A HIGH ACCURACY LAND USE/COVER RETRIEVAL SYSTEM

A. A. Hefnawy

Computer & System Dept., Electronics Research Institute, Cairo, Egypt
E.mail: alaa@eri.sci.eg

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 forty years, most of them were employed in single-resolution image classification. Due to the more heterogeneous spectral-radiometric characteristics within land use/cover units portrayed in high resolution images, many applications of traditional single resolution classification approaches have not led to satisfactory results. In this paper, we propose a fast adaptive content-based retrieval system of satellite images database using relevance feedback. 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. To improve the accuracy of the system, we apply the backpropagation supervised artificial neural network classifier for both the low and high resolution datasets.

1. INTRODUCTION

One of the fundamental characteristics of a remotely sensed image is its spatial resolution; as the basic information contained in the image is strongly dependent on spatial resolution (WOODCOCK, 1987). 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 (Chen, 2005; Foody, 2002). 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 (Marceau, 1994). 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 (WOODCOCK, 1987), 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 (Chen, 2004; Li, 2000; Solberg 1996).

The objective of this paper is to build a high accuracy content-based retrieval system of satellite images. Our proposed human-computer interactive system is based on relevance feedback. A large database of remotely sensed data has been used, which consists of Landsat-7 TM satellite images scenes that cover different areas in Egypt and show land use / land cover (Ezzat, 2006). 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 and algorithm will be presented in section (3). The classification results are shown in section (4), and finally conclusions are given in section (5).

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 21June 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 (Tsai, 1984; Borman, 1998).

Using the nonuniform interpolation SR approach, which takes relatively low computational load (Ur, 1992; Nguyen 2000), we construct a high resolution image from 4 low resolution images (Landsat-7 TM) for the same scene.

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. 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 backpropagation neural network classifier is developed to exploit information obtained from different resolutions and thus, to improve the classification results (Benediktsson, 2007; Paul, 1995; Buddhiraju, 2001). We use information from both resolutions by incorporating them simultaneously in a classification routine (Chen, 2005).

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

$$D(r_k, p_k, r_m, p_m) = \min(D(r_k, p_k, r_i, p_j)) \quad (1)$$

$$\forall i = \{0.....M\} \, and \, j = \{0...P\}$$

Where

$$D(r_1, p_1, r_2, p_2) = \sqrt{\sum_{i=0}^{N} (f(r_1, p_1) - f(r_2, p_2))^2} \quad (2)$$

Since each region has feature vector consisting of the elements $\{f_0,.....,f_N\}$, a radial basis function neural network (RBF) is used to cluster this data (Mc Cullac, 1995). 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 C clusters each with $\{c_0,......,c_Q\}$, the Euclidian distance between a given region's feature vector and each of these clusters is calculated as in equation (2) 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.
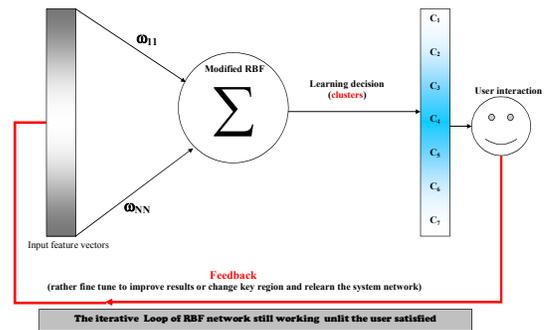

Figure 1. The RBF neural net schematic

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.

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

$$[F(x)]_i = \exp\{ \frac{-1}{2\sigma_i^2} \|x - c_i\|^2 \} \tag{3}$$

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 membership 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 $\{w_{ij}| 1\le i,j\le N\}$ which contain the relationship between group I, and group j are updated, using the correlation matrix M_k

$$M_k = \begin{bmatrix} w_{11} & w_{12} & \cdots & \cdots & \cdots & w_{1N} \\ w_{21} & w_{22} & \cdots & \cdots & \cdots & w_{2N} \\ \cdot & \cdot & & & & \\ \cdot & \cdot & & & & \\ \cdot & \cdot & & & & \\ \cdot & \cdot & & & & \\ \cdot & \cdot & & & & \\ w_{N1} & w_{N2} & & & & w_{NN} \end{bmatrix} \tag{4}$$

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.
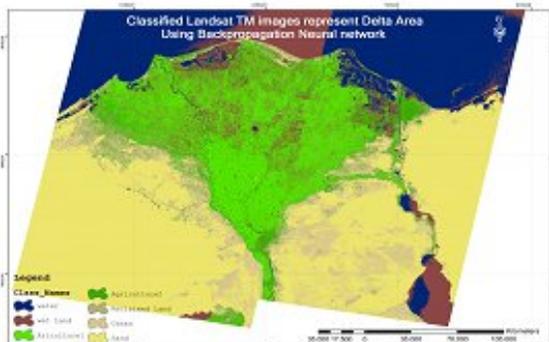
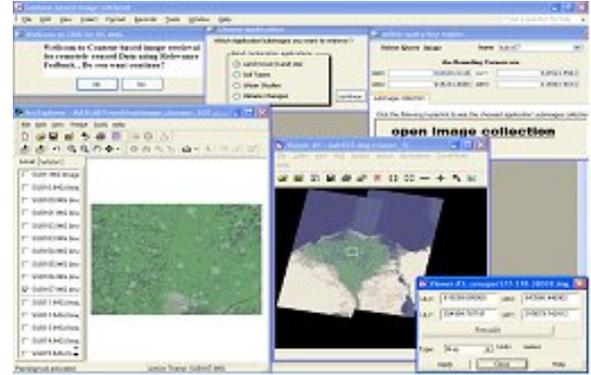Figure 2. Classification result for the Nile Delta of Egypt as an example of supervised classification

Figure 3. A snapshot of the system in the query image chooser stage

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

$$M_k = M_{k-1} + \sum_{i=1}^{n} F(q)F(p_i)^T - \sum_{i=1}^{m} F(q)F(n_i)^T \tag{5}$$

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 equation (5), 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.

Training of the system is done off-line; 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-by128 pixels, and R =band 7, G=band 4, B=band 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 follows

$$\left[F\,(\,x\,)\,\right]_i \;=\; \exp\;\left\{ \frac{-\,1}{2\,\sigma_i^{\,2}}\,\|x\,-\,C_i\|^2 \right\}$$

Where σ^2 is the variance of the ith cluster, C_i is the ith cluster feature space, $[F(x)]_i$ is the ith component of F(x), RBF transform represent the membership function of each image to a group

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, another key region can be chosen.

4. RESULTS

Table 1. 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

Table 2. Classification Results using both the Low & High Resolution Datasets (multi-resolution classification) *83.2% 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

5. 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.

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.

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 (H- and L-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.

REFERNCES

Benediktsson,J.A., Swain, P.H., and Erosy, O., 2007. Neural network Approaches versus Statistical methods in classification of Multisource Remote sensing Data, *IEEE Trans. Geoscience and Remote sensing*, vol. 28,no.4,pp540-552.

Borman S. and Stevenson R.L., 1998. Spatial resolution enhancement of low-resolution image sequences. A comprehensive review with directions for future research, Lab. Image and Signal Analysis, University of Notre Dame, Tech. Dep.,

Buddhiraju K. M., 2001."Contextual Refinement of neural network classification using Relaxation Labeling Algorithms", 22nd Asian Conference on Remote sensing. November

Chen D. M., 2005. A Multi-Resolution Analysis and Classification Framework for Improving Land Use/Cover Mapping from Earth Observation Data, *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* Vol. 34, Part XXX.

Chen D., Stow D., and Gong P., 2004. Examining the effect of spatial resolution on classification accuracy : an urban environmental case. *International Journal of Remote Sensing* 25(00):1-16.

Ezzat H. M., and Hefnawy A. A., 2006. A Fast Adaptive Content-based Retrieval System of Satellite Images Database using Relevance Feedback. *Proceeding of World Academy of Science, Engineering and Technology* vol. 13 May ISSN 1307-6884

Foody, G. M., 2002. Status of land cover classification accuracy assessment. *Remote Sensing of Environment*, 80, pp. 185-201.

Li, J., Gray R.M., Olshen R.A., 2000. Multiresolution image classification by hierarchical modeling with two-dimensional hidden Markov models. *IEEE Transactions on Information Theory*, 46 (5): 1826-1841.

Marceau, D.J., P.J. Howarth, D.J. Gratton, 1994. Remote sensing and the measurement of geographical entities in a forest environment 2: The optimal spatial resolution. *Remote Sensing of Environment*, 49, pp. 105-117.

Mc Cullac and Pits, 1995. *"Artificial Neural Networks :An Introduction Radial Basis Function Neural Network",* Chapter7, pp 236-284,

Nguyen N. and P. Milanfar, 2000. "An efficient wavelet-based algorithm for image super resolution," in Proc. Int. Conf. Image Processing, vol. 2, , pp. 351-354.

Paul L. Rosin and Freddy Fierens, 1995. "Improving Neural Network Generalization", In Proceedings of IGARSS'95,

Solberg, A.H.S.; Taxt, T.; Jain, A.K. 1996, "A Markov random field model for classification of multisource satellite imagery". *IEEE Transactions on Geoscience and Remote Sensing*, 34 (1):100-113.

Tsai R.Y. and Huang T.S., 1984. *"Multipleframe image restoration and registration,"* in Advances in Computer Vision and Image Processing. Greenwich, CT: JAI Press Inc., pp. 317-339.

Ur H. and Gross D., 1992. "Improved resolution from sub-pixel shifted pictures," *CVGIP: Graphical Models and Image Processing*, vol. 54, pp. 181-186, Mar.

WOODCOCK, C. E., and STRAHLER, A. H., 1987, "The factor of scale in remote sensing". *Remote Sensing of Environment,* 21, 311–332.