Semantic-meaningful Content-based Image Retrieval in Wavelet Domain

Yongqing Sun
Department of Science for Open and Environmental Systems
Keio University
3-14-1,hiyoshi,kouhoku-ku,yokohama-shi,Japan
syq@ozawa.ics.keio.ac.jp

Shinji Ozawa
Department of Science for Open and Environmental Systems
Keio University
3-14-1,hiyoshi,kouhoku-ku,yokohama-shi,Japan
ozawa@ozawa.ics.keio.ac.jp

ABSTRACT

In this paper, we propose a semantic-meaningful approach for region-based image retrieval in image database. Our retrieval system is based on wavelet transform for its decomposition property similarity with human visual processing. At first, with the fact that semantic region segmentation desires low frequency resolution, pixel clustering algorithm is applied for image partition in the Low-Low(LL) frequency subband of image wavelet transform . Secondly, with the fact that accurate region identification desires high frequency resolution, the feature vector of segmented region is hierarchically extracted from all the wavelet frequency subbands. Finally, in the distance function for region matching, the weights for feature components of the feature vector are tuned semantically. The experiment results demonstrate that our image retrieval system improves retrieval accuracy, robustness significantly in general-purpose image library.

Categories and Subject Descriptors

H.3.3 [**Information Search and Retrieval**]: clustering.

General Terms

Algorithm, Human Factors.

Keywords

Content-based image retrieval, semantic image segmentation, hierarchical feature vector, wavelet transform.

1. INTRODUCTION

Efficient automatic image retrieval system must make good use of the semantic content of image[1, 2]. Image regions that correspond to meaningful objects are the basic elements in the image to carry semantic information. So semantic segmentation of the image content and concise representation of the essence of the data are crucial steps for image retrieval and lead to the proliferation of a number of region-based image retrieval techniques in the literature [3, 4, 5, 6, 7, 8].

Wavelet transform plays a wide role in image processing and computer graphics because its subband and multiresolution decomposition are extremely adapted in description of image feature[9]. With proper level wavelet decomposition of the original image, the Low-Low(LL) frequency subband preserves image basic content and the other high frequency subbands are added details of image that characterize image variations in different directions.

Based on upon wavelet decomposition property, in this paper, a novel region-based image retrieval algorithm using wavelet transform is proposed. We try to analogize human visual processing system, in which human separates perceptual interest regions mainly through low frequency information of the image and recognizes them with the complement of high frequency information[10]. Semantic image segmentation is done in the Low-Low(LL) frequency subband which shows desired low frequency resolution of the original image, and region feature vector is hierarchically extracted from all the wavelet subbands which depict different spatial-frequency resolution of image content. In our indexing system, the boundaries between regions are deleted, which can make the description of region content more accurately and improve the robustness of region-based image retrieval system against uncertainty of segmentation. The advantages of the region-based image retrieval approach are highlighted in pixel, region, and image levels as the following:

In pixel level, clustering algorithm is applied to the wavelet coefficients in Low-Low(LL) subband. Using LL subband for data simplification is more effective than blocking or color quantization of the original image in other previous methods, which reduces the amount of raw pixel data naturally while preserving the information needed for the image understanding task effectively.

In region level, in contrast to describing region content using one dominant feature vector that is obtained by submerging the feature set, the region feature vector is extracted hierarchically from all the subband in the wavelet domain, in which each feature component is a certain spatial-frequency resolution for region content.

In image level, compared with some existing retrieval systems in which weight adjustment between separated color and texture feature components in distance function are im-

MIR'03, November 7, 2003, Berkeley, California, USA.

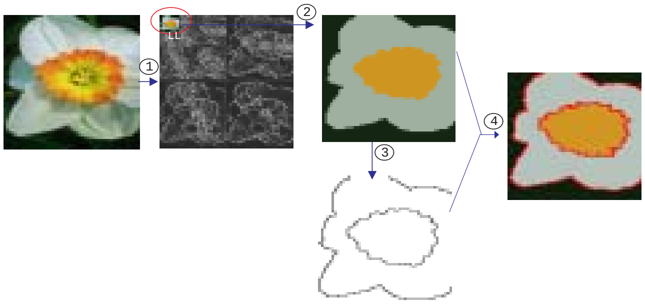

Figure 1: Image segmentation in the LL subband.(1)Image wavelet transform in the original image. (2)Clustering algorithm applied in the LL subband. (3)The localization of boundaries between objects. (4)The boundary deletion(marked by red lines)

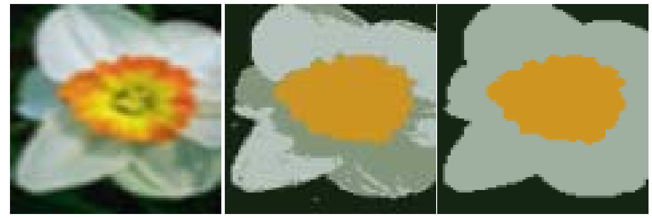

Figure 2: One image segmentation result using the K-means algorithm in the LL subband. Fig.a is original image, fig.b and c are segmentation results in the LL subband. For view conveniently, we enlarge the result. In fig.b, the K=4, and in fig.c, the optimal K=3 obtained by the equation 1.

plemented complexly and difficultly , our system tunes the weight values more semantically and flexibly because in our region feature vector, each component is a combined color and texture feature for region content.

In section 2, we explain how to effectively segment image into regions in pixel level. We propose how to extract hierarchical representative region feature vector in region level in section 3. Section 4 introduces region similarity measurement in image level. The experiment results are discussed in section 5, and the last is conclusion.

2. SEMANTIC-MEANINGFUL IMAGE SEGMENTATION WITHOUT BOUNDARY

In this section, image segmentation conducted in the LL subband is described. The semantic-meaningful process is illustrated using the fig.1. In step 1, according to our experimental experience, the wavelet transform is applied to image by 3 levels for desired division between basic image content and complement details. Secondly the clustering algorithm is applied in the LL subband. At last then the boundaries between the segmented regions are localized and deleted.

In this paper the K-means clustering algorithm [11] is applied in the LL subband for image segmentation. To reflect color and proper texture information, six features are used for clustering . Three features are three color components of each wavelet coefficient in the LL subband. In this paper, the HSV color space is chosen for its perceptual uniformity. To obtain the other three, first for each point in the LL subband, its corresponding points in the same spatial location at the neighboring HL_3, LH_3, HH_3 subbands(referring to fig.3) are localized, and then the H color components of them are taken as the rest three features of the clustered feature vector. The motivation for using these three features is their reflection of texture properties in high frequency to some extent, which is presented in [12].

The cluster number K is critical for the semantic effectiveness of image segmentation. Suppose the set $\{\mu_i, i = 1, \cdots, Q\}$ of Q points is to be clustered into K clusters with centers $\{\hat{\mu_j}, j = 1, \cdots, K\}$. In this paper, we iterate K between minimum cluster number K=2 and maximum cluster number K=12. We do not apply the clustering algorithm to the perceptually homogeneous images, that is, K=1. In our image retrieval system, one image is considered as uniform pattern if mean square deviation of H color values in the whole image is smaller than a threshold. In this paper it is set to 0.08. The maximum value K=12 shows that the image is assumed to have less than 12 different objects. Usually, for the practical images, the image segmentation produces much less number of objects. The maximum cluster number is rarely met. The "optimal" K is determined if it minimizes the *Cluster Dispensation Value* (CDV) defined as:

$$CDV = \frac{1}{K}\sum_{j=1}^{K} R_{C_j} \cdot \sum_{\mu_i \in C_j} D(\mu_i, \hat{\mu_j}) \quad (1)$$

Where $D(\mu_i, \hat{\mu_j})$ is the Euclidean distance between μ_i and the center $\hat{\mu_j}$ of the $j - th$ cluster C_j. R_{C_j} is the smallest rectangular area surrounding the cluster C_j. We can see that too clumped or too scattered region segmentation would all make the CDV value large in equation 1. In figure.2 we give one image segmentation example, which shows the result of the K-means algorithm applied to the LL subband. In fig.2.c, K=3 is the optimal cluster number according to the equation 1. Compared with this, in fig.2.b where K=4, the computed CDV value is larger, and we can also observe that the result of pixel clustering is a little scattered. With minimized CDV value, more semantic-meaningful clusters can be obtained.

Semantically precise image segmentation still keeps a difficult problem now for creating an artificial algorithm whose performance is close to the human visual system. Human can identify distinct objects in an image and give meaningful assignment of the pixel points in the image. Although those pixel points cannot be assigned unambiguously to objects, human visual recognition system always performs well. Based on above observation, in this paper, the boundary points between segmented regions are first localized in the clustered LL subband, and then those boundary points with their 3×3 neighborhood are deleted. The advantage of deleting ambiguous information of boundary in the image is that it can improve the robustness of the image retrieval system against segmentation-related uncertainty. Furthermore, the feature vector of segmented region can be extracted more exactly without including those boundary points.

3. THE EXTRACTION OF HIERARCHICAL FEATURE VECTOR OF REGION

Compared with region segmentation that prefers low frequency resolution, for region identification, high frequency resolution is desired because it includes finer visual features. For describing region content more exactly, it is necessary to combine the information in all the different frequency subbands.

Referring to figure.3 and 4.b, for one segmented region LLR_i in the LL subband, we first find its corresponding regions in the same spatial location at the LH_3,HL_3,HH_3 subbands. Then utilizing spatial orientation tree structure[13, 14] described in figure 3, we localize their corresponding regions at the other high frequency subbands. As a result, referring to fig.4, for one semantic region R_i in the original image, all its corresponding sub-regions in the wavelet domain are achieved. Due to the wavelet decomposition property, all the sub-regions in the wavelet domain are different spatial-frequency resolution of region R_i in the original image. We can expect that combination of color and texture features of these sub-regions will describe R_i content more semantically.

In general, wavelet coefficients obtained by decomposing an image are not invariant with respect to translation, rotation and scaling of the image. In our method, we utilize the energy of wavelet coefficients to define region feature vector. The energies of wavelet coefficients in the high frequency subbands have proven effective for discerning texture in [12]. For every sub-region in the wavelet domain, its weighted sum of local energy in three color channels is computed, which is invariant with respect to image translation or rotation. Such weighted sum of local energy in three color channels can be expected to reflect the color-texture feature of this sub-region and regarded as one feature component which is ordered in the arrow scan direction shown in fig.4.b in our region feature vector. The local energy E_p of one sub-region is calculated from its three color local energy components, E_{pH}, E_{pS}, E_{pV} :

$$E_p = w_H \cdot E_{pH} + w_S \cdot E_{pS} + w_V \cdot E_{pV}, p = 1, 2, \cdots P \quad (2)$$

Where P is the total number of subbands with the wavelet decomposition level L, $P = 3 \cdot L + 1$. It also means that the region feature vector in our approach is P-dimensional. Since L is equal to 3 in this paper(mentioned in section 2), the region feature vector in our indexing algorithm is 10-dimensional. For three color components H, S, V, the hue shows more importance than the saturation and the value for human visual perception[15]. This can be achieved by setting the weight for the hue channel to a higher value than the other weights. In this paper, the weight proportion $1 : 2 : 1 (w_H : w_S : w_V)$ set in equation 2 emphasizes the importance of the hue.

The E_{pH} is calculated in equation 3. The E_{pS} and the E_{pV} are also defined as the similar ways.

$$E_{pH} = \sum_{m,n \in S} (C^H_{m,n})^2 \cdot K(2^{-l}(m_0 - m), 2^{-l}(n_0 - n)) \quad (3)$$

Here, $C^H_{m,n}$ is the H color component of wavelet coefficient $C_{m,n}$ in wavelet domain. S indicate the sub-region in wavelet subband. l indicates which decomposition level

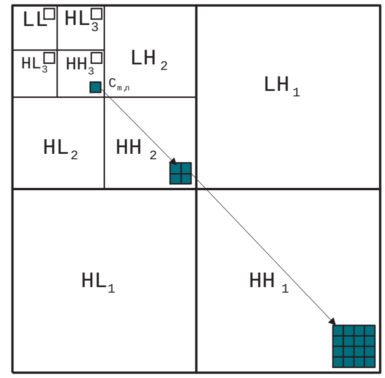

Figure 3: Spatial orientation tree in wavelet domain: for instance, a wavelet coefficient $C_{m,n}$ and all its descendents are shown in the colored parts. The arrow points from the subband of the parents to the subband of the children. In this figure, also shown is that for each point in the LL subband, its corresponding points in the similar spatial location at the HL_3, LH_3, HH_3 subbands are localized.

the sub-region is in. K is smoothing filters and experience has shown that the Gaussian filter is the desired choice [16]. m_0, n_0 denote the sub-region center. [16] also indicates that squaring in combination with a logarithmic nonlinearity normalizing is the best combination for the nonlinearity to describe texture feature. At last we obtain the representative region feature vector FV^{R_i} for one semantic region R_i in the original image as:

$$\begin{aligned} FV^{R_i} &= (fv_1^{R_i}, fv_2^{R_i}, \cdots, fv_P^{R_i}) \\ &= (\log E_1, \log E_2, \cdots, \log E_P) \end{aligned} \quad (4)$$

Our region feature vector definition has two advantages compared with other previous methods. First, such region feature vector can characterize the region content more exactly because it combines all spatial-frequency information of the region R_i, different from the previous methods that lost much semantic information during the submerging process that transfers the feature set of the region to one single feature vector. Secondly, the feature components in the region feature vector can be tuned semantically and effectively when doing region matching in the next section because they are consistent color-texture feature in wavelet domain.

4. REGION-BASED IMAGE SIMILARITY COMPARISON

The region matching is often measured using the distance function. When computing the distance between two region feature vectors, the proper weight adjustment for feature components in the feature vectors is important for region similarity measurement. We observe that, in some existing image retrieval systems, because each component in the region feature vector is a separated feature, for example, in a 6-dimensional region feature vector, the former three components may be three color values and the latter three may be texture features respectively, it is difficult to determine exactly how much more important one component is than

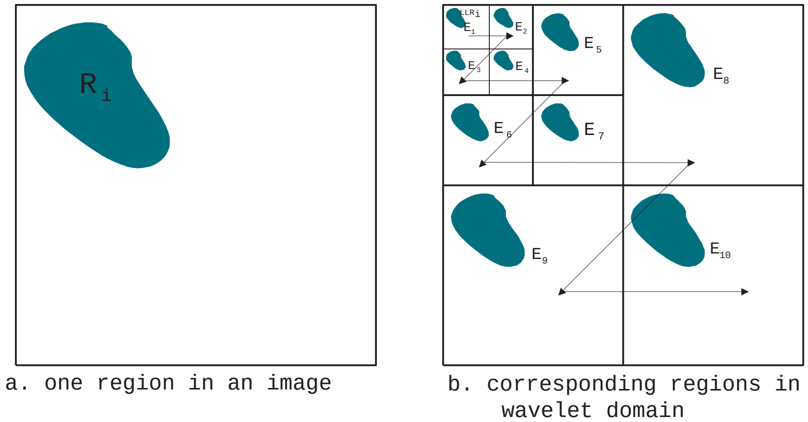

Figure 4: The illustration for the generation of hierarchical region feature vector. The feature vector of one region R_i in the original image is described using all sub-regions of region R_i in the wavelet domain. For each sub-region in the wavelet domain, its weighted sum of local energy in three color channels is computed as one color-texture feature component in region feature vector. At last all these feature components are ordered as the arrow scan direction and constituted as one region feature vector

Table 1: The weights used in the distance function

$w_{LL}:2$	$w_{HL_3}:1$	$w_{LH_3}:1$	$w_{HH_3}:1$	$w_{HL_2}:1$	$w_{LH_2}:1$	$w_{HH_2}:1$	$w_{HL_1}:1$	$w_{LH_1}:1$	$w_{HH_1}:1$
$W_3:3$				$W_2:3$			$W_1:1$		

another.

$$\begin{aligned} Dis(R_i,R_j) \quad = \quad & W_3[w_{LL}(fv_1^{R_i}-fv_1^{R_j})^2+ \\ & w_{HL_3}(fv_2^{R_i}-fv_2^{R_j})^2+ \\ & w_{LH_3}(fv_3^{R_i}-fv_3^{R_j})^2+ \\ & w_{HH_3}(fv_4^{R_i}-fv_4^{R_j})^2]+ \\ & W_2[w_{HL_2}(fv_5^{R_i}-fv_5^{R_j})^2+ \\ & w_{LH_2}(fv_6^{R_i}-fv_6^{R_j})^2+ \\ & w_{HH_2}(fv_7^{R_i}-fv_7^{R_j})^2]+ \\ & W_1[w_{HL_1}(fv_8^{R_i}-fv_8^{R_j})^2+ \\ & w_{LH_1}(fv_9^{R_i}-fv_9^{R_j})^2+ \\ & w_{HH_1}(fv_{10}^{R_i}-fv_{10}^{R_j})^2] \end{aligned} \tag{5}$$

In our image retrieval system, each feature component in region feature vector is a consistent color-texture feature presented by wavelet coefficients that characterize the region content in certain spatial-frequency resolution, therefore, the weight values in the distance function can be tuned more semantically and uniformly.

The distance function between two semantic regions R_i,R_j in this paper is defined as equation 5.

In this paper, referring to fig.3 and fig.4.b, 10-dimensional region feature vector fv_1, fv_2, $\cdots$, fv_{10} respectively correspond to the feature components in the LL,HL_3,LH_3,HH_3(in the third wavelet decomposition level), HL_2,LH_2,HH_2(in the second decomposition level) and HL_1,LH_1,HH_1 (in the first decomposition level) subbands. In [10], we have known that even if given unlimited viewing time, human will not scan all areas of a scene, but will instead attend to several perceptual interest regions which continually attract human attention. With this visual suggestion and wavelet transform property, those feature components in the second and the third wavelet decomposition levels will be given more importance than those in the first level, and the weight for the component in the LL subband will be set larger than the others. The intuition of such weight adjustment is that the main color and texture information of visual significant objects in image is focused on the relatively low frequency subbands. The weight proportion in our implementation is shown in table 1.

Having defined region matching method, we now need to measure the image similarity between two images (for example, the query image Q and one image I in the database) in our image indexing system. First, for each region $R_{i'}$ in image Q, we calculate the distance between it and all the regions in image I. Next, the overall image similarity is measured as a variation of the sum of the similarity between region pairs, which is defined in equation 6.

$$Sim(Q,I)=\sum_{i'\in Q,j'\in I}\lambda_{i',j'}\cdot e^{-Dis(R_{i'},R_{j'})} \tag{6}$$

Here $\lambda_{i',j'}$ is significance value, we consider the effect of region size for the image similarity measurement. The $\lambda_{i',j'}$ for one region pair $R_{i'}$ in query image Q and $R_{j'}$ in image I in the database is defined in equation 7.

$$\lambda_{i',j'}=\frac{size(R_{i'})+size(R_{j'})}{size(Q)+size(I)}\cdot\frac{size(R_{i'})/size(Q)}{size(R_{j'})/size(I)} \tag{7}$$

Here, $size$ is defined as the number of pixels inside the region. The first item takes into account the relative size of the regions with respect to the images they belong to, and the second item takes into account the similarity of size

Table 2: Precision-recall comparison result between image retrieval without boundary and image retrieval with boundary methods. For each semantic category, we do 10 queries in the database and then take the average precision of 10 queries as retrieval precision

The number of retrieved image number	Retrieval precision for six semantic image categories(%): the number in the left means precision for method A, and the right one indicates the result for method B					
	Bus	Africa people and villages	Buildings	Elephant	Flower	Mountains and glaciers
15	94.1/93.2	82.9/83.1	91.1/90.6	96/94.2	91.6/90.5	89.2/88
25	87.8/87	70.3/71.4	84.3/83.7	91.8/89.2	89.9/88.7	78.5/75.6
35	81.5/78.8	60.5/60.8	75.3/71.7	87.5/84.3	76.8/74.2	64.5/63.6
50	75/70.5	59.2/59	70.4/65.7	79.4/77.3	62.3/60.8	59.2/58.1

between two regions.

Then for the query image, the image indexing system measures the image similarity between it and each image in the database, at last sorts the images in descending order by the similarity values.

5. EXPERIMENT

Our image retrieval system is a query by example system. The user do image indexing by giving an example image and retrieve the top N images which are most similar to the query image in the image database. The experimental database includes 10,000 general-purpose photography color images with size 384×256 or 256×384, which are gathered from the COREL image resource. The proposed approach is implemented on PC with Pentium4/Windows2000, RAM 512M.

The first experiment is conducted to verify the effectiveness of deleting ambiguous boundary points in our image retrieval system. Two methods are compared. In both methods the image is segmented into regions in the LL subband and region feature vectors are extracted from all the frequency subbands. The difference is that method A detects boundary points and deletes them, but method B keeps the boundary.

A small sub-database consisting of 300 images of 6 semantic categories is considered in the first experiment, in which each category includes 50 images. We perform total 60 queries (10 queries for each semantic category) in the database to examine precision-recall. A retrieved image is regarded as a match if it belongs to the same category as the query image. For six semantic image categories, retrieval accuracy is computed as the percentage of matched images in the top 15, 25, 35, 50 retrieved images. We take the average value of 10 queries as retrieval accuracy for each category.

Table 2 summarizes the results. We can see that for most image categories method A improves the image retrieval result apparently by eliminating the ambiguous semantic interference of the boundaries. In the case of retrieving 50 images, the average retrieval accuracy of total 60 queries of method A is 2.4% higher than that of method B. For the image showing complex semantic content(e.g., Africa where too much details exist), both methods have the similar performance or even "image retrieval with edge" method performs slightly better. The possible reason is that some useful semantic information is also deleted when doing boundary deletion in complex semantic image. However, in all, by deleting boundaries, the retrieval precision can be improved.

In the second experiment, we want to proof that the information in high frequency subbands is crucial for capturing the region semantic information. Here we compare our image indexing system with the WBIIS (Wavelet-Based Image Indexing and Searching) system[17]. In the WBIIS, the image is first decomposed by a 4-level wavelet transform, then those wavelet coefficients in the four lowest frequency subbands(LL,HL_4,LH_4,HH_4) are stored as the feature vector for the image content description.

In order to provide numerical evaluation results, we test 90 images from 9 categories, each containing 10 of the images and use them as queries. Image categories are shown in table 3. Image matching is performed on the whole database of 10,000 images.

For each individual class, the retrieval accuracy is the average of matched images in the first 100 retrieved images, which is shown in figure.5. The numbers in x-axis correspond to the image categories in table 3.

In fig.5, we can see that since WBIIS utilizes the wavelet coefficients in low frequency subbands to generate the feature vector, for some relatively smooth images classes(e.g., weather, mountain), both the WBIIS and our method perform the image retrieval equally well. However, for the image categories with details crucial to semantics(e.g., flower, food which have fine details), our image indexing method outperforms WBIIS significantly. This is due to the fact that in WBIIS system much useful information in high frequency subbands that means important perceptual details is discarded, in our image retrieval system, a hierarchical region feature vector extracted from all the wavelet frequency subbands can capture finer image features.

We give one image retrieval comparison result in fig.6 and fig.7. For the limitation of space, we only show the top 12 retrieved images. The larger image "meat" on the left is the query image. In fig.6, we can see that except the fifth, eighth and eleventh images, the rest images are all in the same category "food" with the query image, especially, the first, fourth, sixth, ninth, tenth, twelfth images are perceptually similar to the query. But in fig.7, only four images are food. Irrelevant image retrieval in the WBIIS system is due to the lack of semantic detail information of "meat" in the feature vector used in the WBIIS.

In the third experiment, we evaluate the performance of the proposed region-based image retrieval approach through comparing it with a typical region-based indexing method-SIMPLIcity(Semantic-Sensitive Integrated Matching for Pic-

Table 3: Image categories in the second experiment

Image ID	Image Category
1	Flower
2	Elephant
3	Eagle
4	Food
5	Weather
6	Mountain and glaciers
7	Autumn
8	Africa people and village
9	Sports events

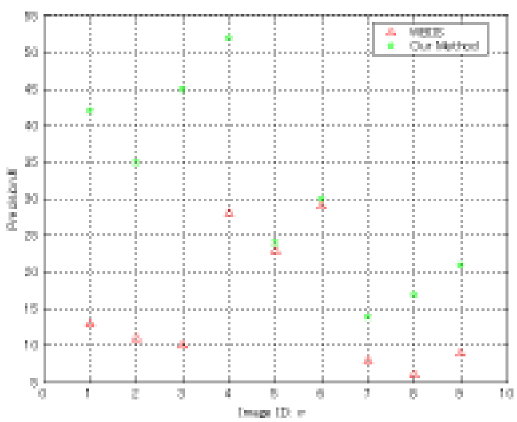

Figure 5: Image retrieval comparison between WBIIS and our image indexing system.The numbers in x-axis correspond to the image categories in table 3, the retrieval accuracy is the average of matched images in the first 100 **retrieved images**

Figure 6: Retrieval result in our image indexing system, the larger image "meat" on the left is the query image. For the limitation of space, only the top 12 retrieved images are shown.

Figure 7: Retrieval result in WBIIS indexing system

ture LIbraries)[4]. SIMPLIcity system does image segmentation in the original image after sub-sampling, then describes region content using 6-dimensional region feature vector (three for color and three for texture), which is the average value of the feature vector set in the region, at last gives one effective image similarity measurement method which incorporate the properties of all the segmented regions in the images.

For fair comparison, our experiment uses the same image database and precision computation method as those in SIMPLIcity. The image database is a subset of the COREL database, formed by 10 image categories, each consisting 100 images. Within this database, each image only belongs to one certain semantic category.

Every image in this database was tested as a query and the retrieval precision is calculated as the average percentage of images from the same category as the query image that are retrieved in the first 100 images. The performance comparison result is shown in table 4.

We can see that our system has demonstrated much improved retrieval accuracy over the SIMPLIcity in most semantic categories. For example, for the "Bus" query, in our system the retrieval rate is about 62%, but in SIMPLIcity it is only 36%. This means our system can retrieve more 26 "Bus" images in the first 100 retrieved images than the SIMPLIcity does. This is because that our method can segment image more effectively in the LL subband, describe the region content more exactly using the hierarchical feature vector and do more semantic-meaningful weight adjustment in the matching distance function. We can also notice that for Africa and Beach image categories, our method does a little worse performance than IRM. We notice that most images in these two classes do not have clear semantic objects but just include many excessively small regions.

It is also verified that our approach is robust to the viewpoint changes. In figure 8 and 9, we represent two query images and their retrieval results from the $10,000$ image database. It is shown that our system can retrieve images in which an identical object is located with various translations, referring to fig 8, and even an identical object is located in different kinds of backgrounds, referring to fig 9.

The robustness can be verified from the region feature vector in our retrieval system, which is obtained based on the local energy in the wavelet subbands. The problem caused by the sensitivity of wavelet coefficients on image translations can be alleviated in our system.

6. CONCLUSION

Table 4: The comparison of average precision for each image category between our approach and IRM. Every image in the database was tested as a query and the retrieval precision is calculated as the average percentage of images from the same category as the query image that are retrieved in the first 100 images.

Category	Average retrieval precision in IRM	Average retrieval precision in our method
Building	0.33	0.370
Buses	0.363	0.622
Dinosaurs	0.981	0.990
Elephant	0.400	0.435
Flowers	0.402	0.800
Horses	0.719	0.764
Moutains	0.342	0.387
Food	0.340	0.641
Africa	0.475	0.401
Beach	0.325	0.300

A novel region-based image retrieval approach using wavelet transform is presented in this paper. The main advantages of our method are summarized as:

1. Region segmentation is conducted in LL subband, which not only significantly lower the computation cost but also makes image segmentation more semantically.
2. Region feature vectors are represented hierarchically using all subbands in wavelet domain in three-color channels. It can describe the region color-texture content more exactly.
3. Because each feature component in the region feature vector is a combined color and texture feature for region content, our method tunes the weight values in the distance function more semantically and flexibly.

In the future work, we want to refine our image segmentation method for more semantic-meaningful description of image content. Other information of images such as shape and context will be considered to enhance our retrieval system. More exhaustive comparison of our image indexing system to other region-based systems should be also investigated.

7. ACKNOWLEGEMENT

This work was partially supported by Optical and Electronic Device Technology for Access Network Project of Grant-in-Aid for Center of Excellence (COE) research in the 21st Century sponsored by Japanese Ministry of Education, Culture, Sports, Science, and Technology.

Figure 8: Retrieval result in our image indexing system, the larger image on the left is the query image. We do the image indexing in the $10,000$ **image database. For the limitation of space, only the top 12 retrieved images are shown, we can pull the vertical scroll bar to view more images**

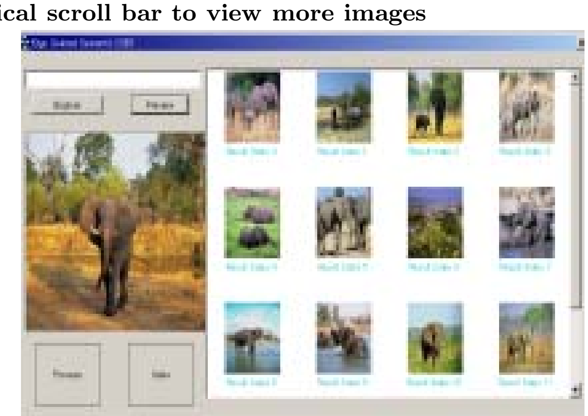

Figure 9: Retrieval result in our image indexing system

8. REFERENCES

[1] R.W.Picard, Content Access for Image/Video Coding: The Fourth Criterion, MIT Media Lab TR No.295,1994

[2] A.W.M.Smeulders, M.Worring, S.Santini, A.Gupta, R.Jain,Content-based image retrieval at the end of the early years, IEEE Trans. Pattern Analysis and Machine Intelligence, vol.22, no.12, pp.1349-1380, December 2000.

[3] Yixin Chen,James Z.Wang, A Region-Based Fuzzy Feature Matching Approch to Content-Based Image Retrieval, IEEE Trans. Pattern Analysis and Machine Intelligence, vol.24, no.9, pp.1252-1267, September 2002

[4] Jia Li,James Z.Wang,Gio Wiederhold, Intergrated Region Matching for Image Retrieval, Proc.ACM multimedia 2000, pp.147-156,Los Angeles,California,USA,Oct. 30-Nov. 3 2000

[5] Wei Wang,Yuqing Song,Aidong Zhang Semantics Retrieval by Content and Context of Image Regions, The 15th International Conference on Vision Interface , pp.17-22,Calgary,Canada ,May. 27-29 2002

[6] Hun-Woo Yoo,She-Hwan Jung,Dong-Sik Jang,Yoon-Kyoon Na, "Extraction of Major Object Features Using VQ Clustering for Content-based Image Retrieval",Pattern Recognition

[7] Feng Jing, Bo Zhang, Fuzong Lin, Wei-Ying Ma, Hong-Jiang Zhang, A Novel Region-Based Image

Retrieval Method Using Relevance Feedback, Proc. 3rd ACM Intl Workshop on Multimedia Information Retrieval (MIR), 2001

[8] G Qiu, Constraint Adaptive Segmentation for Color Image Coding and Content-based Retrieval, Proc. IEEE 4th Workshop on Multimedia Signal Processing, pp. 269-274,Canners,France,Oct.3-5 2001

[9] Charles K.Chui, An introduction to wavelets, Academic Press, San Diego,CA,1992.

[10] Wilfried Osberger, Anthony J.Maeder, Automatic Identification of Perceptually Important Regions in an Image, 14 IEEE Int. Conference on Pattern Recognition, 16-20 August 1998.

[11] T.Gevers, F.C.A.Groen, Segmentation of Color Images, Technical report, Faculty of Mathematics and Computer Science, University of Amsterdam,1990.

[12] Michael Unser, Texture Classification and Segmentation Using Wavelet Frames, IEEE Trans.Image Processing, vol.4, no.11, pp.1549-1560, November,1995

[13] J.M.Shapiro, Embedded image coding using zerotrees of wavelet coeffcients, IEEE Trans.Signal Processing,vol.41,no.12, pp.3445–3462, December 1993.

[14] AmirSaid, William A.Pearlman, A New Fast and Efficient Image Codec Based on Set Partitioning in Hierarchical Trees, IEEE Trans. Circuits and Systems for Video Technology, vol.6,no.3,pp243-250, June 1996.

[15] Foley,J.D.,vanDam,A.,Feiner,S.K.,Hughes,J.F., Fundamentals of Interactive Computer Graphics, Addison Wesley, Reading, MA, 1990

[16] Trtgve Randen, John H.H, Filtering for Texture Classification:A Comparative Study, IEEE Trans. Pattern Analysis and Machine Intelligence, vol.21, no.4, pp.291-310, April 1999

[17] J.Z.Wang, G.Wiederhold, O.Firschein, X.W.Sha, Content-based image indexing and searching using Daubechies' wavelets, International Journal on Digital Libraries, vol.1, no.4, pp.311-328,1998.