ORIGINAL ARTICLE

A high accuracy land use/cover retrieval system

Alaa Hefnawy

Computer & Systems Dept., Electronics Research Institute, Cairo, Egypt

Abstract The effects of spatial resolution on the accuracy of mapping land use/cover types have received increasing attention as a large number of multi-scale earth observation data become available. Although many methods of semi automated image classification of remotely sensed data have been established for improving the accuracy of land use/cover classification during the past 40 years, most of them were employed in single-resolution image classification, which led to unsatisfactory results. In this paper, we propose a multi-resolution fast adaptive content-based retrieval system of satellite images. Through our proposed system, we apply a Super Resolution technique for the Landsat-TM images to have a high resolution dataset. The human–computer interactive system is based on modified radial basis function for retrieval of satellite database images. We apply the backpropagation supervised artificial neural network classifier for both the multi and single resolution datasets. The results show significant improved land use/cover classification accuracy for the multi-resolution approach compared with those from single-resolution approach.

1. Introduction

One of the fundamental characteristics of a remotely sensed image is its spatial (*x–y* domain) resolution; as the basic information contained in the image is strongly dependent on spatial resolution [1]. Improper choice of different spatial resolution can lead to misleading interpretation, e.g. in a Landsat Multi-Spectral Scanner image, the urban residential environment is sensed as a relatively homogeneous entity. However, when observed at finer resolution, the residential area is mostly made of individual houses, roads and plants. With the development of new remote sensing systems, very-high spatial resolution images provide a set of continuous samples of the earth surface from local, to regional scales. The spatial resolution of various satellite sensors ranges from 0.5 to 25,000 m now. Furthermore, high resolution airborne data acquisition technology has developed rapidly in recent years. As an increasing number of high resolution data sets become available, there is an increasing need for more efficient approaches to store, process, and analyze these data sets. The development of efficient analysis methods of using these multiscale data to improve land

use/cover mapping and linking thematic maps generated from high resolution to coarse resolution has become a challenge [2,3]. Several techniques have been employed to assess appropriate (or optimal) spatial resolutions. Although a particular classification can achieve the best result from a single resolution appropriate to the class, there is no single resolution which would give the best results from all classes [4]. Landscape objects (e.g. land cover/use polygons) are not the same size and vary in different structures. Some objects are better classified at finer resolutions while others require coarser resolutions. Therefore, as suggested by Ref. [1], various objects require different analysis scales according to the image scene model. Scene models may be either high (H) resolution with pixels smaller than objects, or low (L) resolution with pixels larger than objects to be mapped. From a practical standpoint, building a framework to represent, analyze and classify images represented by multiple resolutions is necessary in order to capture unique information about mapped classes that vary as a function of scale. Many previous studies show the importance of developing and evaluating spatial analytic methods and models to support multiscale databases [5–7].

The objective of this paper is to build a high accuracy content-based retrieval system of satellite images based on multiscale dataset. The used human–computer interactive system is based on relevance feedback. A large database of remotely sensed data has been used, which consists of 300 Landsat-7 TM satellite images scenes that cover different areas in Egypt and show land use/land cover [8]. By applying the Super Resolution (SR) techniques on this low-resolution Landsat TM dataset, a new high-resolution dataset has been restored. An improvement of the system accuracy has been achieved by applying the backpropagation supervised artificial neural network classifier for both the low and high resolution datasets.

In the next section we will give a brief description of the SR restoration technique used for creating the high resolution dataset. The proposed system will be presented in Section 3. In Section 4 we will demonstrate the used material and methodology. The classification results are shown in Section 5, and finally discussion and conclusions are given in Sections 6 and 7.

2. High resolution dataset

In general, multi-resolution images can be created in two ways: (1) by integrating different resolution images acquired by different sensors; and (2) aggregating fine resolution images into different coarse resolution levels (i.e., image pyramids). Obtaining images of different resolutions from different sensors could have advantage of including more spectral information that can be used to identify different objects, but is expensive. The miss-registration between different images also would increase the processing cost and reduce classification accuracy. It is more efficient to extract spatial information over a range of resolutions from a single high resolution image.

We will use in this paper, only two resolution levels datasets. First one is the low resolution Landsat-7 TM satellite images of different regions of Egypt, acquired on 6 May 1998, and 21 June 2001. Then we construct the second one (high resolution) by applying a SR technique on this dataset.

Super Resolution are techniques that in some way enhance the resolution of an imaging system. These SR-techniques break the diffraction-limit of the digital imaging sensor. There are both single-frame and multiple-frame variants of SR, where multiple-frame are the most useful. The basic idea behind Super-Resolution is the fusion of a sequence of low-resolution noisy blurred images to produce a higher resolution image or sequence. The information that was gained in the SR-image was embedded in the LR images in the form of aliasing. That is, LR images are sub-sampled (aliased) as well as shifted with sub-pixel precision. If the LR images are shifted by integer units, then each image contains the same information, and thus there is no new information that can be used to reconstruct an HR image. If the LR images have different sub-pixel shifts from each other and if aliasing is present, however, then each image cannot be obtained from the others. In this case, the new information contained in each LR image can be exploited to obtain an HR image.

Generally to obtain different looks at the same scene, some relative scene motions must exist from frame to frame via multiple scenes or video sequences. Multiple scenes can be obtained from one camera with several captures or from multiple cameras located in different positions. These scene motions can occur due to the controlled motions in imaging systems, e.g., images acquired from orbiting satellites. The same is true of uncontrolled motions, e.g., movement of local objects or vibrating imaging systems. If these scene motions are known or can be estimated within sub-pixel accuracy, and if we combine these LR images, SR image reconstruction is possible [9,10].

The first step to comprehensively analyze the SR image reconstruction problem is to formulate an observation model that relates the original HR image to the observed LR images as follows

$$y_k = DB_kM_kX + n_k \quad \text{for } 1 \leqslant k \leqslant z \quad (1)$$

where X is the desired HR image and yk are the z LR images, M_k is a warp matrix of size $L_1N_1L_2N_2 \times L_1N_1L_2N_2$, B_k represents a $L_1N_1L_2N_2 \times L_1N_1L_2N_2$ blur matrix, D is a $(N_1N_2)^2 \times L_1N_1L_2N_2$ subsampling matrix, and n_k represents a lexicographically ordered noise vector.

Most of the SR image reconstruction methods proposed in the literature consists of the three stages illustrated in Fig. 1: registration, interpolation, and restoration (i.e., inverse procedure). These steps can be implemented separately or simultaneously according to the reconstruction methods adopted. The estimation of motion information is referred to as registration, and it is extensively studied in various fields of image processing. In the registration stage, the relative shifts between LR images compared to the reference LR image are estimated with fractional pixel accuracy. Obviously, accurate subpixel motion estimation is a very important factor in the success of the SR image reconstruction algorithm. Since the shifts between LR images are arbitrary, the registered HR image will not always match up to a uniformly spaced HR grid. Thus, nonuniform interpolation is necessary to obtain a uniformly spaced HR image from a nonuniformly spaced composite of LR images. Finally, image restoration is applied to the upsampled image to remove blurring and noise.

Using the nonuniform interpolation SR approach, which is the most intuitive method for SR image reconstruction [11,12], the low-resolution observation image sequence is registered, resulting in a composite image composed of samples on a nonuniformly spaced sampling grid. These non-uniformly spaced sample points are interpolated and re-sampled on the high-resolution sampling grid (see Fig. 2). Applying this relatively
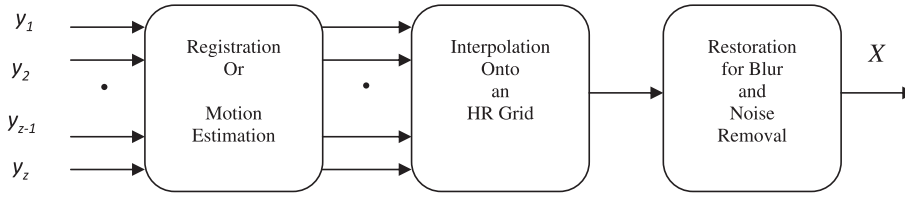
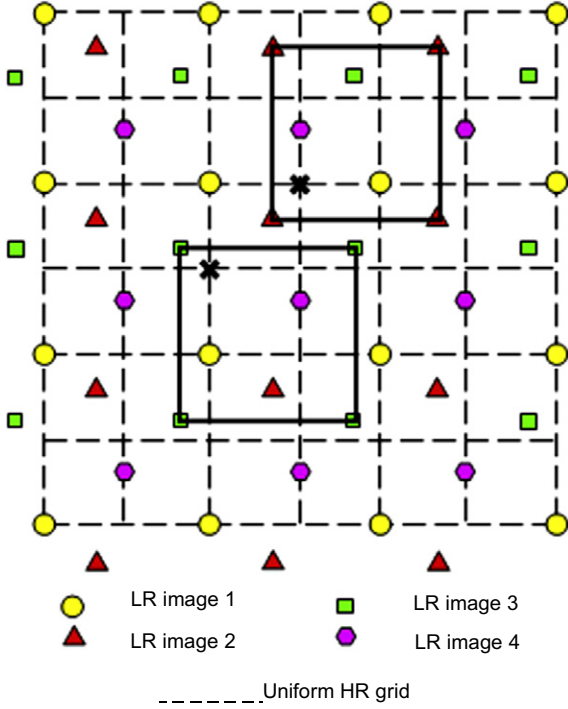**Figure 1**   The scheme for Super Resolution.

Figure 2   Nonuniform interpolation from LR grid into HR grid.

low computational load approach, we construct a high resolution image from four low resolution images (Landsat-7 TM) for the same scene (the same segment of the geo reference data).

Training of the classification model is takes place by dividing both the LR & HR dataset scenes into small subimages of 128-by-128 pixels. The classification problem involves the identification of seven land cover types (see Tables 3 and 4). Each scene is rectified and consists of seven bands. We choose the suitable band combination that reflect the desired land cover types such as water, vegetation and urban. As the application here is land use/cover, we choose the band combination to be bands (1, 4, 7).

For the two resolution datasets, the subimages feature vectors are extracted for each subimage regions, which based, for example, on color, shape, mean, variance, location of the subimage four corners. These extracted feature vectors have been stored and indexed in the database in a way that helps the retrieval stage. This is done by attaching to each subimage some indicators that help to decide if the subimage is classified to its right cluster correctly or not. A simple strategy for the back-propagation neural network classifier is developed to exploit information obtained from different resolutions and thus, to improve the classification results [13–15]. We use information from both resolutions by incorporating them simultaneously in a classification routine [2].

3. Content-based retrieval system

A query initiated by the selection of the region of interest from a key image. This identifies the object or the scene's element, which should be present in the retrieved subimages. The system selects a preliminary set of images by minimizing the Euclidian distance measure from the region's feature vector to those of potentially similar regions. Let the feature vector dimensionality to be N. Given that region r_k from image p_k is chosen as the key, then the best match in the initial query will be region r_m chosen from image p_m if

$$D(r_k, p_k, r_m, p_m) = \min(D(r_k, p_k, r_i, p_j)) \tag{2}$$
$$\forall\, i = \{0 \dots M\} \text{ and } j = \{0 \dots P\}$$

where

$$D(r_1, p_1, r_2, p_2) = \sqrt{\sum_{i=1}^{N} (f_i(r_1, p_1) - f_i(r_2, p_2))^2} \tag{3}$$

Since each region has feature vector consisting of the elements $\{f_1, \dots, f_N\}$, a radial basis function neural network (RBF) is used to cluster this data [16,17]. Centroids of RBF are determined in the initialization. The number of clusters varies according to the volume of the input data but with t training examples, it usually returns between $t/3$ and $t/2$ clusters. According to the locality of the feature vectors for the user's classified examples they are classified as relevant (positive examples) or irrelevant (negative examples). Then to get the next group of subimages, feature vectors of all regions in all subimages in the database are compared to the vectors describing the node centroids. Assume that there are Q clusters each with $\{c_1, \dots, c_Q\}$, the Euclidian distance between a given region's feature vector and each of these clusters is calculated as in Eq. (3) hence the cluster $C_{\min}$ with minimum distance found. The user identifies a variable threshold θ of the cluster radius. The iterative refinement continues until the user is satisfied with the resulting subimages. If, at any stage, the user is unhappy with the direction of the system, then the user can take a new key region that added to the dataset. This has been found to avoid the local minima in the training stage (see Figs. 3–5).

Each image group can be viewed as a node in a feedback neural network characterized by its centroid and its variance i.e. there exist a transformation so that every feature vector can be expressed in terms of the centroid and variance of all the image groups. The RBF is a nonlinear transform that provides a set of functions, which constitute a basis for the input feature vector. This transform can be modified such that, each component represents the membership function of a subimage to a group.
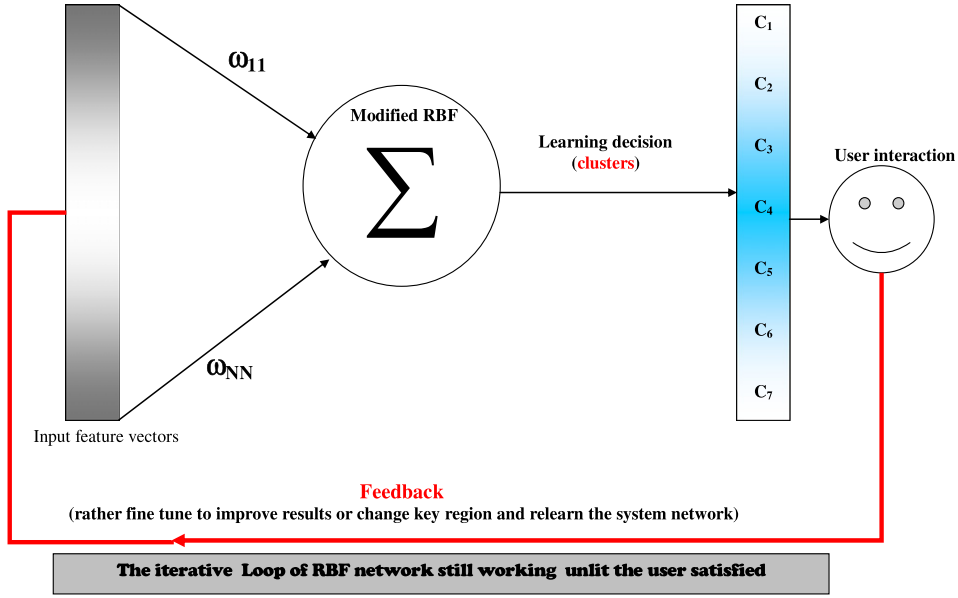
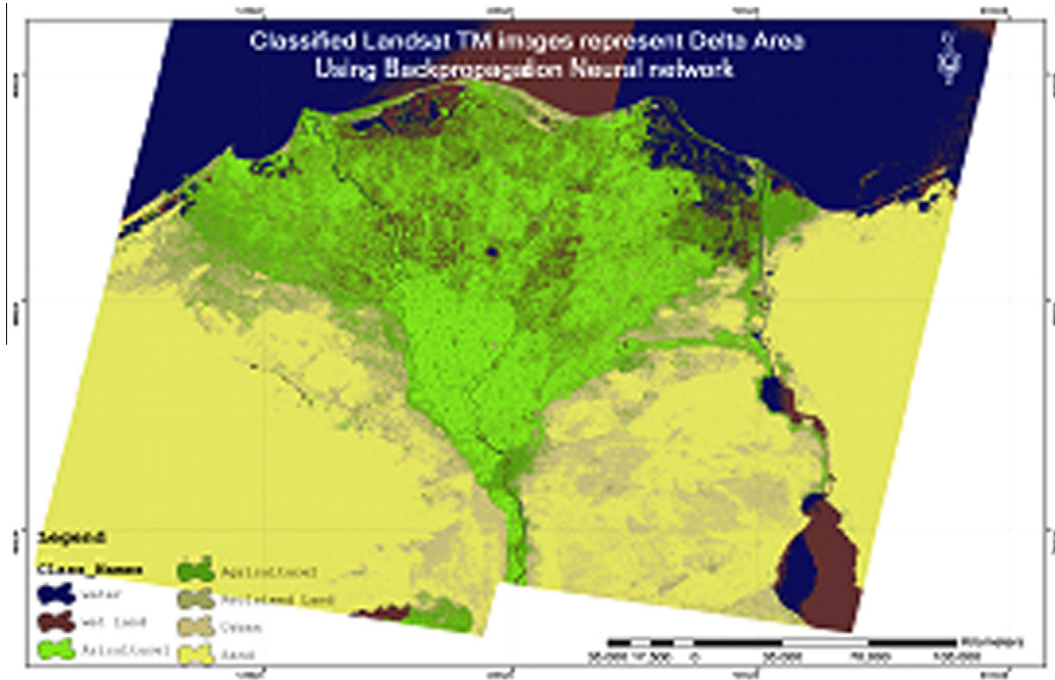**Figure 3**   The RBF neural net schematic.

Figure 4   Classification result for the Nile Delta of Egypt as an example of supervised classification.

Let x be an arbitrary image feature vector, c_i the centroid of the ith cluster feature space and $\mathbf{S}$ number of image clusters. The modified RBF transform maps x to $F(x)$ according to the equation

$$[F(x)]_i = \exp\left\{\frac{-1}{2\sigma_i^2}\|x - c_i\|^2\right\} \qquad (4)$$

where $[F(x)]_i$ is the ith component of $F(x)$ and σ_i^2 is the variance of the ith cluster. RBF transform represents the mem-

bership function of each image to a group. The proposed system transforms each subimage region feature vector x to $F(x)$ by applying the modified RBF transform utilizing the feedback information in the form $F(x)$, the weights in the network are updated using a correlation matrix. In order to embed relevance feedback information into the system, the weights $\{w_{ij}|1 \leqslant i, j \leqslant S\}$ which contain the relationship between group i, and group j are updated, using the correlation matrix M_k
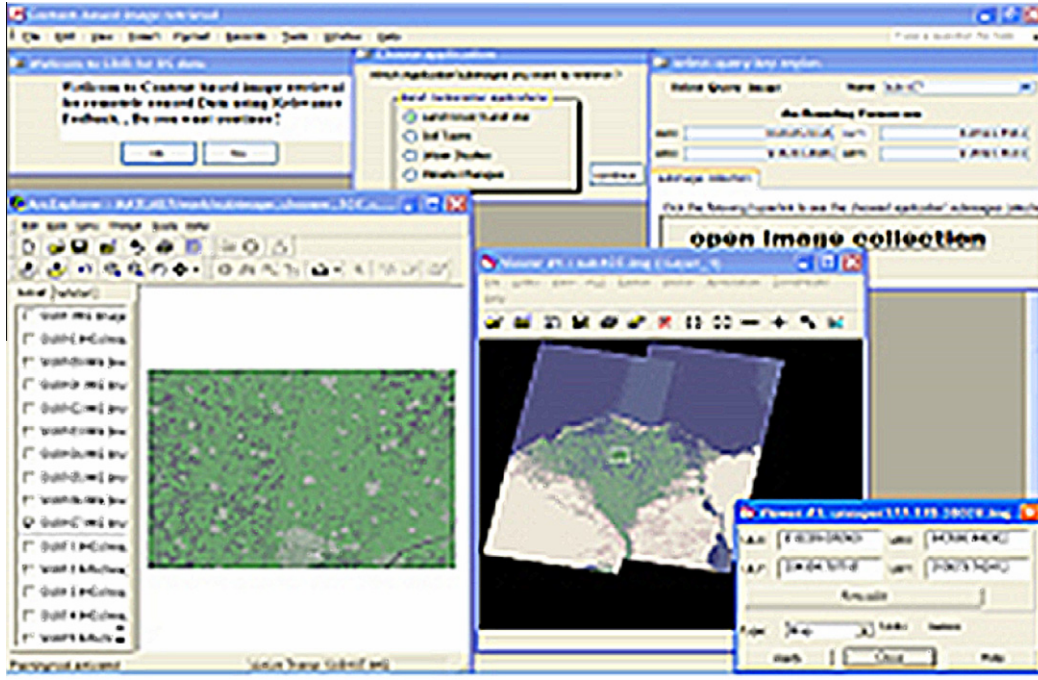
Figure 5 A snapshot of the system in the query image chooser stage.

$$M_k = \begin{bmatrix} w_{11} & w_{12} & \cdots\cdots\cdots\cdots\cdots & w_{1S} \\ w_{21} & w_{22} & \cdots\cdots\cdots\cdots\cdots & w_{2S} \\ \cdot & \cdot & & \\ \cdot & \cdot & & \\ \cdot & \cdot & & \\ \cdot & \cdot & & \\ \cdot & \cdot & & \\ w_{S1} & w_{S2} & & w_{SS} \end{bmatrix} \qquad (5)$$

In addition, k is the current iteration. Suppose for a given iteration, $n + m$ images are displayed and the user marks n images as being relevant, then the rest m images are considered as irrelevant to the query.

Let q be the query feature vector, $\{p_i | 1 \leqslant i \leqslant n\}$ the set of positive feedback vectors and $\{n_i | 1 \leqslant i \leqslant m\}$ the set of negative feedback vectors .the correlation matrix is updated as follow:

$$M_k = M_{k-1} + \sum_{i=1}^{n} F(q)F(p_i)^T - \sum_{i=1}^{m} F(q)F(n_i)^T \qquad (6)$$

where M_{k-1} represent the previous estimate of the weight matrix, M_k is the updated weight matrix based on the relevance feedback provided by the user, and $F(x)$ is the membership function of the feature vectors. Computing correlation as in Eq. (6), the weights between positive clusters are increased and the weights between negative clusters are decreased.

The system correlation matrix saves updates, and correlates the subimage groups to make the system learn progressively with each new session and become less dependent on the initial settings. The cluster splitting and merging process eventually breaks the feature space into semantically related clusters. For non-neighboring clusters that contain semantically related subimages, the correlation weights between those clusters of subimages are large in value. Thus, the correlation matrix is used to guide the system search process for retrieval, so rather than searching nearby clusters, the system is allowed to jump across clusters of subimages to search for semantically related clusters.

4. Materials and methods

Both LR and HR database of remotely sensed data has been used, which consists of the original 300 Landsat-7 TM satellite images scenes that cover different areas in Egypt and show land use/land cover, plus the constructed corresponding high resolution one.

Training of the system is done off-line for both LR and HR images; the used algorithm is given as follows:

(1) Layer stacking and rectifying the images.
(2) Choose the suitable band for the application (in our case we choose layers that reflect land use/cover bands 7, 4, 1).
(3) Divide each image scene into subimages with 128-by-128 pixels, and bands 7, 4, 1.
(4) Classify subimages to get segmented subimages.
(5) Extract the feature vector from each subimage region.
(6) Build database to store classified (segmented) images.
(7) Compute the Euclidean distance between the feature vector of the query subimage key region, and the stored feature vectors of the subimages regions in the database to get preliminary candidate cluster of subimages that contain all the subimages with regions of minimum Euclidean distance values as initialization.
(8) Calculate redial basis functions neural network centroids.
(9) Use the modified radial basis function transform that maps the feature vector X to $F(X)$ as in Eq. (4).
(10) Update the RBF weights by updating the correlation matrix M_k.

(11) Take the user's feedback to mark images as relevant or irrelevant then update the subimage groups by merging and splitting groups, and update the correlation matrix too.

(12) Fine-tune the system results by re-clustering the database images, if user is not satisfied with the system's results direction (i.e., the system not converges to the interest region) another key region can be chosen.

Using 14 hidden nodes for back-propagation neural network; Table 1 shows the learning results. We noticed that after 50,000 iterations the learning rate and the total error, they both almost negligible. By running 11 test simulations; Table 2 shows the resulted classification accuracy confidence. Fig. 6

shows comparison between the confidence accuracy for both the learning phase and the testing one.

5. Results

Finally after learning the neural network off line for the whole data sets, and fine tuning the system by applying the previous algorithm, the system now is ready to be tested by using any query image from the data set. Fig. 7 shows an example of the resulted top 20 candidates of the retrieved subimages for the giving query image.

A 1767 different test image samples has been used to evaluate the system performance. The samples were distributed over the seven predetermined ground categories (water, old agriculture, new agriculture, sand, wet land, urban, Reclaimed land). Using only the LR data set, the system achieved 81.1% classification accuracy. Table 3 shows the detailed results for each category. A significant improvement of the system classification accuracy has been done by using both the LR and HR data sets. Table 4 shows that the accuracy of the system has been improved to be 89.4% for the same test samples. Comparing the results of both Tables 3 and 4, it is obvious that the main contribution comes from the two categories (urban and roads), while the other categories have not been affected by applying the multi-resolutions data set.

Table 1 Learning schedule for back-propagation neural network (with 14 hidden units).

Learn count	10,000	30,000	50,000
Learning rate	0.3	0.15	0.0375
Momentum	0.4	0.2	0.05
No. of inputs and No. of outputs	3 and 7	3 and 7	3 and 7
No. of hidden layers	1	1	1
Total error	0.89	0.35	0.01

Table 2 Confidence accuracy for backpropagation neural network classification.

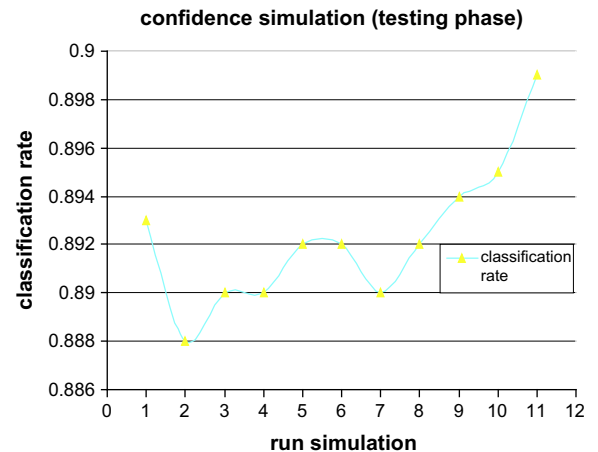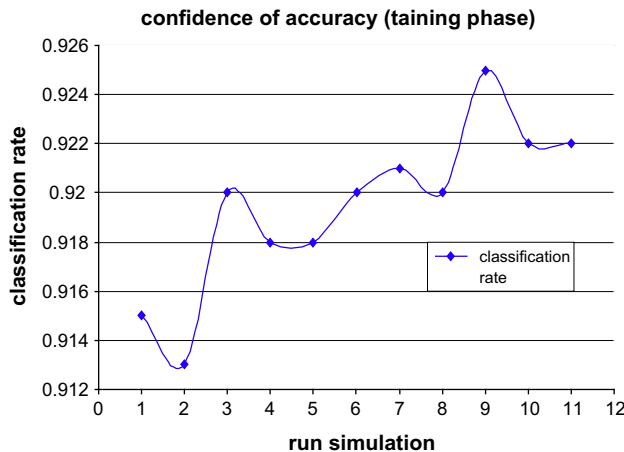| Run simulations	Classification rate	Train set average	Std. dev.	Classification rate	Test set average	Std. dev.
1	0.915	0.922	0.001	0.893	0.895	0.004
2	0.913			0.888		
3	0.92			0.89		
4	0.918			0.89		
5	0.918			0.892		
6	0.92			0.892		
7	0.921			0.89		
8	0.92			0.892		
9	0.925			0.894		
10	0.922			0.895		
11	0.922			0.899		
Confidence interval of simulations 9%	0.919454545	0.922	0.998	0.892272727	0.895	0.903

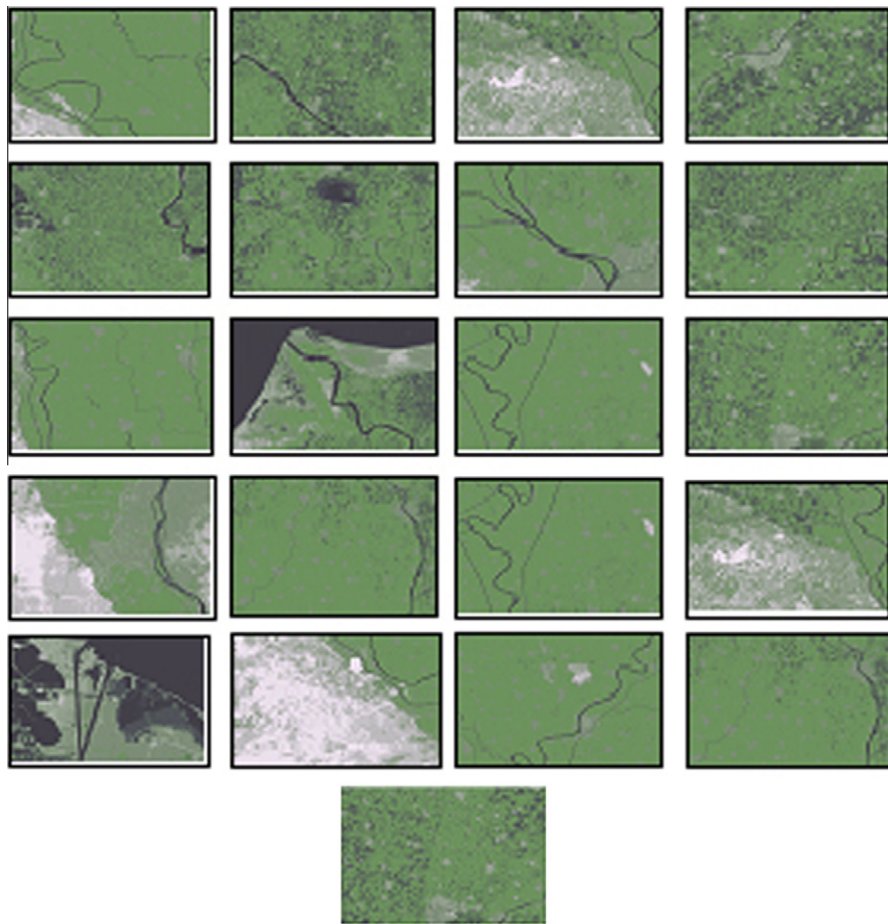**Figure 6** Confidence accuracy (a) training, (b) testing phases.

Figure 7 Example of the final results of the system that shows the top best 20 retrieved subimages for the following query image.

Table 3 Classification results using the low resolution dataset only (single resolution classification) *81.1% classification accuracy at 0.9 confidence level.*

Ground categories	Neural network classified classes							Total
	Water	Agriculture 1	Agriculture 2	Sand	Mixed grass	Urban	Roads	
Water	159	9	0	2	0	0	0	170
Old agriculture	2	384	6	1	5	0	0	398
New agriculture	0	0	147	0	0	5	0	152
Sand	5	0	0	450	9	0	0	464
Wet land	4	0	5	7	93	0	3	112
Urban	0	0	3	0	0	263	30	296
Reclaimed land	0	0	0	0	0	78	97	175
Total	170	393	161	460	107	346	130	1767

6. Discussion

One of the fundamental considerations when using remotely sensed data for land use/cover mapping is that of selecting appropriate spatial resolution(s). With the increased availability of very high resolution multi-spectral images spatial resolution variation will play an increasingly important role in the employment of remotely sensed imagery. The correct application of image classification procedures for mapping land use/cover requires knowledge of certain spatial attributes of the data to determine the appropriate classification methodology and parameters to use. In general, traditional single-resolution classification procedures are inadequate for understanding the effects of the chosen spatial resolution. They have difficulty discriminating between land use/cover classes that have complex spectral/spatial features and patterns.

We can notice from the previous results that, using a single "low" resolution as in Table 3, the most errors concentrated in only two categories (urban and roads) which need a finer resolution. On the other hand by using multiple resolutions "low and high" as in Table 4, the same two categories have been

Table 4 Classification results using both the low and high resolution datasets (multi-resolution classification) *89.4% classification accuracy at 0.9 confidence level.*

Ground categories	Neural network classified classes							Total
	Water	Agriculture 1	Agriculture 2	Sand	Mixed grass	Urban	Roads	
Water	158	8	0	3	1	0	0	170
Old agriculture	2	388	4	1	3	0	0	398
New agriculture	0	0	142	0	7	3	0	152
Sand	6	0	2	445	11	0	0	464
Wet land	3	0	4	5	100	0	0	112
Urban	0	0	1	0	0	277	18	296
Reclaimed land	2	0	4	12	0	34	123	175
Total	171	396	157	466	122	314	141	1767

significantly improved, while the rest of categories (which need coarser resolution) have not been affected much.

7. Conclusions

In this paper, we presented a content-based retrieval system of large database of satellite images. We used the modified RBF transform for clustering because of its varied values of the variance.

The multi-resolution framework proposed in this paper recognizes that image classification procedure should account for image spatial structure to minimize errors, and increase efficiency and information extraction from the classification process. Selection of the training scheme and classification decision rules should be guided by specification of the type of scene model (high and low resolution) and level of spatial variance represented by the image to be classified. A Super Resolution approach has been used to generate a high resolution image dataset. Different spatial analysis methods can provides the above information to allow resolution effects on individual classes examined. Different strategies can be used to incorporate information from multiple resolutions.

The results illustrated the potential of multi-resolution classification framework. Using a simulated multi-resolution dataset and one multi-resolution strategy, it was demonstrated that multi-resolution classification approaches developed could significantly improve land use/cover classification accuracy when compared with those from single-resolution approaches. Multiscale data analysis can provide useful information to ensure that subsequent classification methods and parameters are suited to the spatial characteristics of the features (or classes). The results confirm the validity and efficiency of the proposed framework.

References

[1] Lu D, Weng Q. A survey of image classification methods and techniques for improving classification performance. Int J Remote Sens 2007;28(5):823–70.

[2] Chen D. A multi-resolution analysis and classification framework for improving land use/cover mapping from earth observation data, the international archives of the photogrammetry. Remote Sens Spatial Inform Sci 2005;34(Part XXX).

[3] Foody GM. Status of land cover classification accuracy assessment. Remote Sens Environ 2002;80:185–201.

[4] Davidson G, Ouchi K, Saito G, Ishitsuka N, Mohri K, Uratsuka S. Single-look classification accuracy for polarimetric SAR. Int J Remote Sens 2006;27(22):5073–80.

[5] Chen D, Stow DA, Gong P. Examining the effect of spatial resolution on classification accuracy: an urban environmental case. Int J Remote Sens 2004;25(00):1–16.

[6] Li J, Gray RM, Olshen RA. Multiresolution image classification by hierarchical modeling with two-dimensional hidden Markov models. IEEE Trans Inform Theory 2000;46(5):1826–41.

[7] Solberg A, Taxt T, Jain AK. A Markov random field model for classification of multisource satellite imagery. IEEE Trans Geosci Remote Sens 1996;34(1):100–13.

[8] Ezzat H, Hefnawy A. A fast adaptive content-based retrieval system of satellite images database using relevance feedback. In: Proceeding of world academy of science, engineering and technology, vol. 13, May, 2006. ISSN 1307-6884.

[9] Costa GH, Bermudez JC. Statistical analysis of the LMS algorithm applied to super resolution image reconstruction. IEEE Trans Signal Process 2007;55:2084–95.

[10] Borman S, Stevenson RL. Super-resolution from image sequences—a review. In: Proceeding midwest symposium circuits and systems, vol. 5, April, 1998.

[11] Chan JC, Jianglin M, Kempeneers P, Canters F. Superresolution enhancement of hyperspectral CHRIS/proba images with a thinplate spline nonrigid transform model. IEEE Trans Geosci Remote Sens 2010;48(6):2569–79.

[12] Pestak TC. Development of an efficient super-resolution image reconstruction algorithm for implementation on a hardware platform. Master of Science in Engineering (MSEgr), Wright State University; 2010.

[13] Benediktsson JA, Swain PH, Erosy O. Neural network approaches versus statistical methods in classification of multisource remote sensing data. IEEE Trans Geosci Remote Sens 2007;28(4):540–52.

[14] Shafri H, Suhaili A, Mansor S. The performance of maximum likelihood, spectral angle mapper, neural network and decision tree classifiers in hyperspectral image analysis. J Comput Sci 2007;3(6):419–23.

[15] Buddhiraju KM. Contextual refinement of neural network classification using relaxation labeling algorithms. In: 22nd Asian conference on remote sensing, November, 2001.

[16] Senapati MR, Vijaya I, Dash PK. Rule extraction from radial basis functional neural networks by using particle swarm optimization. J Comput Sci 2007;3(8):592–9.

[17] El-Sherbiny M. Particle swarm inspired optimization algorithm without velocity equation. Egypt Inform J 2011;12(1):1–8.