WaveCluster for Remote Sensing Image Retrieval

Priti Maheshwary
Research Scholar, Department of Computer Application
Maulana Azad National Institute of Technology
Bhopal, MP, India

Namita Srivastava
Department of Mathematics & Computer Application
Maulana Azad National Institute of Technology
Bhopal, MP, India

***Abstract*— Wave Cluster is a grid based clustering approach. Many researchers have applied wave cluster technique for segmenting images. Wave cluster uses wave transformation for clustering the data item. Normally it uses Haar, Daubechies and Cohen Daubechies Feauveau or Reverse Bi-orthogonal wavelets. Symlet, Biorthogonal and Meyer wavelet families have been used in this paper to compare its clustering capacity and then results of this are used to retrieve the remote sensing image. The results are promising suggesting that Cohen Daubechies or Reverse Bi-orthogonal gives best result as compared to other four wavelet families.**

Keywords-Wave Cluster, Image Retrieval, Wavelet Families

I. Introduction

Wavelet transform is a signal processing technique that decomposes a signal into different frequency sub bands (for example, high frequency sub band and low frequency sub band). The wavelet model can be generalized to n-dimensional signals in which one dimensional transform can be applied multiple times. Methods have been used to compress data [1] or to extract features from signals (images) using wavelet transforms [2, 3, 4]. There are a number of basis functions that can be used as the mother wavelet for Wavelet Transformation. Haar wavelet is one of the oldest and simplest wavelet. Daubechies wavelets are the most popular wavelets. They represent the foundations of wavelet signal processing and are used in numerous applications. The Haar, Daubechies, Symlets and Coiflets are compactly supported orthogonal wavelets. These wavelets along with Meyer wavelets are capable of perfect reconstruction. The Meyer, Morlet and Mexican Hat wavelets are symmetric in shape. Bi-orthogonal family of wavelets exhibits the property of linear phase, which is needed for signal and image reconstruction. By using two wavelets, one for decomposition and the other for reconstruction instead of the same single one, interesting properties are derived.

Wave cluster applies wavelet transform on the featured space. It is a grid based approach which is efficient for large data bases [5]. Wavelet transform is applied iteratively multiple times, each time ignoring some details the average sub band and effectively increase the size of units neighborhood, resulting in clusters with different degree of details in each cluster. The processing time in this approach is independent of the number and distribution of the spatial objects in the space [6].

II. Methodology

Images are **acquired** and **features** of images are extracted such as color, texture and spectral. Then **Wave clustering** is applied to group similar images and stored in the **database**. When a query image is given then query image is matched with similar images in that particular group and on the basis of distance matrix most similar images are **retrieved and accuracy assessment** is done with the help of recall and precision value. Figure 1 shows the overall process of Content Based Remote Sensing Image Retrieval.

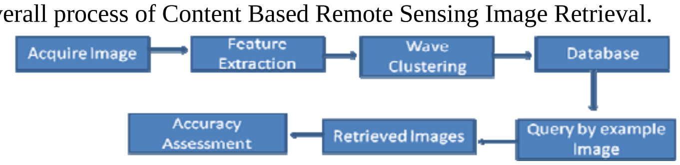

Figure 1: Content Based Remote Sensing Image Retrieval

III. Experimental Plan

For our experiments, we use 2 LISS III multi-spectral satellite images with 23.5m resolution. We choose to support 3 semantic categories in our experimental system, namely fields, water, and vegetation. In consultation with an expert in satellite image analysis, we choose near-IR (infra-red), red and green bands as the three spectral channels for classification as well as to display. The reasons for this choice are as follows:

Near-IR band is selected over blue band because of a somewhat inverse relationship between a healthy plant's reflectivity in near-IR and red, i.e., healthy vegetation reflects high in near-IR and low in red. Near-IR and red bands are keys to differentiating between vegetation types and states. Blue light is very abundant in the atmosphere and is diffracted all over the place. It therefore is very noisy. Hence use of blue band is often avoided. Visible green is used because it is less noisy and provides unique information compared to Near IR and red.

The pixels dimensions of each satellite image are used in our experiments are 5760x3597.The choice patch size is critical. A patch should be large enough to encapsulate the visual features of a semantic category, while being small enough to include only one semantic category in most cases. We choose patch size 128×128 pixels. We obtain 2800 patches from the images in this manner. These patches are stored in a database along with the identity of their parent images and the relative location within them. Ground truth categorization is not available readily for our patches.

We use six wavelet families namely Haar, Daubechies, Symlet, Biorthogonal, Reverse Biorthogonal or Cohen Daubechies and Meyer. We apply their transformation to find out six clusters. Though number of clusters is not required to be given in wave clustering but for similarity purpose, we use three clusters. Numbers of images in each cluster vary from different wavelet transformation. Table 1 shows the cluster number of first 25 images in different wavelet transform as Haar, Daubechies (db2), Symlet (sym3), Biorthogonal (bior1.3), Reverse biorthogonal (rbior1,3) and Meyer(Dmey). As Daubechies itself has different forms as db2, db4 etc., thus all forms of individual wavelet families are also experimented and most significant ones are then compared.

Table 1: Cluster numbers by each wavelet transform.

Imageid	Haar	db2	sym3	bior1.3	rbio1.3	Dmey
1	1	1	2	1	1	1
2	1	1	2	1	1	1
3	1	1	2	1	1	1
4	1	2	2	2	2	1
5	2	2	1	2	2	2
6	1	1	2	1	1	1
7	1	1	2	1	1	1
8	2	2	2	2	2	2
9	1	2	2	2	2	1
10	1	1	2	1	1	1
11	1	1	2	1	1	1
12	1	2	2	2	2	1
13	1	2	2	2	2	1
14	1	2	2	2	2	1
15	1	2	2	2	2	1
16	1	2	2	2	2	1
17	2	2	2	2	2	2
18	1	2	2	2	2	1
19	1	2	2	2	2	1
20	2	2	2	2	2	2
21	1	2	2	2	2	1
22	2	2	2	2	2	2
23	2	2	2	2	2	2
24	2	2	1	2	2	2
25	2	2	1	2	2	2

Then retrieval is done on the basis of Manhattan Distance, i.e., images of same cluster is retrieved on the basis of their distances from other images of same cluster. The evaluation of the different trial runs is done using precision/recall criteria. Recall and Precision are set based measures. They evaluate the quality of an unordered set of retrieved images [8].

Precision is a measure of the ability of a system to present only relevant images and is calculated as

$$\text{Precision} = \frac{\text{Number of Relevant images Retrieved}}{\text{Total Number of Retrieved Images}}$$

Recall is a measure of the ability of a system to present all relevant images and is calculated as

$$\text{Recall} = \frac{\text{Number of Relevant images Retrieved}}{\text{Number of Relevant Images in the database}}$$

Table 2: No. of images retrieved on one execution

Wavelet Transform	No.of relevant images retrieved	No. of images retrieved	Precision
Haar	15	20	0.75
db2	18	20	0.9
sym3	10	20	0.5
bior1.3	16	20	0.8
rbio1.3	19	20	0.95
Dmey	14	20	0.7

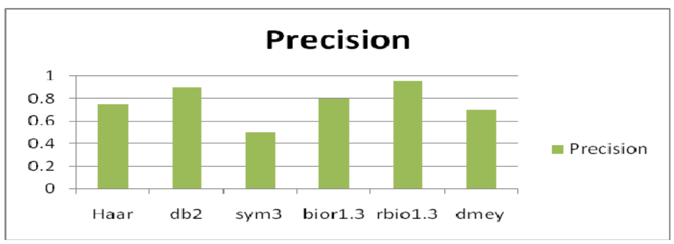

Figure 2: Precision graph for Haar, db2, sym3, bior, rbior and Meyer.

Table 3: Recall for images in Cluster 1

Number of Relevant *I*mages in the Database					
Wavelet	Cluster 1	Cluster 2	Cluster 3	No.of Relevant images retrieved on the basis of Cluster 1	Recall on the basis of Cluster 1
Haar	1008	1484	308	304	0.3
db2	532	2100	168	213	0.4
sym3	700	1932	168	322	0.46
bior1.3	588	2044	168	382	0.65
rbio1.3	588	2044	168	235	0.4
dmey	1000	1450	350	550	0.55

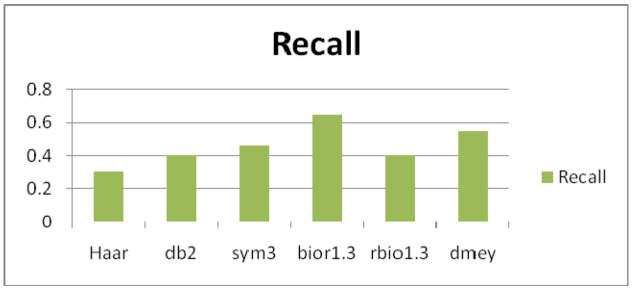

Figure 3: Precision graph for Haar, db2, sym3, bior, rbior and Meyer.

For the analysis Haar, Deubuchies, Symlet, Biorthogonal, Reverse Biorthogonal and Meyer wavelet transformation of which are applied, the result is shown in table 2 and 3 and figure 2 and 3. Table shows the number of retrieved images in each run and number of relevant images on each run of different wavelet transformation and the value of Precision is calculated on the basis of formula shown above. Out of the six transformations the reverse biorthogonal or Cohen Deubuchies Wavelet gives best results for image retrieval while Deubuchies (dbN) and Biorthogonal are next to this.

IV. Conclusion

In this paper, we have presented Wave Cluster for remote sensing image retrieval. Our experimental results demonstrated that Wave Cluster can be used for image retrieval. We have applied Haar, Deubuchies, Symlet, Biorthogonal, Reverse Biorthogonal and Meyer wavelet transformation, as the result indicate Reverse Biorthogonal gives best result and after that Deubuchies and Biorthogonal gives best result. On the Recall and Precision graph the value of precision for Reverse Biorthogonal transform is 0.95 and for Deubuchies transform it is very close to Reverse Biorthogonal that is 0.9 and recall for the two transformation is also 0.4 which are very much bettere than other transforms. This new approach will provide extremely valuable information to GIS, digital city constructions and other urban applications.

REFERENCES

[1] Michael L. Hilton, Bjorn D. Jawerth, and Ayan Sengupta. Compressing Still and Moving Images with Wavelets. Multimedia Systems, 2(5):218-227, December 1994

[2] J. R. Smith and S. Chang. Transform Features For Texture Classification and Discrimination in Large Image Databases. In Proceedings of the IEEE International Conference on Image Processing, pages 407- 411,1994.

[3] G. Sheikholeslami and A. Zhang. An Approach to Clustering Large Visual Databases Using Wavelet Transform. In Proceedings of the SPIE Conference on Visual Data Exploration and Analysis IV, San Jose, February 1997.

[4] G. Sheikholeslami, A. Zhang, and L. Bian. Geographical Data Classification and Retrieval. In Proceedings of the 5th ACM International Workshop on Geographic Information Systems, pages 58-61, 1997.

[5] Berthold Klaus Paul Horn. Robot Vision. The MIT Press, forth edition, 1988.

[6] Sheikholeslami, G. and Chatterjee, S. and Zhang, Wavecluster: A multi-resolution clustering approach for very large spatial databases, Proceedings of the International Conference on Very Large Data Bases,428—439, 1998

[7] Common evaluation measures by Trec-Evaluation Report.