONS

In this paper, we present a region-based image retrieval system with high-level concepts obtained from low-level region features using a fuzzy-ID3 decision tree. By reducing the 'semantic gap', the proposed system improves image retrieval accuracy. In addition, it can support not only query-by-sample, but also query-by-keyword which is more user-friendly.

In this work, we specify only one interested region in the retrieval process. We expect the retrieval performance to be further improved by considering multiple regions/concepts. In addition, the decision tree can be more robust by including a complete set of the available concepts in the database, and by learning through a larger sample dataset.

## 6. REFERENCES


[1]F. Jing, Mj Li,, L. Zhang, H-J Zhang, B. Zhang, "Learning in Region-based Image Retrieval", Proc. Inter. Conf. on Image and Video Retrieval(CIVR2003), 2003.

[2]A. Vailaya, M.AT Figueiredo, A.K. Jain, H.J. Zhang, "Image Classification for Content-based Indexing," *IEEE Trans. on Image Processing*, 10(1), pp117-130, 2001.

[3]Ishwar K.Sethi, Ioana L.Coman, "Mining Association Rules Between Low-Level Image Features and High-Level Concepts," Proc. of SPIE Data Mining and Knowledge Discovery, III, pp.279-290, 2001.

[4]J. R. Quinlan. *Discovering Rules by Induction from Large Collections of Examples*. In Expert Systems in the Micro-Electronic Age. Edinburgh University Press, 1979.

[5]J.Z.Wang, J.Li, G. Wiederhold, "SIMPLIcity: Semantics-Sentitive Integrated Matching for Picture Libraries," *IEEE Trans. Pattern and Mach.Intell*.Vol 23, no.9, pp947-963, 2001.

[6]W.Y.Ma,B.S. Manjunath, "A Texture Thesaurus for Browsing Large Aerial Photographs," *J. of the American Society for Information Science*, vol.49, (no.7), pp.633-48, May 1998.

[7]H. Tamura, S. Mori and T.Yamawaki, "Texture Features Corresponding to Visual Perception," *IEEE Trans. On Systems, Man and Cybernetics*, vol.8, No.6, pp460-473,1978.

[8]W..Leow, S.Lai, "Scale and Orientation-Invariant Texture Matching for Image Retrieval," *World Scientific*, 2000.

[9]Y.Deng, B.S.Manjunath and H.Shin "Color Image Segmentation," Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR '99, Fort Collins, CO, vol.2, pp.446-51, June 1999.

[10]Y.Liu, X.Zhou, W.Y.Ma, "Extraction of Texture Features from Arbitrary-Shaped Regions for Image Retrieval," Inter. Conf. on Multimedia and Expo (ICME04), Taipei, June 2004.

[11]Rubner, Y., Tomasi, C., and Guibas, L., "A Metric for Distributions with Applications to Image Databases," Proc. of IEEE Inter. Conf. on Computer Vision, Jan. 1998.

**Figure 2. Regions and their IR** (marked by red line)

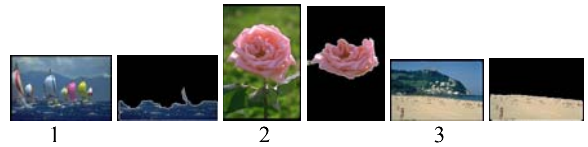


**Figure 5. Query image/region examples**

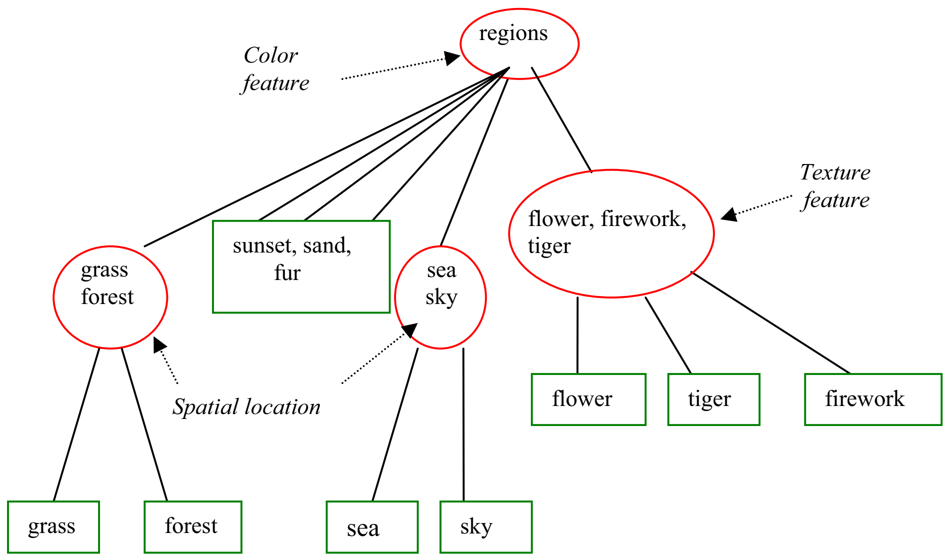


**Figure 3. Decision tree**

(Red circles contain concepts which have not been successfully classified yet. Green boxes contain concepts which are successfully classified. Italic terms are the attributes used in decision making)

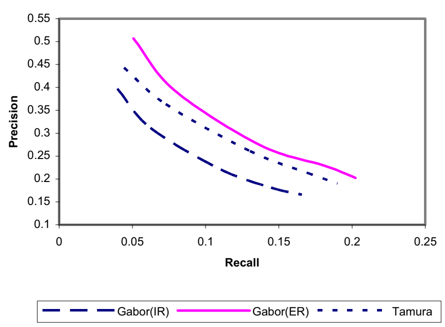


**Figure 4. Using different texture features**

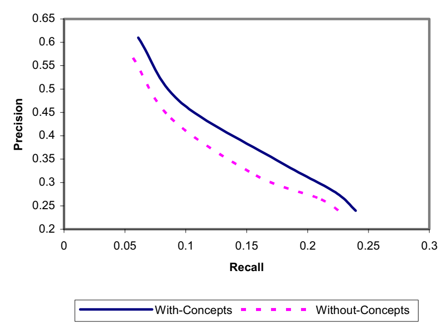


**Figure 6. Retrieval with/without concepts**

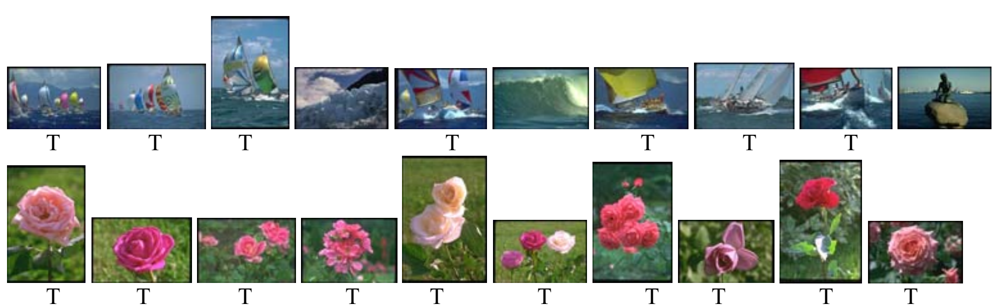


**Figure 7. Retrieval with concepts sea, flower** (‘T’ refers to the relevant images selected)