A Hybird Image Retrieval System with User's Relevance Feedback Using Neurocomputing

Dianhui Wang and Xiaohang Ma
Department of Computer Science and Computer Engineering
La Trobe University, Melbourne, VIC 3086, Australia
E-mail: csdhwang@ieee.org

Keywords: Image retrieval, intelligent systems, neurcomputing, relevance feedback

Received: May 17, 2004

This paper aims at developing a hybrid scheme for intelligent image retrieval using neural nets. Each item in an image database is indexed by a visual feature vector, which is extracted using color moments and discrete cosine transform coefficients. Query is characterized by a set of semantic labels, which are predefined by system designers and associated with domain concerns. The proposed hybrid image retrieval (HIR) system utilizes the image content features as the system input, and the semantic labels as its output. To compensate the deficiency of semantics modelling, an on-line user's relevance feedback is applied to improve the retrieval performance of the HIR system. The neural net acts like a pattern association memory bank that maps the low-level feature vectors to their corresponding semantic labels. During the retrieval process, the weights of the neural net are updated by an interactive user's relevance feedback technique, where the feedback signal comprise the neural net actual output, semantic labels provided by users and the given query. A prototype HIR system is implemented and evaluated using an artificial image database. Experimental results demonstrate that our proposed techniques are promising.

Povzetek: Hibridni algoritem z nevronsko mrežo je uporabljen za iskanje slik.

1 Introduction

With the development of the Internet and database techniques, information retrieval (IR) becomes very popular [1]. As a powerful form of delivering information, multimedia data are frequently used in many domain applications. Techniques for effectively dealing with multimedia databases management are useful and in demand. In the past, a lot of efforts on content-based image retrieval (CBIR) have been devoted to achieving this goal [2, 3]. A direct motivation for developing CBIR systems is to release the workload of manually annotating image data using text-based keywords. The existing CBIR systems can be categorized into two classes [4, 5]. The first scheme extracts low-level visual features from images, then uses a similarity measure to calculate the distance between a query and images from the database using the feature vectors for items rank. The second scheme is a semantic content-based approach, where semantics are automatically extracted from raw images, and then a construction key is made from these semantic items. The query is characterized using some combinations of the semantics extracted from the images, and the retrieval is achieved by counting the semantic items occurrence frequency. Currently, most CBIR systems fall into the first class, where semantic information of the image is not utilized during retrieval.

There exists a big gap between image semantic content and its corresponding representation using low-level visual features. This is one of the main reasons why the present CBIR systems cannot fully satisfy users' requirements. Users perceive images and measure their similarity using high-level semantic concepts which sometimes are hard to directly relate to low-level features. Even though there are many sophisticated algorithms to describe colour, shape and texture features, those visual features do not adequately model image semantics. However, because the low-level features can be extracted automatically and calculated efficiently, they are widely used in CBIR systems. To overcome the gap between the low-level visual features and the high-level semantics, pattern recognition or computer vision techniques can be used to extract semantics. To obtain high-level semantics, which is desirable in image retrieval, region information is not enough. Also, the automatic segmentation is not always reliable and time consuming. For some applications, object extraction can be ignored in design of CBIR systems. This is because the objective of the CBIR system is to retrieve some semantic relevant images from databases rather than to recognize objects from images. The underlying assumption is that semantically relevant images have similar visual characteristics or features. Consequently, the CBIR system may understand semantics within images by analysing those features instead of extracting object information. Therefore, the CBIR system does not need to understand images in the way as human beings do, but merely to assign images to semantic classes. Another remarkable difference between computer vision, pattern recognition systems and CBIR systems is that human is an indispensable part of the retrieval systems.

User's relevance feedback (RF) is a mechanism for enhancing retrieval performance of CBIR systems by using user's opinions on the retrieved items [7]. Generally, the RF techniques can be used to adjust parameters involved in CBIR systems so that the updated system may perform better in terms of some criteria. There are various ways to express and use the RF, for example, it can be encoded in binary form or discrete values to describe the degree of similarity between a retrieved item and a specific query. For more details and some new developments on RF techniques, readers may refer to [8]. At present, the adjustment of system parameters mainly takes place in similarity measure. Due to the capabilities of neural nets for pattern memory, generalization and adaptation, some promising results on learning similarity metrics using neural nets for CBIR systems have been developed [9, 10, 11, 12]. Indeed, the concept of learning similarity is closely related to visual feature classification [13]. Recently, it has been reformulated as a problem of pseudo metric approximation using neural nets [14]. In some existing neural nets based CBIR systems, the weights of neural nets are obtained through two phases: off-line training followed by on-line updating. These two steps correspond to the processes of pattern memory and neural similarity metric adaptation. Although some problems in this scheme still remain open, for example, the impact of the subjective RF on retrieval performance, the reported results indicate the usefulness of the RF techniques.

Neural nets, as a powerful modeling tool, have demonstrated its good potential for image retrieval tasks. It has been successfully applied in intelligent image retrieval systems, especially for semantics recognition and learning similarity measure. To further explore the power of neural nets based intelligent image retrieval systems, we present a hybrid scheme for image retrieval in this paper. Our proposed hybrid image retrieval (HIR) system takes low-level visual features as the system input and the semantic labels as its output. Off-line learning takes place before performing retrieval tasks. A modified cost function for "error back-propagation" training algorithm is presented to implement the RF, where feedback signal comprises the neural net actual output, semantic labels provided by users and the query. The remainder of this paper is organized as follows. Section 2 describes our hybrid intelligent image retrieval system in details. Section 3 evaluates the proposed techniques and reports our experimental results. Section 4 concludes this paper with some remarks on further work.

2 System Description

2.1 System Architecture

An intelligent image retrieval system may be viewed as a computing platform with a friendly user interface that allows users to represent, store and retrieve images from a given database. In addition, a good retrieval system should provide several modules to perform automatic feature extraction and selection, database update and user's interaction. Figure 1 shows the flow chart of our proposed HIR system.

The components and their functions in the HIR system are outlined below:

1. Feature Extraction Module - takes image pixel values as input and outputs the visual features. The feature extraction of the images should be done automatically or semi-automatically.

2. Database Module - All images and corresponding features data are stored here. Usually, it contains the following repositories:

   - Image Repository - consists of raw image data.
   - Feature Repository - holds the features that are extracted from image repository.
   - Links Repository - The connections between images and features are recorded in this repository.
   - Other Repository - saves the other accessorial data. For example, some parameters involved in learning and similarity measure, and the data used to accelerate the retrieval process.

3. Matching and Ranking Module - measures the similarity between the query image and images in the database and ranks the query results.

4. User Module - provides an interactive user interface to let users input the query and view the result. Although the user is generally thought of as a human agent, it is also possible that the module is an interface to another information system.

5. Interactive Feedback Module - The system can provide a mechanism to adjust the matching and ranking module. So, using the user feedback, the system can refine the retrieval results.

2.2 Features Extraction

Visual features, denoted by $F = [f_1, f_2, \dots, f_n]$, used in image retrieval systems are crucial because they directly affect the system performance. Although there are various criteria and techniques available in literature, so far, it is still hard to tell which feature is necessary and sufficient to result in a high performance retrieval system. In our HIR system, we adopt the RGB color model, calculate the color moments and some local statistics of the discrete cosine transform (DCT) coefficients of the images to construct the feature vectors. It is well-known that the DCT coefficients corresponding to lower frequencies contribute more information to an image than those ones associated with higher

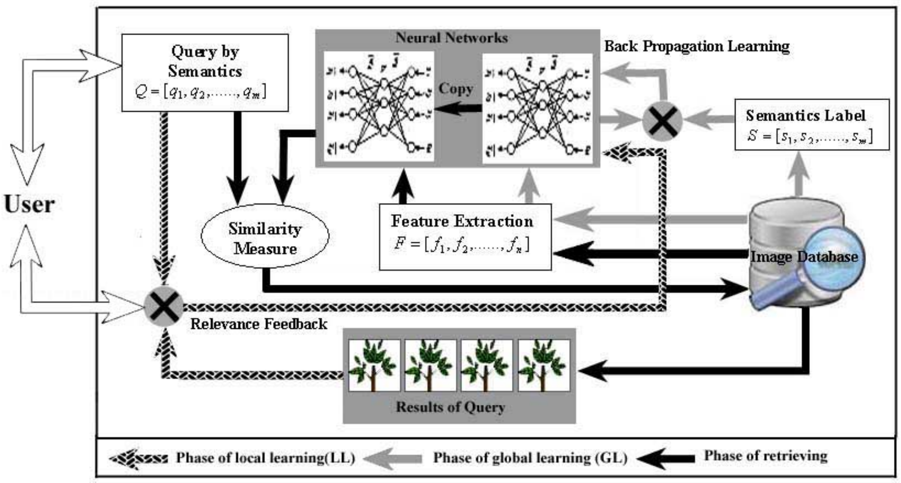


Figure 1: The proposed intelligent HIR system using neural nets

frequencies. Considering a tradeoff between the dimensionality curse and fidelity of information, we apply a partition technique as shown by an example in Figure 2 where the coefficients are grouped into 7 categories based on location attribute, for computing the local statistics of the DCT coefficients.

Semantics within an image can be extracted manually, and we use a semantic label vector, $S = [s_1, s_2, \ldots, s_m]$, to represent the presence or absence of semantics within the image, where m is the number of semantics concerned and predefined by domain experts, and s_i takes binary values. For example, $S = [1, 0, 1]$ can be interpreted as that the image contains semantics 1 and 3, but it does not contain semantics 2.

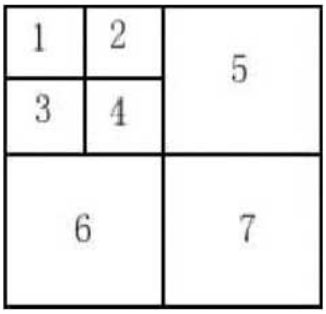


Figure 2: DCT transformation

2.3 Off-line Semantics Modelling

The purpose of off-line semantics modeling is to associate the low-level visual features with the semantic concepts contained in images. A feedforward neural net is employed to implement this task because of its power of learning, generalization and adaption [15].

Let $G = \{(F, S)\}$ be a collection of feature-semantic-label pairs. The neural net used in this study is with three layer architecture, i.e., input layer, hidden layer and output layer. Sigmoid activation function, $\sigma(x) = [1 + \exp(-x)]^{-1}$, is used at both the hidden layer and the output layer. In order to improve the generalization capability of the neural net, a regularized hybrid training algorithm is adopted [16]. The objective function in this algorithm is given by

$$\begin{aligned} E_1 = \textstyle\sum_G \|\sigma(\sigma(Fw_N)w_L) - S\|^2 \\ +\lambda Tr(w_N^T w_N + w_L w_L^T), \end{aligned} \tag{1}$$

where w_N is the hidden layer weights; w_L is the output layer weights; λ is a regularizing parameter; and $Tr(M)$ represents the trace of a matrix M.

2.4 Query Representation and Similarity Measurement

Distinguishing from other neural net based image retrieval systems, the query in our proposed system is specified by an indicator vector $Q = [q_1, q_2, \ldots q_m]$, where q_k takes binary values with 1(0) representing the presence (**don't care**) of a semantic concept contained in the target images. During the retrieval process, the low-level features are fed into the well-trained neural net, and it gives real number outputs in [0,1], denoted by $O = [o_1, o_2, \ldots, o_m]$. We define a similarity measure $D(Q, O)$ by a weighted dot product (DP), that is,

$$D(Q, O) = \sum_{k=1}^{m} \alpha_k q_k o_k, \tag{2}$$

where $\alpha_k \geq 0$, subjected to $\sum \alpha_k = 1$, is a weighting factor which reflects the emphases on different semantics related to a specific query. In our simulations, we take these factors equally.

2.5 On-line Memory Bank Updating with User's Relevance Feedback

One of the important characteristics of the CBIR systems is that human is an indispensable part of the systems. In the HIR system, the RF is applied for refreshing the pattern association memory so that an improved recognition rate for relevant semantic images may be achieved. In such a way, the synapse between visual features and corresponding semantic labels may be reconstructed by the items retrieved.

From each retrieved image, the user can assign a feedback vector, denoted by $U = [u_1, u_2, \ldots, u_m]$, where u_k takes binary values with 1(0) representing presence (absence or **don't care**) of the semantics. The feedback vector U and the query vector Q may not be identical. This mismatching can be caused by various reasons, such as insufficient training time, inappropriate low-level features used or limited generalization capability of the neural net model. Generally, there are four cases:

1. As $q_k = u_k = 0$, the user "do not care" this semantic item. The retrieved image does not contain the corresponding semantic item or it contains this item but the user does not mark it. For this case, the updating of the weights associated to the *k*-th output of the neural net is irrelevant to further improve the retrieval performance.

2. As $q_k = 0,\ u_k = 1$, the user "do not care" this semantic item, but the retrieved image contains this item and the user also marks it in the feedback vector. It is like a byproduct to supervise the neural net to refine its memory.

3. As $q_k = 1,\ u_k = 0$, this semantic item is what the user expects. Unfortunately, the retrieved image does not contain it.

4. As $q_k = 1,\ u_k = 1$, this semantic item is what the user expects and the retrieved image satisfies the user's requirement.

Summarizing the above four cases, the on-line learning will take place only for the weights associated to the outputs with $q_k \vee u_k = 1$, where "$\vee$" represents the logic "OR" operator. In such a way, the objective function (1) for online learning is modified as:

$$E_2 = \sum_{G'} Z\Theta Z^T + \lambda Tr(w_N^T w_N + w_L w_L^T), \qquad (3)$$

where $Z = \sigma(\sigma(F'w_N)w_L) - U; G' = \{(F', U)\}$ represents a collection of feature-semantic-lable pairs from the retrieved images; $\Theta = diag\{q_1 \vee u_1, q_2 \vee u_2, \ldots, q_m \vee u_m\}$.

3 Performance Evaluation

3.1 Experimental Setup

The proposed HIR system has been implemented using C++ and evaluated by an artificial image database with 355 nature scene images containing following semantic concepts: Rock, Water, Tree, Sky and Human. Table 1 and Table 2 show some basic statistics of the database. For example, in Table 1 , there are 165 images containing rocks; in Table 2 , there are 125 images containing 2 semantic concepts.

Table 1: Database Statistics I

Semantic	Image Number
Rock	165
Water	144
Tree	220
Sky	173
Human	80

Table 2: Database Statistics II

Semantics Number	Image Number
1 Semantic	70
2 Semantics	125
3 Semantics	99
4 Semantics	50
5 Semantics	11

For evaluation purpose, each image was transformed into 9 different images. The transformation detail is given in Table 3. So totally there are 3,550 images in the testing database. Because a single train-and-test experiment may generate misleading performance estimates when the sample size is relatively small, a 5-fold cross validation scheme was used to evaluate the performance of the retrieval system. For each fold, we partition the database into training dataset (1,775 images) and test dataset (1,775 images) randomly.

Table 3: Transformation Methods

	Method	Parameters
1	Resize	0.8
2	Resize	1.2
3	Rotation (Clockwise)	90°
4	Rotation (Clockwise)	180°
5	Rotation (Clockwise)	270°
6	Salt-Pepper Noise	0.03
7	Gaussian Noise	$\mu = 0, \sigma = 0.06$
8	Vertical Mirror	
9	Horizontal Mirror	

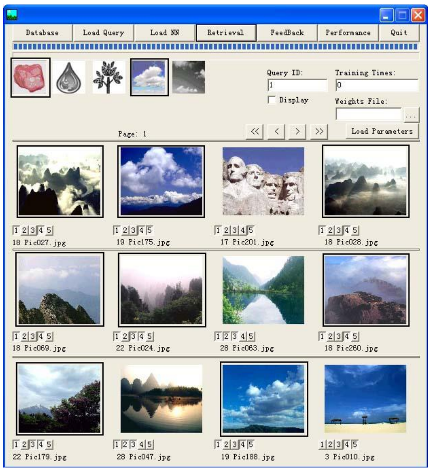
The visual features used in our system are comprised of 9 color moments and 42 local statistics of the DCT coefficients for three-color channels, i.e., the mean and the variance of the DCT coefficients in the 7 sub-regions for three color channels. A neural net with an architecture 51-30-5 is employed in the HIR system. The training program runs 10,000 epochs without the momentum term and with a learning rate as 0.1 for the off-line learning.

There are two statistical measures commonly used in IR: *recall* and *precision*. *Recall* is the ratio between the number of retrieved relevant images and the total number of retrieved images, given a certain window size for simulations. *Precision* is the ratio between the number of retrieved relevant images and the number of relevant images in the database. A higher value of *precision* therefore indicates that the top-ranked hits more target images. Because of the database used in this study, the standard *recall* and *precision* calculation formulas cannot be directly applied to characterize the system performance. Therefore, we used a modified *recall* calculation formula where a variable window size is used for each group of images. The window size takes the number of images having the same semantics as the query. Another measure adopted in this evaluation is the so-called mean rank μ, which is defined by

$$\mu = \frac{N(N+1)}{2\sum_{i=1}^{N} rank(i)} \tag{4}$$

where $rank(i)$ is the rank of the relevant image i (i.e., position of retrieved relevant image i in the retrieved images), which is an integer between 1 and 1,775 in this case study, and N is the number of total relevant images in the test database.

3.2 Results and Analysis



Query ID: 1

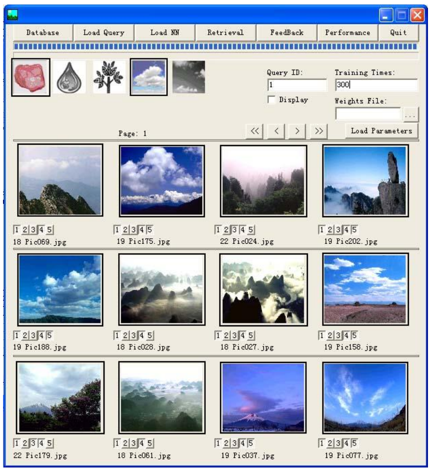
□ Display

Training Times: 0

Weights File:

Figure 3: Performance without any RF

The results shown in this section for a certain number of semantic concepts are the average rates for all possible combination, for example, 2 S (2 semantic concepts) is the average performance for all 2 semantic concepts combination, including $< Rock, Tree >$, $< Rock, Water >$ and etc. During the retrieval process, 5 times of RF were applied using the variable windows for each group images, and the neural net was trained for 300 epochs without the momentum term and with a learning rate as 0.1 for online updating. Figures 3 and 4 show the system performance for a specific query utilizing two semantic concepts: "Rock" and "Cloud". Tables 4 and 5 show *recall* and μ performances of the proposed HIR system for the training datasets. It indicates the semantics modelling power of the neural net. Tables 6 and 7 show *recall* and μ performances for the test datasets with various times of RF. It can be seen that the *recall* performance has been gradually enhanced through the use of the RFs. It is observed that the average *recall* rates for queries with different semantics monotonically decrease with the increased number of semantics in the images. This sounds logical and reasonable because adding complexity will result in some loss of system performance. For μ, there is little improvement as the RF is applied. The main reason for this is of the lack of positive examples from the retrieved items for the RF. The feedback activity only happens for the first N retrieved images. The relevant images located in the rear of retrieval queue is the "dead zone" and can not be activated for the RF. Therefore, it is a significant technique to inspire some relevant images from the rear of retrieval queue in the database for on-line learning.



Query ID: 1

□ Display

Training Times: 300

Weights File:

Figure 4: Performance after 5 times of RF

Table 4: The *Recall* performance for the training datasets (%)

recall	**1 S**	**2 S**	**3 S**	**4 S**	**5 S**
Fold 1	96.58	92.18	89.02	87.27	86.67
Fold 2	96.15	91.83	87.59	79.25	66.67
Fold 3	96.42	91.32	85.87	79.19	66.67
Fold 4	96.66	92.85	91.52	90.88	86.67
Fold 5	94.54	88.65	82.61	78.67	80.00
Avg.	96.07	91.36	87.32	83.05	77.33

Table 5: The μ performance for the training datasets

μ	**1 S**	**2 S**	**3 S**	**4 S**	**5 S**
Fold 1	0.931	0.764	0.635	0.414	0.113
Fold 2	0.923	0.735	0.568	0.368	0.084
Fold 3	0.935	0.759	0.588	0.366	0.086
Fold 4	0.943	0.787	0.697	0.633	0.388
Fold 5	0.912	0.732	0.577	0.464	0.342
Avg.	0.929	0.755	0.613	0.449	0.203

We investigate the impact of the number of feedback images on the system performance. Tables 8 and 9 show results of *recall* and μ with different RF window sizes, respectively. The reported figures are the average values from

Table 6: The *Recall* performance vs. relevance feedback times (%)

Recall	**1 S**	**2 S**	**3 S**	**4 S**	**5 S**
FB 0	80.79	64.69	53.10	45.11	34.67
FB 1	79.54	67.05	59.50	53.51	42.67
FB 2	81.15	69.59	61.63	56.24	50.67
FB 3	82.07	70.32	62.90	57.77	53.33
FB 4	82.97	71.52	63.91	58.90	53.33
FB 5	83.40	72.03	65.45	59.86	54.67
Avg.	81.65	69.20	61.08	55.23	48.22

Table 7: The μ performance vs. relevance feedback times

μ	**1 S**	**2 S**	**3 S**	**4 S**	**5 S**
FB 0	0.780	0.454	0.250	0.125	0.033
FB 1	0.757	0.459	0.258	0.133	0.037
FB 2	0.773	0.473	0.271	0.136	0.038
FB 3	0.779	0.473	0.269	0.139	0.037
FB 4	0.786	0.486	0.280	0.140	0.038
FB 5	0.789	0.487	0.279	0.145	0.038
Avg.	0.777	0.472	0.268	0.136	0.037

*FB m - m times of relevance feedback

1 to 5 times of RF, which are believed as being objective. The *recall* and μ performance increase obviously when the feedback window size varies from 1 to 2. However, when the window size keeps increasing, the system performances tend to be flat. We also observed that the feedback window size has no obvious impact on 5 semantic concepts. This could be related to the limited number of images to be retrieved in the database.

Table 8: Average *Recall* performance vs. the number of feedback images (%)

Recall	**1 S**	**2 S**	**3 S**	**4 S**	**5 S**
WSR 1	81.65	69.20	61.08	55.23	48.22
WSR 2	88.10	77.54	67.42	59.13	48.22
WSR 3	87.95	78.07	68.64	58.93	48.22
WSR 4	87.74	78.36	69.66	59.77	47.33
WSR 5	87.81	78.51	69.12	61.21	48.22

3.3 A Comparative Study

In this comparative study, we investigated the effect of the features and similarity metrics used in the HIR system. Three different feature sets, that is, the colour moment (CM) features, DCT features and the mixed features, are examined under the HIR system framework, respectively. The same datasets are used and the performance results are obtained by averaging a 5-fold runs. Uisng the colour moment features, a 9-dimension feature vector is extracted from each image. Correspondingly, a neural net with an architecture of 9-20-5 is employed. For the DCT features, a neural net with an architecture of 42-30-5 is employed, since 42 local statistics are extracted from the coefficients of the DCT in the 7 sub-regions. The neural nets were trained off-line for 10,000 epochs with learning rate as 0.1.

Table 9: Average μ performance vs. the numbers of feedback images

μ	**1 S**	**2 S**	**3 S**	**4 S**	**5 S**
WSR 1	0.777	0.472	0.268	0.136	0.037
WSR 2	0.819	0.516	0.299	0.145	0.037
WSR 3	0.818	0.528	0.304	0.148	0.037
WSR 4	0.814	0.538	0.316	0.155	0.035
WSR 5	0.813	0.544	0.319	0.166	0.037

*WSR m - Ratio between the number of images used for RF and the number of images the number of images contained in the groups with the same number of semantics

Table 10: *Recall* performance comparison for different features using the training datasets (%)

recall	**1 S**	**2 S**	**3 S**	**4 S**	**5 S**
Mixed	96.07	91.36	87.32	83.05	77.33
CM	67.39	30.53	17.47	8.29	4.00
DCT	66.47	21.12	8.73	4.53	0.00

Tables 10 and 11 show the system performances for the training datasets with the different features. Tables 12 and 13 show the system performances for the test datasets with the different features. It is remarkable that the system performance using the mixed features is much better than that obtained by the separated ones.

The system performances produced by the DP similarity measure and the $L_p(p = 1, 2, \infty)$ norms are compared. Notice that the "0" elements in a query do not means "absence" but "do'not care", therefore a non-standard L_p norm is applied, that is,

$$D_{L_p}(Q, O) = (\sum_{k=1}^{m} q_k |q_k - o_k|^p)^{1/p}, \tag{5}$$

where $Q =< q_1, q_2, \ldots, q_m >$ and $O =< o_1, o_2, \ldots, o_m >$ are the query indicator vector and the neural net's output, respectively.

Table 11: μ performance comparison for different features using the training datasets

μ	**1 S**	**2 S**	**3 S**	**4 S**	**5 S**
Mixed	0.929	0.755	0.613	0.449	0.203
CM	0.652	0.270	0.115	0.049	0.013
DCT	0.630	0.233	0.085	0.033	0.007

Table 12: *Recall* performance comparison for different features using the test datasets (%)

recall	**1 S**	**2 S**	**3 S**	**4 S**	**5 S**
Mixed	80.79	64.69	53.10	45.11	34.67
CM	67.21	30.65	18.80	8.61	1.33
DCT	63.70	19.93	7.16	3.13	0.00

Table 13: μ performance comparison for different features using the test datasets

μ	**1 S**	**2 S**	**3 S**	**4 S**	**5 S**
Mixed	0.780	0.454	0.250	0.125	0.033
CM	0.651	0.270	0.115	0.049	0.012
DCT	0.596	0.222	0.083	0.033	0.008

Tables 14 and 15 show the retrieval performances for the training datasets with the different similarity metrics. Tables 16 and 17 report the results for the test datasets. It can be seen that the system performances obtained by the $Dot\ product$, L_2 and L_∞ norms are comparable, whereas the results from the L_1 norm is poor compared with others.

3.4 Robustness Analysis

Model reliability or robustness with respect to the model parameters shift is meaningful. A higher reliability implies a relaxed requirement to the solution constraints. Conversely, if the model reliability is weak, then the variation scope of the parameters becomes limited. This makes the process of achieving a feasible solution more complicated or difficult. For neural nets, the model parameters are the weights, the solution refers to a set of specified weights obtained through learning, and the constraints may be the learning rate and/or the terminal conditions. To investigate the HIR system reliability, we generate a random matrix, namely, M_{noise}, whose size equals the weight matrix and its elements are uniformly distributed in $(-1, 1)$. Then, perturbed weight matrices can be obtained by $W_{noise} = (I + \delta M_{noise}). * W$ at 10 different levels, i.e., the δ varies from 1% to 10%. Figures 5 and 6 show the effects of the model noise to the memory bank for the training datasets. A comparative study on the functionality of the RF to different levels of noise was investigated. Three times of RF were applied for the queries with different number of semantic concpets. Figures 7 and 8 depict the performances for 2 semantic concepts. They show that the HIR system is more robust to the model noise as the RF technique is employed.

Table 14: *Recall* performance comparison for the training datasets (%)

recall	**1 S**	**2 S**	**3 S**	**4 S**	**5 S**
DP	96.07	91.36	87.32	83.05	77.33
L_1	61.17	40.61	43.37	61.00	77.33
L_2	96.07	91.38	87.38	83.31	77.33
L_∞	96.07	91.39	87.39	83.31	77.33

Table 15: μ performance comparison for the training datasets

μ	**1 S**	**2 S**	**3 S**	**4 S**	**5 S**
DP	0.929	0.755	0.613	0.449	0.203
L_1	0.598	0.341	0.283	0.332	0.203
L_2	0.929	0.758	0.618	0.4455	0.192
L_∞	0.929	0.722	0.495	0.253	0.046

Table 16: *Recall* performance comparison for the test datasets (%)

recall	**1 S**	**2 S**	**3 S**	**4 S**	**5 S**
DP	80.79	64.69	53.10	45.11	34.67
L_1	56.93	34.42	30.04	36.61	34.67
L_2	80.79	64.70	52.99	44.85	34.67
L_∞	80.79	64.68	52.95	44.93	34.67

4 Concluding Remarks

A hybrid scheme for content-based intelligent image retrieval is proposed in this paper. Our main technical contributions are (i) the framework of a new intelligent image retrieval scheme using neural nets; (ii) the modified objective function for on-line memory bank updating using user's relevance feedback and query information; and (iii) the robustness analysis and comparative studies. Simulation results demonstrate that the interactive relevance feedback with on-line learning strategy could enhance the recall performance in the HIR system. However, it is quite limited for improving the μ performance. This may be largely caused by the lack of suitable teacher signals (images) during the feedback learning process, and/or the scale constraint of our image database used in this study. It is believed that the μ performance of the HIR system will be increased by using some typical images from "dead zone", i.e., the set of images in the database whose elements have no chance to be retrieved for some specific queries.

It is interesting to see the effects of false relevance feed-

Table 17: μ performance comparison for the test datasets

μ	**1 S**	**2 S**	**3 S**	**4 S**	**5 S**
DP	0.780	0.454	0.250	0.125	0.033
L_1	0.544	0.266	0.161	0.116	0.033
L_2	0.780	0.462	0.257	0.128	0.033
L_∞	0.780	0.464	0.250	0.125	0.034

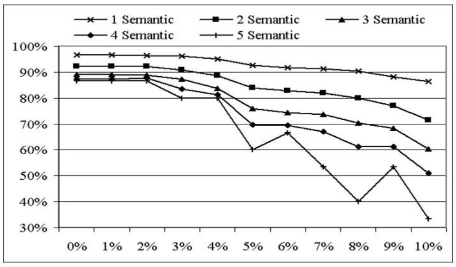


Figure 5: $recall$ performance vs. 10 noise levels for the training datasets

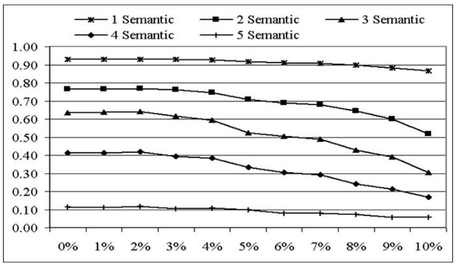


Figure 6: μ performance vs. 10 noise levels for the training datasets

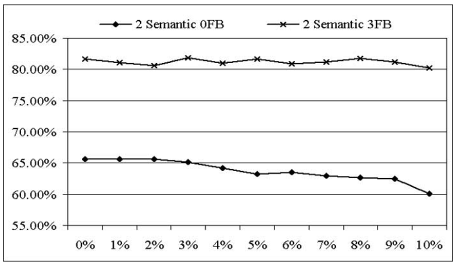


Figure 7: $recall$ performance vs. 10 noise levels for the test datasets with 2 semantics

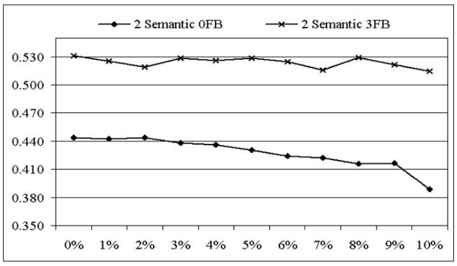


Figure 8: μ performance vs. 10 noise levels for the test datasets with 2 semantics

back information on the retrieval performances. Also, it is critical to resolve the "dead zone" problem for retrieval systems. A study on the use of the lower-bounding lemma [1] in the HIR system for speeding up the retrieval process will be very necessary. Finally, further developments of the HIR system in both theoretical aspects and real world practices are being expected.

References

[1] B. Y. Ricardo and R. N. Berthier (1999) *Modern Information Retrieval*, ACM Press, Addison-Wesley.

[2] J. Wu (1997) Content-based indexing of multimedia databases, *IEEE Trans. On Knowledge and Data Engineering*, vol. 6, pp. 978–989.

[3] Y. Rui, T. S. Huang, and S. F. Chang (1999) Image retrieval: Current techniques and promising directions and open issues, *Journal of Visual Communication and Image Representation*, vol. 10, pp. 39–62.

[4] F. Crestani and G. Pasi (2000) *Soft Computing in Information Retrieval: techniques and applications*, Physica Verlag (Springer Verlag).

[5] A. Yoshitaka and T. Ichikawa (1999) A survey on content-based retrieval for multimedia databases, *IEEE Trans. On Knowledge and Data Engineering*, vol. 11, pp. 81-93.

[6] S. Santin, and R. Jain (1999) Similarity measures, *IEEE Trans. On Pattern Analysis and Machine Intelligence*, vol. 21, pp. 871–883.

[7] Y. Riu, T. Hunag, M. Ortega, and S. Mehrotra (1998) Relevance feedback: A power tool for interactive content-based image retrieval, *IEEE Trans. On Circuit and Systems. for Video Technology*, vol. 5, pp. 644–656.

[8] X. S. Zhou and T. S. Huang (2002) Relevance feedback in content-based image retrieval: some recent advances, *Information Sciences-Applications-An International Journal*, vol. 148, pp.129–137.

[9] H. Lee and S. Yoo (2001) Intelligent image retrieval using neural network, *IEICE Trans. on Information and Systems*, vol. 12, pp. 1810–1819.

[10] T. Ikeda and M. Hagiwara (2000) Content-based image retrieval system using neural networks, *International Journal of Neural Systems*, vol. 5, pp. 417–424.

[11] J. H. Lim, J. K. Wu, S. Singh, and D. Narasimhalu (2001) Learning similarity matching in multimedia content-based retrieval, *IEEE Trans. On Knowledge and Data Engineering*, vol. 13, pp. 846–850.

[12] G. D. Guo, A. K. Jain, W. Y. Ma, and H. J. Zhang (2002) Learning similarity measure for natural image retrieval with relevance feedback, *IEEE Trans. On Neural Networks*, vol. 13, pp. 811–820.

[13] V. Aditya, A. T. Mario,K. J. Anil, and H. J. Zhang (2001) Image classification for content-based indexing, *IEEE Trans. On Image Processing*, vol. 10, pp. 117–130.

[14] D. Wang and X. Ma (2004) Learning pseudo metric for multimedia data classification and retrieval, *Proc. Knowledge-Based Intelligent Information & Engineering Systems, LNAI 3213*, vol. 1, pp. 1051–1057.

[15] D. E. Rumelhart, G. E. Hinton and R. J. Willianms (1986) Learning representations of back-propagation errors, *Nature*, vol. 323, pp. 533–536.

[16] S. McLoone and G. Irwin (2001) Improving Neural Network Training Solutions Using Regularisation, *Neurocomputing*, vol. 37, pp. 71–90.