CLOUD FREE MOSAIC IMAGES

T. Hosomura, P.K.M.M. Pallewatta
Division of Computer Science
Asian Institute of Technology
GPO Box 2754, Bangkok 10501, Thailand

ABSTRACT

Certain areas of the earth's surface are constantly covered by clouds during most of the time of the year. Obtaining cloud free images of such areas is an extremely difficult task. Neural networks can be meaningfully used as classifiers in situations where the data to be classified is of non parametric nature. This will be the situation encountered when we consider all the clouds as one class and all the clear sky areas as another class. In this study we make a relatively cloud free mosaic, by using a neural network classifier for cloud detection. This experimental part of this study has been carried out with MOS-1 data.

KEY WORDS: Neural Networks, Classification, Mosaics

I. INTRODUCTION

Remote sensing of the Earth's surface is of prime importance today. Remote sensing is applied today in a number of fields. Environmental monitoring, crop estimation, land use classification , cartography are only a few examples. But unfortunately some areas of the earth are constantly covered by clouds and obtaining cloud free images of such areas is a difficult task. Present day earth resource observation satellites have comparatively high revisit capabilities, which allows us to obtain images of the same areas of the earth more frequently.

The neural network approach to pattern recognition is attractive because it has the capability to learn patterns whose complexity makes them difficult to define using more formal approaches. Also it has the ability to incorporate new features without degrading prior learning.

Recently there has been a great resurgence of research in neural networks. New and improved neural network models have been proposed, models which can be successfully trained to classify complex data. The generalized delta rule is one such. Neural network models have as an advantage over statistical methods that they are distribution free and no prior knowledge about statistical distributions of classes is needed to apply these methods for classification.

In this study an effort has also been made to develop image processing techniques to make cloud free mosaic images of such areas using daylight images of the same area obtained at different times (multitemporal images). Clouds and shadow areas have been detected using a backpropogation neural network. To make cloud free mosaic images the following steps should be done.

1) Detecting clouds and shadows in several multitemporal images.

2) Registering the above images.

3) Making a cloud mask.

4) Intensity matching of the segments of the mosaic.

5) Constructing the mosaic.

2. NEURAL NETWORK CLASSIFIERS

2.1 A Brief Description of Artificial Neural Networks

Neural network models have an advantage over statistical methods that they are distribution free and no prior knowledge is necessary about the statistical distributions of the classes in the data sources in order to apply these methods for classification. The neural network methods also take care in determining how much weight each data source should have in classification. A set of weights describe the neural network and these weights are computed in a iterative training procedure. On the other hand neural network models can be very complex computationally and need a lot of training samples to be applied successfully and their iterative training procedures are slow to converge. Also in practice the performance of the neural network models in classification is more dependant on the training samples, whereas statistical methods provide a model for each class. To overcome this difficulty a good generalization of the network is necessary so that the network can correctly respond to patterns not presented during training.

A neural net work is a network of neurons, wherein a neuron can be described in the following way: a neuron has many input signals x_j , $j=1,2,3,....N$ which represent the activity at the input or the momentary frequency of neural impulses derived by another neuron to this input. The following is a simplest formal model of a neuron.

$$o=K\phi\left(\sum_{j=1}^{N} w_j x_j-\theta\right) \qquad (1)$$

where $\underline{k}$ is a constant ø is a non linear function which takes 1 for positive arguments and -1 for negative arguments. The $\underline{w_j}$ are called weights and θ is a threshold.

In the neural network approach to pattern recognition, the neural network operates as a black box which receives a set of input vectors **x** and produces responses $\underline{o_i}$ from it's output units $\underline{i}$ ($\underline{i = 1,2,....L}$). The general idea followed in neural network theory is that if a neuron $\underline{i}$ is active $\underline{o_i}=1$ and $\underline{o_i}=-1$ (or 0) if inactive. The process is then to learn weights through an adaptive (iterative) training procedure. The training procedure is ended when the network is stabilized , i.e the weights do not change from one iteration to another or change less than a threshold amount. Then the data is fed in to the network to perform classification, and the network provides the output class number of each pixel.

2.2 <u>The Generalized Delta Rule</u>

The generalized delta rule or the back propagation of errors, involves two phases. During the first phase the input is presented and propagated forward through the network to compute the output value o_{pj} in presentation $\underline{p}$ for each unit $\underline{j}$; i.e.,

$$o_{pj}=f_j(net_{pj}) \qquad (2)$$

where $net_{pj}=\sum_i w_{ji}o_{pi}$, w_{ji} is the weight of the connection from unit $\underline{i}$ to unit $\underline{j}$, and f_j is the semilinear activation function at unit $\underline{j}$ which is differentiable and non decreasing. A widely used choice for a semilinear activation function is sigmoid function :

$$f_i(net_{pj})=\frac{1}{1+\exp-(net_{pj}+\theta_j)} \qquad (3)$$

The second phase involves a backward pass through the network (analogous to the initial forward pass), during which the error signal δ_{pj} is passed to each unit and the appropriate weight changes are made according to

$$\Delta_{pw_{ij}}=\eta\delta_{pj}o_{pi} \qquad (4)$$

This second backward pass allows the recursive computation of δ_{pj}. Δ_{pwij} also gives the negative value of the gradient of error at the outputs of neurons multiplied by η which is known as the learning rate.

The learning procedure requires only that the change in weight be proportional to the weight error derivative. True gradient descent requires that infinitesimal steps be taken. The constant of proportionality or the learning rate η should be set as large as possible for rapid learning, but it still should not lead to oscillation. One way to increase the learning rate without leading in to oscillation is to modify the back propagation learning rule to include a momentum term μ. This can be accomplished by the following rule,

$$\Delta_{pw_{ij}}=\eta\delta_{pj}o_{pi}+\mu\Delta_{p-1w_{ij}} \qquad (5)$$

The derivation of the back propagation algorithm is found in RUMMELHART, McCLELLAND et al. (1986 v.1).

3. REGISTRATION AND MOSAICS

Proper registering of the source images is necessary for making cloud free mosaics. Conventional approaches to image registration needs a certain degree of human involvement. The process of image registration can be divided in to four steps.

1) Control point selection.

2) Control point matching.

3) Estimation of the mapping function using matched control points.

4) Spatially registering one image with another using the mapping function computed in step 3.

Steps 3 and 4 can be fully automated and a variety of commercial software packages are available for this purpose, although for our research purposes we will be using our own software for this. Steps 1 is the most difficult to automate in general and requires a substantial amount of human involvement.

3.1 Control Point Matching

The task here is to determine the corresponding structures in the images to be registered, in order to derive the transformation function between them. In this study we will be using manual control point matching, due to certain problems described later.

3.2 Transformation Functions

Selection of right transformation function is another important factor in registering digital images. The best transformation function for registering two images which have only transnational differences is very simple and has only two unknown parameters. Applying a transformation function with more parameters make the registration process costly and less accurate.

Usually more control points are used than the number of unknown parameters in the transformation functions, and a least mean square solution is obtained.

4. EXPERIMENTAL RESULTS

4.1 Detecting Clouds Using a Neural Network Classifier

In this study a neural network classifier has been selected be cause it is capable of forming disjoint and complex decision regions in feature space and does not make any assumptions about the distributions of the under laying classes. This will be very useful when we consider all the clouds and shadows as one class, and the rest as another class, since the clouds are in the bright end of Visible and near IR bands and the shadows are in the other extreme. It is necessary to employ an near IR band in order to distinguish between water and cloud because especially where there is sun glint in the visible band they tend to have similar spectral signatures.

By the nature of the classification problem to be solved it can be seen that the problem is not linearly separable. It is a widely known fact that the Delta Rule perceptrons are unable to solve this type of problems. So it was decided to build the classifier based on Generalized Delta Rule mulitilayer perceptons.

Since our original scenes did not include such high albedo surfaces as snow it was decided to limit the input information to two bands of spectral data only. Since the two visible bands and the two near IR bands were highly correlated among themselves, it was decided to employ one band from visible range and the other from near IR. The choices were band 2 and band 4 , since these bands showed the highest contrast upon visible inspection. Keeping these constraints and limitations in mind the first multilayer perceptron with sigmoid output function was made for cloud/shadow detection. This mulitilayer perceptron had 2 input nodes, 8 nodes in first hidden layer, 2 hidden nodes in second hidden layer and one output node. All the neural network simulations were carried out using the software package provided with the book Explorations in Parallel Distributed Processing (McCLELLAND, RUMELHART 1988), which was modified and included as an integral part of our image processing system. In training the neural network the training data was obtained form obviously cloud covered areas, from areas cover by shadows created by the clouds and from different types of land areas and water bodies. It is important that the training data set included data from each sub-class within a given complex class, in order to achieve good generalization. The data thus obtained was fed to the neural network and the total error measured in the network was minimized.

The learning rate of the network was initially set to 0.3, but it was observed that with real data this learning rate led to oscillatory behavior of the network's total error. So the learning rate was reduced to 0.1 , but with some data some decaying oscillatory behavior was observed.

The main problem encountered here was that the neural network was unable to train itself given real data from the training areas. The first step towards the solution of the problem was to get the data from the training area via a 3x3 pixel averaging filter. The well known law of large numbers in statistics show us that this reduces the variability of the data and hence the variance. So the training data thus collected would not be truly representing real data, and this method was dropped , since it can lead to poor classification accuracies. It must be mentioned that even with the averaging filter the neural network was unable to train itself given real data taken from the images.

Another phenomenon observed was that the neural network was able to train itself with a small set of training data. But such arbitrary selection of a limited number of samples of training data would lead to poor classification accuracies since they are hardly representative of the dispersion of the data in the feature space. So the neural network, was trained with a small set of data obtained by having the mean vector of each training area to represent each class.

But as mentioned previously this data did not represent any information about the dispersion in the feature space and resulted in poor accuracies.

Keeping the above facts in mind a subset of training data was selected consisting of 28 pixels covering approximately all classes under consideration. These pixels were selected in such a way so that the neural network would be able to form decision regions in feature space using these pixel values (pixels in the boundary).

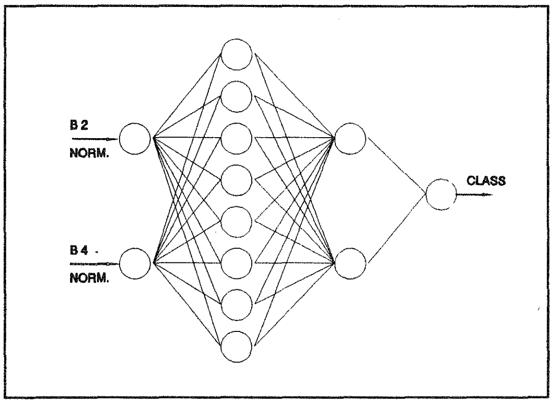

Fig. 4.1 Neural network model used for cloud and shadow detection.

The neural network was trained with this training data and after approximately 13000 epochs over the whole data set the network converged. The critical value of error was set to 0.09 and the learning rate was fixed at 0.1 with a momentum of 0.9 (1.0 - learning rate). But it is clear that selecting training data by manually inspecting them in feature space is unacceptable. It was also observed that the classification result was not very accurate due to the misclassifications at the boundaries between the certain cloud areas and clear sky areas.

The next step was to give the network input data not as normalized continuous values, but as binary data. So each data input (MOS band 2 and band 4) was split in to 8 bit vectors. This made the number of input nodes 16. The number of nodes in the first hidden layer was selected to be 48 and in the second hidden layer to be 5. Such an approach has been followed by other researchers also. With a 100 sample training data set it was possible to converge the network to an error of 0.09 in 150 epochs and to a minimum error of 0.01 in 450 epochs. This vast improvement can be attributed to the increase in the degrees of freedom in the network, which is caused by the increase in the number of weights. But it should be noted that this leads to poor generalization. The learning rate was fixed at 0.1 and the momentum parameter was (1.0 - learning rate).

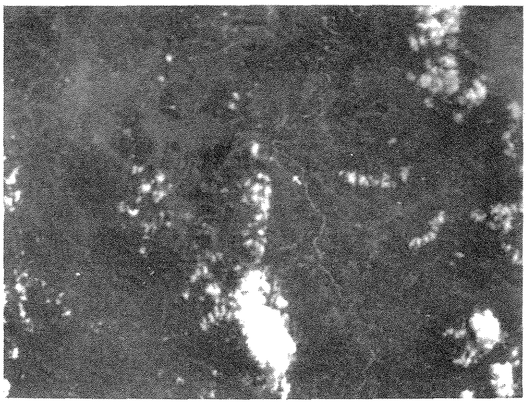

Fig. 4.2 Band 2 of image ait-3.img.

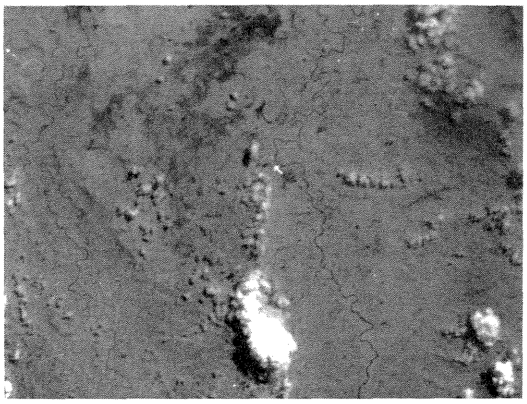

Fig. 4.3 Band 4 of image ait-3.img.

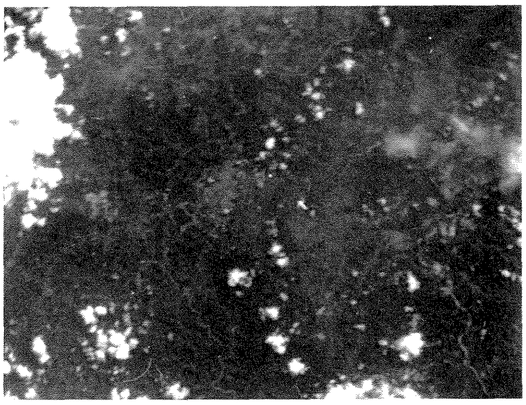

Fig. 4.4 Band 2 of image ait-4.img.

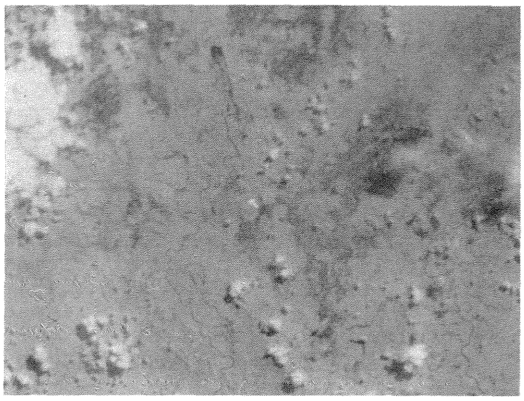

Fig. 4.5 Band 4 of image ait-4.img.

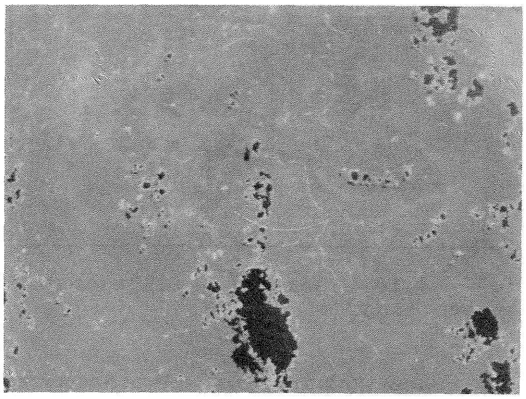

Fig. 4.6 Detected cloud and shadow areas superimposed on ait-3.img band 2.

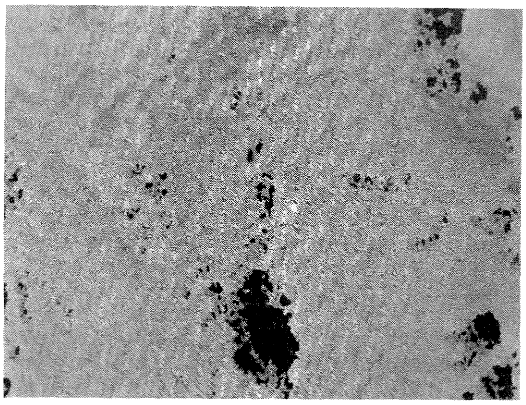

Fig. 4.7 Detected cloud and shadow areas super imposed on ait-3.img band 4.

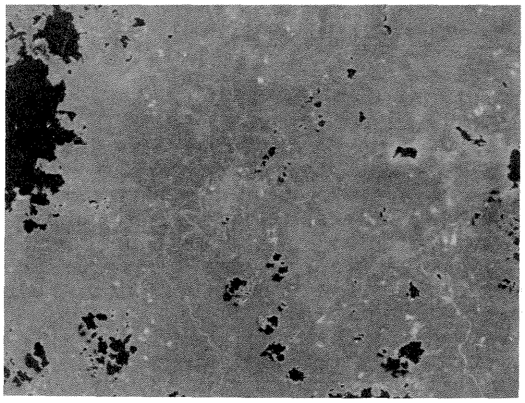

Fig. 4.8 Detected cloud and shadow areas superimposed on ait-4.img band 2.

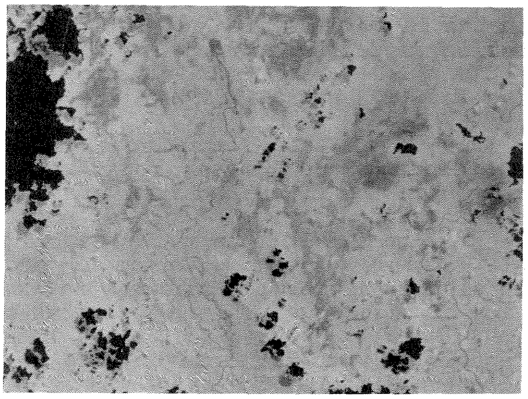

Fig. 4.9 Detected cloud and shadow areas superimposed on ait-4.img band 4.

4.2 Registering Multitemporal Images

The images used for making a cloud free mosaic are sub scenes of MOS-1 path 40 row 111w. The size of each image is 1024x800 pixels covering a ground area of approximately 5 km x 4 km. By visual inspection it was found that the images are not registered with each other.

Both images contained a significant amount of clouds and some rivers. The presence of bright clouds created additional problems. It eliminated the possibility of using simple thresholding techniques to separate bright objects in the images such as bridges and concrete structures. In such cases the thresholding mechanism could have been inhibited by using the cloud map created by the cloud detector, but any undetected cloud or water with sun glint (which is present in the image) would have still created problems. Upon observation it was found that there were not much bright objects that can be used as control points.

Water bodies presented an even worse problem. There was some wet land present in both images. Any multispectral classification method would have detected these wet land areas also as water bodies due to the similarity in their spectral signatures. As the main water bodies present in both images were rivers in principle they could have been detected using their elongatedness as a feature. A subsequent thinning and possible segmentation at points where the second derivative is zero would have produced segments that could have been matched automatically. But the spill of water out of the rivers made such techniques very complicated or almost impossible.

Under conditions where the true control point pairs are little compared with the false pairs any automatic point matching algorithm such as one proposed by TON and JAIN, would not have converged (TON and JAIN 1989).

Due to the above complications it was decided to use manual registration techniques. As a preliminary analysis three matching pairs of control points were selected in both images. Since in their vicinity there were no objects with similar spectral signatures as them in any of the bands, the area covered by them were extracted by local thresholding. Two of these objects were concrete bodies and the third was an old isolated river segment. The concrete bodies were extracted from band 2, and the river segment was extracted from band 4. The thresholds for bright objects were set at 92% of the local histogram (band 2) and for dark objects it was set at 8% of the local histogram (band 4). Their areas and centroids were calculated with respect to the image coordinate system in each image using a simple object tracking program. The results of these calculations are given in table 4.2.

By inspecting the two images it was seen that the control points in the two images match in pairs in the same order in which they are given above. Now the test given in table 4.3 was carried out in order to find the differences between the images. Also note that the differences in object areas are due to differences in scene brightness.

Since the difference in x and y coordinates almost remain constant for all three control points, it indicates that only translational differences are present in the two images. So the following functions were defined for the registering of the images.

if (x3,y3) is a point in image ait-3.img and the corresponding point in image ait-4.img is (x4,y4),

$$\begin{aligned} x4 &= x3 + dx \\ y4 &= y3 + dy \end{aligned} \qquad \textbf{(6)}$$

By using the control points given above, a least squares solution was obtained for equation (5.1) and it was found that dx = 273.6567 and dy = 184.9625. This solution was obtained using a least squares solution program provided with the book C Tools for Scientists and Engineers (BAKER 1989).

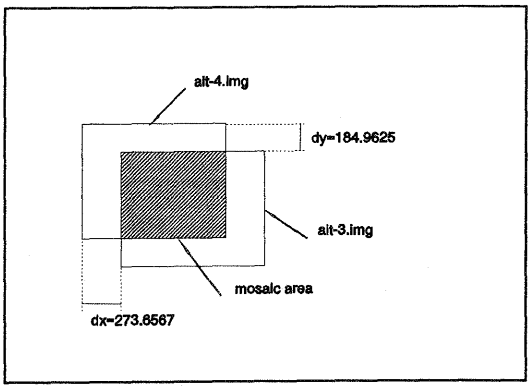

Fig. 4.10 Spatial relationship between images.

Using these values for displacements along x and y directions, the images were registered. Upon visual inspection it was found that the images appear to have registered very accurately. So no further effort was made to register the images.

4.3 Making Mosaic Images

To make the mosaic images cloud masks were formed for each image using the results of the classifications. As it was mentioned in the previous section, the two images ait-3.img and ait-4.img had only translational differences, and registering the two images with nearest neighbor interpolation, amounts to just matching only two corresponding points in the two images according to the nearest neighbor approximation and adjusting the other points accordingly.

The cloud mask of an image contained a bit for each pixel in the image and was set to 1 or 0 depending on whether a cloud or shadow had been detected at that pixel or not. A third mask called a stencil mask was formed by using the two cloud masks of the images by applying the following operations.

First the cloud masks of both images

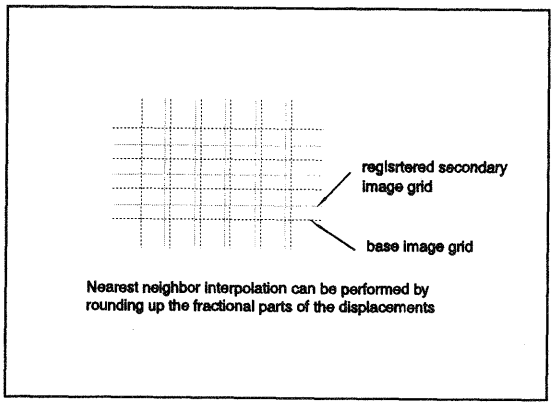
regisrtered secondary image grid
base image grid
Nearest neighbor interpolation can be performed by rounding up the fractional parts of the displacements

Fig. 4.11 Interpolating to a new grid.

were analyzed and the least contaminated image was selected as the base image. The contaminated parts of this image will be replaced by the corresponding parts of the other image whenever it is possible. By selecting the image that is cloud free as possible as the base image, we get a better image than if it was done the other way, since generally it reduces the number of mosaic operations required to be performed. In our case ait-3.img had 50798 pixels flagged as cloud/shadow contaminated and ait-4.img had 87039 pixels flagged as contaminated. So the base image chosen was ait-3.img. Then the cloud masks of the two images were registered and a new mask was formed with the following operation.

$$PIX_{st.m}=PIX_{b.m}\wedge\neg PIX_{sec.m} \qquad \textbf{(7)}$$

$PIX_{b.m}$ -pixel of the base image mask
$PIX_{sec.m}$-pixel of the secondary image mask
$PIX_{st.m}$ -pixel of the stencil mask

It is this mask that will be used to stencil in pixels from the secondary image. Note that the this stencil mask has 1, only at positions where the base image is contaminated and the secondary is not. Therefore we will not be replacing cloudy pixels with some other cloudy pixels.

The mosaic was constructed by replacing the cloud/shadow contaminated areas of the base image by the corresponding areas of the secondary image. In doing so the pixels in the replacement areas of the secondary image were themselves checked for cloud/shadow contamination, and the replacement was done only if they were uncontaminated. The main problem encountered here was that there were areas in which both images were cloud contaminated and in those areas the replacement was not carried out. Another problem was that the two images did not completely over lap and thereby reducing the area where the mosaic operation is possible. Although these shortcomings were present some cloud ares in the middle of the base image have been successfully removed.

TABLE 4.1 ACCURACY ANALYSIS OF NON-CONTEXTUAL CLASSIFIER

Scene Description	Total Num. of Pixels.	Num. of Cl/Sh. Pixels Detect.	Detection Accuracy
Bright thick cloud	280	280	100.00
Light cloud	154	154	100.00
Thin cloud	100	77	77.00
Shadow (dark)	195	192	98.46
Shadow (light)	143	109	76.22
Vegetation	342	0	100.00
River + Vegetated Land	456	0	100.00
Total			93.10

TABLE 4.2 POTENTIAL CANDIDATES FOR CONTROL POINTS

Image	Object description	Area in pixels	Centroid x coordinate	Centroid y coordinate
ait-3.img	Concrete structure (small)	48	140.166667	179.937500
	Concrete structure (large)	87	258.712644	269.390805
	Old river segment	83	616.771084	374.421687
ait-4.img	Concrete structure (small)	36	413.444444	365.361111
	Concrete structure (large)	69	532.637681	455.101449
	Old river segment	74	890.000000	560.364865

TABLE 4.3 AN ANALYSIS OF CONTROL POINTS

Object description	Difference in x coordinate	Difference in y coordinate
Concrete structure(small)	273.2778	185.4286
Concrete structure (large)	273.9250	185.7106
Old river segment	273.2290	185.9432

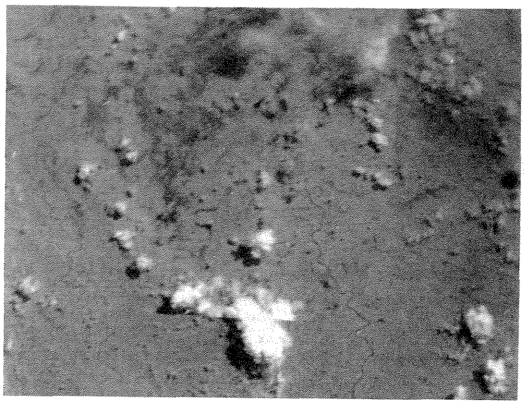

Fig. 4.12 Registered images

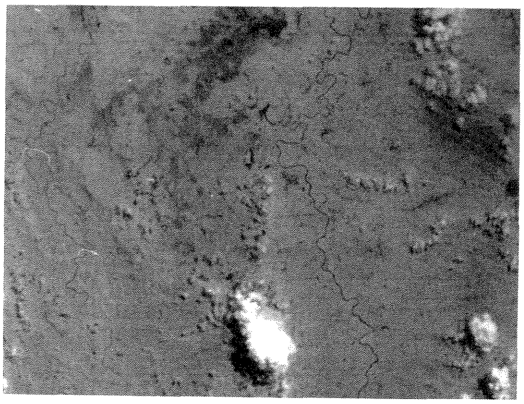

Fig. 4.13 A relatively cloud free mosaic

5. CONCLUSIONS

This paper has presented a method to make cloud free mosaic images from multitemtoral satellite images. The use of neural network classifiers to detect clouds has been explained. Further research can be done in achieving stability in training neural networks and predicting the optimal structure of such networks for good generalization. Research should also be directed at automatic registration under difficult conditions, like the conditions experienced by us.

REFERENCES

+References from journals

BENEDIKTSSON J.A., SWAIN P.H., ERSOY O.K.,(1990), Neural Network Approaches Versus Statistical Methods in Classification of Multisource Remote Sensing data, IEEE Transactions on Geoscience and Remote Sensing , Vol. 28, No. 4, pp. 540-552.

GOSHTABY A., JAIN A.K., ENSLIN W.R.,(1982), Automatic Image Registration , Symposium on Machine Processing of Remotely Sensed Data, Purdue University.

LIPPMAN R.L. (1987), An Introduction to Computing Using Neural Nets, IEEE ASSP Magazine No-4, pp. 4-22.

TON J., JAIN A.K.,(1990), Registering Landsat Images By Point Matching, IEEE Transactions on Geoscience and Remote Sensing , Vol. 27, No. 5, pp. 642-651.

WIDROW B., LEHR M.A. (1990), 30 Years of Adaptive Neural Networks : Perceptron, Madaline, and Backpropagation. Proceedings of the IEEE , Vol. 78, No. 9, pp. 1415-1442.

+References from books

BAKER L. (1989), C Tools for Scientists and Engineers, McGraw-Hill Inc.

McCELLAND J.L. and RUMMELHART D.E. (1988), Explorations in parallel distributed processing , MIT press , Cambridge, MA.

McCELLAND J.L. and RUMMELHART D.E. (1986), Parallel Distributed Processing, Explorations in the Microstructure of Cognition, Vol. 1, MIT press , Cambridge, MA.