Multiscale methods and their applications

Challa S. Sastry and Rao Vemuri

Abstract

As in the decimal system of representation of numbers, details are added successively to coarser level information in wavelet decomposition to improve upon the accuracy in representation. This progressive reconstruction feature and the localization property of wavelets are critical in many scientific applications. The present article being tutorial in nature discusses applications of multiscale techniques provided by wavelets in such areas as network security, data mining, information retrieval and computerized tomography.

1. Introduction

Wavelets [2] became popular because they permit a totally new perspective in data analysis; they permit analysis according to scale. The fundamental idea behind wavelets is similar to that of Fourier analysis where one strives to approximate functions using superposition of sines and cosines. However, in wavelet analysis, the scale we use to look at data plays a special role. Wavelet methods permit us to look at data at different scales or resolutions. If we look at a signal with a large "window," we would notice gross features. If we look at a signal with a small "window," we would notice small features.

By definition, sines and cosines are non-local (they stretch out to infinity). They therefore do a very poor job in approximating sharp spikes. With wavelet analysis, we can use approximating functions that are contained in finite domains. Therefore, wavelets are well-suited for approximating data with sharp discontinuities.

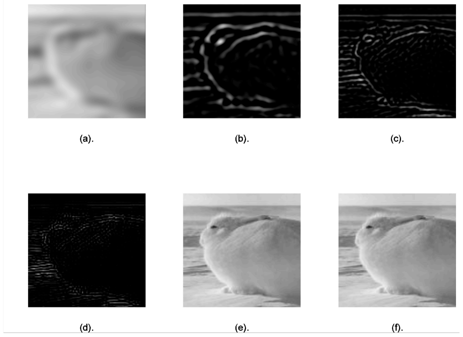
In wavelet methods one starts with a prototype wavelet function, called an analyzing wavelet or mother wavelet. For analysis in the time-domain, one works with a contracted, high-frequency version of the prototype wavelet, while frequency domain analysis is performed with a dilated, low-frequency version of the same wavelet. Because the original function can be represented in terms of a wavelet expansion, data operations can be performed using just the corresponding wavelet coefficients. If you further choose the best wavelets adapted to your data, or truncate the coefficients below a threshold, your data is sparsely represented. This sparse coding makes wavelets an excellent tool in the field of data compression [10][11]. Due to easy accessibility of software packages that contain fast and efficient algorithms to perform wavelet transforms, wavelets have quickly gained popularity among scientists and engineers working on both theoretical issues and applications. Above all, wavelets have been widely applied in such computer science areas as image processing, computer vision, network management and data mining [4][12]. In the subsequent sections, the paper discusses wavelets and some of its real life applications in fields such as information retrieval, computerized tomography, data networks and data mining.

2. A brief introduction to wavelets

Wavelets have many favourable properties, such as compact support, and can generate

Figure 1: Decomposition of an image in (f) into wavelet basis. Images in (a) to (d) are approximation and detail parts, while the image in (e) is sum of images in (a) to (d).

different classes of (wavelet) bases. The property of wavelets being finite in duration implies the localization feature of them. The main advantage of



this feature is that the presence of local error (noise) in the data reflects local changes in wavelet coefficients, unlike the Fourier technique wherein the local change in data has global effect on the Fourier coefficients. This feature along with multiresolution feature is widely being used in image/signal/pattern analysis [6][8][9]. To add to this, in contrast to the Fourier technique wherein one uses *sine* and *cosine* functions to generate representation of a suitable function, in wavelet technique, different wavelet bases like orthonormal, biorthogonal (symmetric), multiwavelets, wavelet packets etc are constructed [2] to tackle various applications. The decomposition of an image into wavelet basis is shown in Figure 1.

The choice of wavelet basis can be made depending on the requirement. For example, in image compression, boundary value problems, biorthogonal (symmetric) wavelets are found to be useful. In some of the feature extraction algorithms, orthonormal wavelets are found to be useful [2]. Besides the stated properties, wavelet bases possess other properties like zero moments, hierarchical and multi resolution framework, and `decorrelated' coefficients. These features could provide considerably more efficient and effective solutions to many practical problems.

3. Application of wavelets

The present section studies some of the applications of wavelets in fields such as data mining, data networks, computerized tomography and information retrieval.

3.1 Application in image retrieval

Content-Based Image Retrieval (CBIR) refers to the recall of images from a database that are relevant to a query, using information derived from the images themselves, rather than relying on accompanying text indices or other annotation. CBIR has received increasing attention as a result of the availability of large image databases in medicine, science, commerce, and the military. CBIR has been proposed to overcome the difficulties encountered in textual annotation for large image databases. Like a text-based search engine, a CBIR system aims to retrieve information that is relevant (or similar) to the users query. Extracting features from raw images, the CBIR system generates a ‘compact’ (called *signature*) vector form of them. Then using some similarity measure between a query image and database images based on these features, it retrieves the images that are similar in content to the query image. A typical CBIR system [6] is shown in Figure 2.

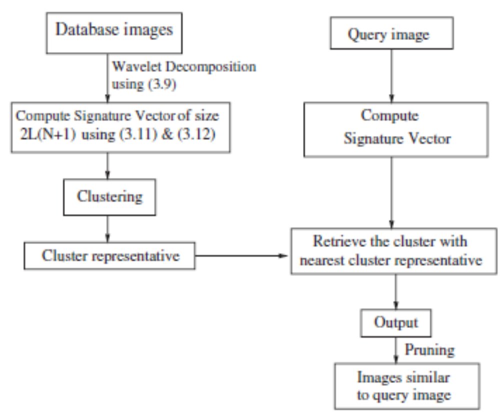


Figure 2: A typical CBIR system

The wavelet transform with its localization and multiscale frame work has been found to be a very useful tool for CBIR applications. There have been different methods proposed for CBIR using wavelets [6][7], wherein the compact representation of an image is generated by taking the statistical elements, mean and variance, of wavelet coefficients at different levels. The use of the features in CBIR requires the use of an appropriate similarity measure for comparing the query image with those in the database. Several similarity measures based on common distance functions such as Euclidean, Mahalonobis etc are defined in literature [4][12]. The retrieval performance of a wavelet based CBIR system proposed in [6] is shown in Figure 3.

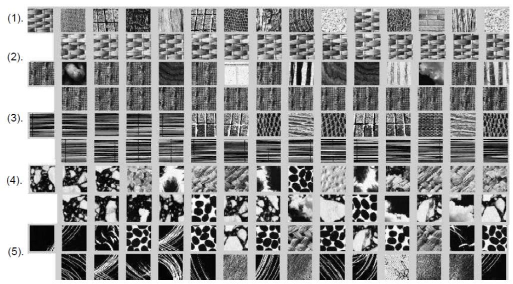


Figure 3: Comparison of retrieval performances of Gabor and wavelet based methods. Images on the very first column represent query images. Corresponding to each query image, the images on first and second rows are respectively those retrieved by Gabor and wavelet based methods (see [6], for details). This figure concludes that the wavelet based methods show relatively better performance.

But in applications like remote sensing, same image may be present in image database with different orientations. Hence, when classifying or retrieving images by query, all those images that are present at different orientations must be retrieved or classified into one class. Therefore, the method of extraction of features must be invariant to the presence in images of rotation. Radial type wavelets could be used in making signature vectors rotation invariant [6][7].

3.2 Application in Computerized Tomography

The basic objective in Computerized Tomography (CT) is to obtain the high quality images from projection data obtained using different scanning geometries (Figure 4) such as parallel, fan and cone or spiral cone beam geometries with as less exposure and utmost efficiency.

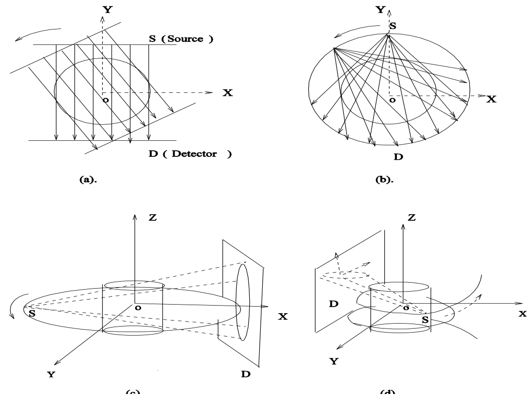


Figure 4: Data acquisition in tomography in different scanning geometries.

For a large class of medical imaging problems, one needs to reconstruct only a small interior portion of a cross section using the data pertaining almost to that part of the interior region. This problem, called local tomography, throws a need for the development of suitable algorithms that can be used for the reconstruction of local regions with local data and low computational power and complexity. It is, however, proved [5] that the exact local reconstruction from localized data is not possible. In addition, the local reconstruction has null space problem, in the sense that even if the underlying density function is not identically zero over an interior region, the associated projection data can be zero identically. These problems could be addressed to certain extent using wavelet framework. The local reconstruction problem and the way wavelets carry out the localization of data is shown in the Figure 5.

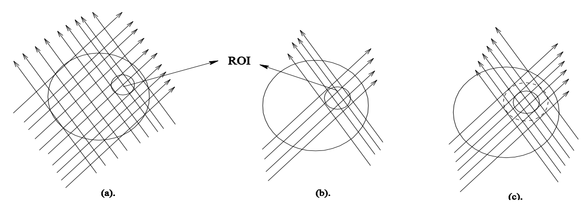


Figure 5: (a). Standard Fourier based reconstructions require global data even for local recovery, (b). Ideally localized data is desired to be used local recovery and (c). Wavelet based methods provide good approximation to local recovery through almost localized data.

The Fourier based techniques are popularly called convolution backprojection (CBP) methods. The ramp filter associated with CBP is responsible for global data to be used for local recovery. The wavelet weighted ramp filters have reduced duration and hence provide local reconstructions using localized projection data. A wavelet based local reconstruction is shown in Figure 6 (See [9] for details).

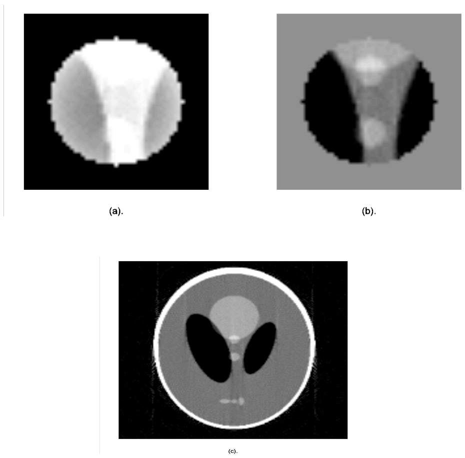


Figure 6: Local reconstruction using (a). localized data through Fourier based method (CBP), (b). almost localized data through wavelet based method and (c). Shepp Logan Phantom (test) image.

3.3 Application in Data Mining

In this subsection, we briefly talk about some applications of wavelets in data mining, which are useful in the analysis of data networks.

The basic objective in dimensionality reduction is to keep the information content of a larger data set in a smaller data set. As the wavelet transform breaks up a function or a data set into different frequency components, wavelets can achieve [4][12] the dimensionality reduction by projecting the data set into frequency spaces of lower dimension or by retaining significant wavelet coefficients. The first case involving the projection of data into lower resolution spaces is equivalent to taking first few coefficients in the wavelet representation of data set. Although, this approach is useful for easy indexing, it works well when the information content of the data set is present in first few levels and that in higher resolution levels is insignificant. While, the second case, involving retaining few larger coefficients in wavelet domain results in very little loss of information in data [12]. This process involves arranging coefficients in decreasing order and then taking first few, as dictated by the preassigned error tolerance between the energy of data set and that of retained coefficients. One may use data sets in wavelet domain after reducing dimension for similarity search. An excellent overview on the application of wavelets to similarity search has been given in [4].

The aim of data clustering methods is to group objects in databases into meaningful subclasses. Due to the huge amount of data in use, an important challenge for clustering algorithms is to achieve good time efficiency. Using the multiresolution property of wavelet transforms, G. Sheikholeslami et al. [12] proposed an algorithm, called WaveCluster, for clustering very large databases. The WaveCluster considers the multidimensional data as a multidimensional signal, and applies wavelet transform to convert data into the frequency domain. It then convolves the wavelet domain data with an appropriate kernel function, which results in a transformed space where the natural clusters in the data become more distinguishable. Finally it identifies the clusters by finding the dense regions in the transformed domain. It has been experimentally observed that the WaveCluster outperforms some of the standard clustering algorithms.

3.4 Application in network security

Design of intrusion detection systems (IDSs) has been an active area of research for more than a decade due to the increasing rate of attacks on computer systems. There are two families of techniques to build an IDS [1][3][8]: The first is Misuse-based IDS which works on the signatures of known attacks and thus can not capture new attacks. The second is anomaly-based IDS that learns the normal behavior of a system (viz. users, computer networks or programs), and any deviation in this behavior is considered as a probable attack. Intrusion detection systems based on the later technique are generally capable of detecting new attacks. Based on the data being analyzed by the IDS to detect an intrusion, there are host-based IDS (HIDS) and network-based IDS (NIDS). HIDS collects data from the system it is protecting, while NIDS collects data from the network, usually in the form of packets.

It has been observed that Internet traffic is self-similar in nature. Self-similarity is the property that is associated with the objects whose structure is unchanged on different scales. It is pointed out [1] that the self-similarity of Internet traffic distributions can often be accounted for by a mixture of the actions of a number of individual users, and hardware and software behaviors at their originating hosts, multiplexed through an interconnection network. It has been observed that the traffic related to attacks, especially DoS attacks, is bursty in nature. Traffic that is bursty on many or all time scales can be described statistically using the notion of self-similarity.

In view of its multiscale framework and localization aspects, the wavelet technique is capable of being used for the analysis of scaling and local ``burstiness'' features of a function. In recent literature on wavelet based IDS, the wavelet analysis has been used both to describe and analyze data network traffic [9]. Besides, it has been used to characterize network related problems like congestion, device failure etc. The basic philosophy in wavelet based IDS is that self-similarity is prevalent in the network traffic under normal conditions, and therefore can be considered as a signature for normal behavior. The loss of self-similarity, signifying a possible attack, can be taken as a deviation from normal behavior.

The basic idea in the wavelet based approaches [3][9] is that, when data satisfies scaling property, the corresponding wavelet coefficients and hence the energies of them at different levels also satisfy the same scaling property (with different exponents). Analyzing the exponents of scaling property of energies at different levels, one concludes whether or not the data satisfies the scaling property.

References:

[1] Crovella M and Bestavros A , “Self-Similarity in World Wide Web Traffic: Evidence and Possible Causes," IEEE-ACM Transactions on Networking, 5:835-846, 1997.

[2]. I. Daubechies, “Ten Lectures on Wavelets,” Capital City Press, Montpelier, Vermont, 1992.

[3] A. C. Gilbert, “Multiscale Analysis and Data Networks," Applied and Computational Harmonic Analysis, 10:185-202, 2002.

[4] T. Li and Q. Li, S. Zhu and M. Ogihara, “A Survey of Wavelet Applications in Data Mining," SIGKDD Explorations, 4(2):49, 2002.

[5]. F. Natterer, “Mathematical methods in image reconstruction,” SIAM 2001.

[6]. C. S. Sastry, A. K. Pujari, B. L. Deekshatulu, C. Bhagvati, “A wavelet based multiresolution algorithm for rotation invariant feature extraction,” Pattern Recognition Letters, Vol. 25, Issue 16, pp: 1845-1855, 2004.

[7]. C. S. Sastry, M. Ravindranath, A.K. Pujari, B. L. Deekshatulu, “A modified Gabor function for content based image retrieval," Pattern Recognition Letters, Vol. 28, 2, pp: 293-300, 2007.

[8]. C.S. Sastry, S. Rawat, A.K. Pujari, V.P. Gulati, “Network traffic analysis using singular value decomposition and Multiscale transforms," Information Sciences, Vol. 177, pp: 5275-5291, December, 2007.

[9]. C.S. Sastry, P.C. Das, “Wavelet based multilevel backprojection algorithm for parallel and fan beam scanning geometries,” Int. J. of Wavelets, Multiresolution and Information Processing, Vol. 4, N0. 3, pp: 523-545, 2006.

[10]. S. Saha and V. Vemuri, "Image Categorization and Coding using Neural Nets and adaptive Wavelet Filters,” Invited paper, IEEE ICIT-2000, pp 438-443, Goa, India, January 19-22, 2000.

[11]. S. Saha and Rao Vemuri, "Use of adaptive integer-to-integer wavelet transforms in lossless image coding," in Proc. IEEE-CAS, Sydney, Vol. 2, pp. 393-396, 2001.

[12]. G. Sheikholeslami, S. Chatterjee and A. Zhang “WaveCluster: A Multiresolution Clustering Approach for Very Large Spatial Databases," Proc. 24th Int. Conf. Very Large Data Bases, VLDB, pp 428-439, 1998.