W-transform method for feature-oriented multiresolution image retrieval

Man Kam Kwong and Biquan Lin

Mathematics and Computer Science Division

Argonne National Laboratory

Argonne, IL 60439-4844

E-mail: kwong@mcs.anl.gov, blin@mcs.anl.gov

ABSTRACT

We present a local-feature-oriented image indexing and retrieval method based on Kwong and Tang's *W*-transform. Multiresolution histogram comparison is an effective method for content-based image indexing and retrieval. However, most recent approaches perform multiresolution analysis for whole images but do not exploit the local features present in the images. Since *W*-transform is featured by its ability to handle images of arbitrary size, with no periodicity assumptions, it provides a natural tool for analyzing local image features and building indexing systems based on such features. In our approach, the histograms of the local features of images are used in the indexing system. The system not only can retrieve images that are similar or identical to the query images but also can retrieve images that contain features specified in the query images, even if the retrieved images as a whole might be very different from the query images. The local-feature-oriented method also provides a speed advantage over the global multiresolution histogram comparison method. The feature-oriented approach is expected to be applicable in managing large-scale image systems such as video databases and medical image databases.

Keywords: *W*-transform, multiresolution, histogram, image indexing and retrieval, feature

1. INTRODUCTION

Image database management is important in the development of multimedia technology. Since an enormous amount of digital images is likely to be generated within the next few decades in order to integrate computers, television, VCR, cables, telephone and various imaging devices, effective image indexing and retrieval systems are urgently needed so that images can be easily organized, searched, transmitted, and presented.

Traditional image retrieval methods are built on indexing systems that use keywords to describe image objects and attributes. This approach becomes ineffective in large-scale image

DISCLAIMER

Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.

database management. The reasons are as follows. First, building such indexing systems is tedious because it requires enormous database entries to describe image objects and attributes and the building processes cannot be automated. Second, it is difficult (and almost impossible in some cases) to design a scene description language to describe the rich information contained in images. The solution to overcome these difficulties is to build indexing and retrieval systems directly based on the content of images.

Recently, several content-based image retrieval methods have been studied. Among them, Pentland et al. [5] used a Karhunen-Loeve (KL) transform to obtain a compact description of the most significant features present in the image, such as brightness, edges, and textures. Pentland and his colleagues were then able to search images, find objects inside the images, and compare textures of the images directly based on the KL transform coefficients. Since KL transform has an extremely high energy compaction, a KL transform-based image indexing and retrieval method effectively preserves most of the energy of the images. One drawback of this approach, however, is its large amount of time required to obtain KL transform coefficients since KL transform is computationally demanding. Lee and Dickinson [4] built a video indexing system by exploiting histograms of multiresolution decompositions of images. Histogram comparison is a method robust to image displacement, orientation, and some amount of distortion. Multiresolution decomposition gives a hierarchical image structure so that one can manipulate images in different levels according to different requirements. The method combining these two important techniques is anticipated to be efficient. The approach, however, performs multiresolution decomposition for whole images but does not use local image features. Since local image features are important for understanding and analyzing images, it is natural to extend the approach to design a local-feature-based image indexing and retrieval system.

The main difficulty in building a local-feature-oriented multiresolution analysis is the lack of suitable tools. Most existing wavelets, such as Haar wavelets and Daubechies' wavelets [1] require that the size of images be a power of 2 and assume that periodical extension is possible beyond the boundaries of images. To satisfy the image size constraint, one needs to pad zeroes before performing multiresolution decomposition. Doing so not only introduces more storage requirements and computational burdens but also dilutes the features of the images. Further, while assuming periodicity for the whole images is not harmful, it is unreasonable to assume that all local areas are periodically extendable. What is needed is a wavelet that does not have a size requirement or periodicity constraint.

Recently, Kwong and Tang [2] introduced W -transforms for multiresolution analysis. The W -transform is featured by its smoothness, its simple form of representation, its ability to handle arbitrary image sizes, and its exemption of periodicity assumptions. The last two properties make W -transform a natural tool for analyzing local image features and for building indexing systems based on the local features. For those who are interested in experimenting with W -transforms by hand, the MATLAB version of the W -transform source code can be found in [3].

This paper presents a local-feature-oriented image retrieval method using Kwong and Tang's W -transforms. In this approach, the multiresolution histograms of the local feature images are used as an index. The images with the specified features are retrieved by comparing the multiresolution histograms of the local features of the query images with the indexing histograms. Our numerical experiments in implementing this method indicated that the feature-oriented approach is practical and efficient.

The remainder of this paper is organized as follows. Section 2 introduces Kwong and Tang's W -transforms for multiresolution analysis. Section 3 discusses the multiresolution histogram comparison method for image analysis. Section 4 presents our new local-feature-oriented image indexing and retrieval method. Section 5 gives some numerical experiments. Finally, Section 6 summarizes the advantages of our new method.

2. W -TRANSFORMS AND MULTIREOLUTION ANALYSIS

First, we illustrate multiresolution decomposition and reconstruction of signals in the one-dimensional case. Let $f = (f_0, f_1, f_2, \dots, f_{N-1})$ denote a one-dimensional signal. Suppose G and H are $K \times N$ matrices and G^+ and H^+ are $N \times K$ matrices with the property that

$$G^+G + H^+H = I, \quad (1)$$

where I denotes an $N \times N$ identity matrix. Then, the signal f can be decomposed into two components u and v by linear transforms

$$\begin{aligned} u &= Gf \\ v &= Hf \end{aligned} \quad (2)$$

and f can be reconstructed from the components u and v by

$$f = G^+u + H^+v. \quad (3)$$

For some specifically selected G and H , the components u and v are often referred to as the lowpass component and highpass component respectively. If we apply some matrices in a form similar to G and H to the lowpass component u , we obtain lowpass and highpass components of u . Repeating this process, we obtain a sequence of highpass components and one lowest component. On the other hand, the original signal can be reconstructed from these components in the order that is the reverse of the preceding decomposition process. This is the basic idea of multiresolution decomposition and reconstruction.

Many matrices can be used for such a decomposition and reconstruction. Some matrices G and H are associated with important transforms in signal processing. Here we give two examples, one is called Haar transform and the other is called Daubechies D_4 transform.

A Haar matrix pair G and H is a pair of $N/2 \times N$ matrices defined by

$$G = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 1 & 1 & \dots & 0 & 0 \\ & & & & \dots & & \\ 0 & 0 & 0 & 0 & \dots & 1 & 1 \end{bmatrix} \quad (4)$$

and

$$H = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 1 & -1 & \dots & 0 & 0 \\ & & & & \dots & & \\ 0 & 0 & 0 & 0 & \dots & 1 & -1 \end{bmatrix}. \quad (5)$$

The associated matrices G^+ and H^+ are defined by

$$\begin{aligned} G^+ &= G, \\ H^+ &= -H. \end{aligned} \quad (6)$$

It is easy to verify that such G , H , G^+ and H^+ satisfy (1) so that it is possible to perform decomposition and reconstruction as (2) and (3).

The Haar transform has a very simple form. A feature of the Haar transform is its orthogonality, that is, $GH^T = HG^T = 0$. The orthogonality is important in some signal processing applications. For example, it can provide a complete representation of the original signal without redundancy, and it will not introduce correlation of noise after performing transforms for signals that are contaminated by independent noise. Drawbacks of the Haar transform include its lack of smoothness and its size requirement for the original signal. The lack of smoothness means that its associated scaling function is not continuous; hence, a smooth signal will be mapped to a signal with sharp cusps. The size requirement means that the original signal is required to have a size of power of 2, thereby making the Haar transform unsuitable for analyzing local features.

Daubechies [1] proposed a method to generate high-order orthogonal matrices that can be used to perform multiresolution decomposition and reconstruction. Specifically, she gave a pair of matrices with order 4 (called D_4 matrices in literature). We refer the reader to [1] and [2] for the detailed expression of such matrices. In contrast to the Haar transform, the scaling function associated with the D_4 transform is continuous. But the D_4 transform still requires that the original signal have a size of power of 2. In addition, it requires a periodic extension beyond the boundary of the original signal to obtain a complete reconstruction. Therefore, the D_4 transform is not a good choice for analyzing local feature images either.

Now, we turn to the W -transform. According to [2], an arbitrary vector $g = [g_1, g_2, g_3, g_4]$ and two constants c and d are first chosen. Then,

$$h = [h_1, h_2, h_3, h_4] = [g_1/c, g_2/c, g_3/d, g_4/d] \quad (7)$$

together with g is used to construct W -matrices for signals with even or odd sizes. For a signal with an even size, the corresponding W -matrix pair G and H is

$$G = \begin{pmatrix} g_1 + g_2 & g_3 & g_4 & & & \\ & g_1 & g_2 & g_3 & g_4 & \\ & & & g_1 & g_2 & g_3 & g_4 \\ & & & & & \ddots & \\ & & & & & & g_1 & g_2 & g_3 + g_4 \end{pmatrix} \quad (8)$$

and

$$H = \begin{pmatrix} h_1 + h_2 & h_3 & h_4 & & & \\ & h_1 & h_2 & h_3 & h_4 & \\ & & & h_1 & h_2 & h_3 & h_4 \\ & & & & & \ddots & \\ & & & & & & h_1 & h_2 & h_3 + h_4 \end{pmatrix}. \quad (9)$$

For a signal with an odd size, the corresponding W -matrix pair G and H is

$$G = \begin{pmatrix} g_1 + g_2 & g_3 & g_4 & & & \\ & g_1 & g_2 & g_3 & g_4 & \\ & & & g_1 & g_2 & g_3 & g_4 \\ & & & & & \ddots & \\ & & & & & & g_1 & g_2 & g_3 & g_4 \\ & & & & & & & h_1 & h_2 + h_3 + h_4 \end{pmatrix} \quad (10)$$

and

$$H = \begin{pmatrix} h_1 + h_2 & h_3 & h_4 & & & \\ & h_1 & h_2 & h_3 & h_4 & \\ & & & h_1 & h_2 & h_3 & h_4 \\ & & & & & \ddots & \\ & & & & & & h_1 & h_2 & h_3 & h_4 \end{pmatrix}. \quad (11)$$

An interesting and useful property of such W -matrix pair is that its corresponding G^+ and H^+ can also be generated by a pair of vectors of length 4. Therefore, the original signal can be reconstructed by the decomposed local components. A specific set of such matrices given in [2] is built by $g = [-1, 3, 3, -1]$ and $h = [-1, 3, -3, 1]$.

We can see here that W -matrices possess two interesting features among many others. First, the W -matrices can be applied to signals with arbitrary sizes, even signals with odd sizes. This feature is not possessed by classic wavelet matrices such as Haar matrices and D_4 matrices. Second, W -matrices do not require periodic extension for the original signal. The transformed signal in some location is related only to the signals at its neighboring locations. Thus, W -matrices are suitable for analyzing local signals.

A two-dimensional signal (an image) can be expressed by a matrix. The 2D W -transform of a matrix is obtained by first applying the 1D W -transform to each column of the matrix and then applying the same W -transform to each row of the resulting matrix. This procedure generates four components (GG, GH, HG, HH) for a matrix. For example, the GG component is obtained by applying G to each column and each row of the matrix. For the component GG, we can also obtain its four components by applying W -transform to it. Repeating this process, we obtain a multiresolution decomposition of the original image, which can be used to reconstruct the original image by the reverse process. In the next section, we use these decompositions to analyze the original images and build a fast retrieval method for large-scale images.

3. MULTIREOLUTION HISTOGRAMS FOR IMAGE RETRIEVAL

Multiresolution decomposition of images generated by a wavelet transform such as the W -transform has been applied to several image processing areas such as image compression and image restoration. This paper explores a new application of wavelet transform. We apply wavelet transform for large-scale image management and analysis.

Recently, RAID (Redundant Arrays of Inexpensive Disks) has become popular for meeting high-capacity storage requirements for multimedia technology. In RAID, a dual data set is stored to avoid unexpected data loss. For large-scale images, we believe, it will be more effective if the dual data set is stored as multiresolution decompositions of images instead of the exactly same images. The reasons are as follows. First, the original image can be quickly and completely reconstructed from the multiresolution decompositions. Therefore, saving the multiresolution decompositions of images is equivalent to saving the original images. Second, the multiresolution decompositions can be exploited for quick image retrieval.

A typical medical image such as a mammogram has $2K \times 2K$ pixels and each pixel requires 2 bytes (usually only 10 bits are used). To store such an image requires 10 Mbytes. Usually, a radiologist needs to compare 10 to 100 images at one time. Suppose a radiologist requests 20 mammograms through Ethernet, whose transmission rate is 1.25 Mbytes per second. The radiologist must wait for about 4 to 5 minutes to get the requested images. Also, some or all of the retrieved images often are not what is really wanted. So the radiologist needs wait another several minutes to get another group of images. This situation can be solved by storing multiresolution decompositions of images. For example, if the images are decomposed up to scale 4, then the lowest components have size of 128. A strategy to solve the preceding problem is retrieving the lowest components first and seeing whether the images are exactly what is wanted. If they are, the original images may be retrieved. Since the time for retrieving the lowest components is 256 times less than that for retrieving the whole images, the response time for retrieving 20 the lowest components will be reduced to 2 to 3 seconds, clearly within the tolerance level. Therefore, using this strategy can avoid unnecessary waiting.

Histograms are a useful tool for image comparison. Let $\mathbf{f} = \{f(i, j), i, j = 1, 2, \dots, N\}$ denote an image with pixel values from T_{min} to T_{max} . Let $T_{min} = t_0 < t_1 < t_2 < \dots < t_K = T_{max}$. The histogram $\mathbf{h} = \{h(k), k = 0, 1, 2, \dots, K-1\}$ of $\mathbf{f}$ is defined as the number of (i, j) 's with $t_k \leq f(i, j) < t_{k+1}$. Let $\mathbf{h}_1$ and $\mathbf{h}_2$ be the histograms of images $\mathbf{f}_1$ and $\mathbf{f}_2$ respectively. The difference of histograms (DOH) of $\mathbf{f}_1$ and $\mathbf{f}_2$ is defined by

$$DOH = \sum_{k=0}^{K-1} |h_1(k) - h_2(k)|.$$

Two images will have the same histogram if one is obtained by displacing or rotating the other. Histograms of two images will not differ very much if one of the images is degraded from the other by noise. But calculating the histogram of a $2K \times 2K$ image is still very costly. Since the lowest component of the multiresolution decomposition of an image gives a compact description about the image features, comparing histograms of the lowest components should distinguish images in most cases. If a $2K \times 2K$ image is decomposed to scale 4, the cost of calculating the histogram of the lowest component is reduced by 256 times.

However, this approach does have two major drawbacks. First, it cannot search such images that are significantly different from the query image as a whole but contain local features identical with or similar to those in the query image. On the other hand, the local-feature-based search is desirable in some applications. For example, in computer aided diagnosis (CAD), radiologists often compare local image features rather than whole images. Second, The multiresolution histogram comparison method still requires a large amount of computational time to compute histograms for an image database that contains many large images. For example, for each $2K \times 2K$ mammogram, we still need to compute histogram for the coarsest component with size of 128×128 if we decompose the image up to scale 4. If the database contains 10,000 such mammograms, the computational cost for computing histograms for all coarsest components can be overwhelming. In this paper, we develop a local-feature-based indexing and retrieval method to overcome these drawbacks.

4. FEATURE-ORIENTED INDEXING AND RETRIEVAL ALGORITHMS

In this section, we present a local-feature-oriented image indexing and retrieval algorithm using Kwong and Tang's W -transform. The main difference between our algorithm and the algorithm in [4] is that our algorithm exploits local features rather than computing a multiresolution histogram for the whole image.

The feature-oriented image indexing system is built by the following steps:

1. Select image features that are either regions of interest (ROIs) or regions used to identify the image.

2. Take W -transforms for each selected feature from scale 1 to scale J , where J usually takes an integer value between 3 and 5.
3. Compute histograms for multiscale decomposed feature images, and save them as indices.

The algorithm for retrieving images where there exist the same or similar features as those in the query image is as follows:

1. Click features in the query image.
2. Take W -transforms for the selected features from scale 1 to J .
3. For each scale,
 - compute histograms for the multiscale decomposed feature images,
 - compute DOH between the histogram of the query multiscale feature images and the histogram in the index, and
 - select the image if the DOHs are less than given thresholds.
4. Display the retrieved images.

The threshold in Step 3 of the retrieving algorithm is used to measure the degree of similarity: the less the threshold is, the more similar the feature contained in the retrieved image and that in the query image are. (If the threshold is zero, only the images with the exactly same features as in the query image will be retrieved.) In our numerical experiments, we found that histograms of multiscale feature images will not differ very much if we click a feature that has some displacement from that in the indexing system. Therefore, we recommend a slightly relaxed threshold to compensate for the histogram error caused by the slight displacement if one tries to retrieve the images with the same features.

Several different loop strategies can be used in Step 3 of the retrieving algorithm, depending on how much accuracy is required. One may compare all DOHs or just the first one or two DOHs from coarse scale to fine scale to decide whether to terminate searching. In our numerical experiments we determined that, for most cases, the desired images can be retrieved by considering DOHs of the coarsest scale if we carefully select a threshold.

5. EXPERIMENTAL RESULTS

Using MATLAB, we conducted simulation experiments to test the local-feature-oriented indexing and retrieval method for an MRI image database.

First, we built a feature-oriented index for the image database, using the algorithm described in the previous section. The process of building the index is an automatic process: we pick the features in the image, and the program builds the index itself. This is an important advantage of a content-based approach over a scene description language based approach.

Figure 1 to 5 illustrate the retrieving process. Figure 1 is a typical MRI image with a size of 256×256 and an 8-bit gray scale. Figure 2 is a selected feature inside Figure 1 and consists 60×80 pixels. Here we see that the size of the selected image is not required to be a power of 2. Figure 3 is the multiresolution decomposition of Figure 2 using W -transforms for the first scale. Figure 4 is the multiresolution decomposition of Figure 2 using W -transforms up to scale 3. Figure 5 is the histogram of the left upper corner of Figure 4. The histogram is compared with the indexing histograms, and those images will be displayed if the difference between the query histogram and indexing histogram is less than some given threshold.

Selecting the threshold is crucial in the retrieving algorithm. If the threshold is too small, the program will not tolerate any displacement of the selected features; if the threshold is too large, many undesired images will be retrieved. We tested this observation by examples. Figure 6 shows the average accumulative number of retrieved images for different DOH thresholds based on four sets of 22 feature images. Here we see that if a threshold is set larger than 300, two false images will be retrieved on average. A test on a larger image database is to be done in the future.

6. CONCLUSIONS

The feature-oriented image indexing and retrieval method not only can retrieve images that are similar or identical to the query images but also can retrieve images that contain features specified in the query images, even if the retrieved images as a whole might be very different from the query images. The feature-oriented method also provides a speed advantage over the global multiresolution histogram comparison method. Our approach seems promising for managing large-scale image systems such as video databases and medical image databases. How to build the technique into different hierarchical storage systems remains a future research topic.

7. ACKNOWLEDGMENTS

This work was supported by the Office of Scientific Computing, U.S. Department of Energy, under Contract W-31-109-Eng-38.

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

1. MRI Query Image

3. L1 Decomposition

5. Histogram (L3)

2. MRI Feature Image

4. L3 Decomposition

6. Threshold Selection

References

- [1] Daubechies, I., *Ten Lectures on Wavelets*, CBMS-NSF Series Appl. Math., SIAM, 1991.
- [2] Kwong, M. K., and Tang, P. T., *W-matrices and Nonorthogonal Multiresolution Analysis of Finite Signals of Arbitrary Length*, Argonne National Laboratory Preprint Series MCS-P449-0794, 1994.
- [3] Kwong, M. K., *MATLAB Implementation of W-Matrix Multiresolution Analyses*, Argonne National Laboratory Preprint Series MCS-P449-0794, 1994.
- [4] Lee, J., and Dickinson, B. W., "Multiresolution video indexing for subband coded video databases", *SPIE Proc. of Storage and Retrieval for Image and Video Databases II*, pp. 162-169, 1994.
- [5] Pentland, A., Picard, R., Davenport, G. and Haase, K., "Video and image semantics: Advanced tools for telecommunications", *IEEE Multimedia*, pp. 73-75, 1994.