1-1-2001

Image segmentation in the wavelet domain using N-cut framework

Dongsheng Yao

University of Nevada, Las Vegas

Follow this and additional works at: https://digitalscholarship.unlv.edu/rtds

Repository Citation
https://digitalscholarship.unlv.edu/rtds/1267

This Thesis is brought to you for free and open access by Digital Scholarship@UNLV. It has been accepted for inclusion in UNLV Retrospective Theses & Dissertations by an authorized administrator of Digital Scholarship@UNLV. For more information, please contact digitalscholarship@unlv.edu.
INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

ProQuest Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

UMI®

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
IMAGE SEGMENTATION IN THE WAVELET DOMAIN
USING N-CUT FRAMEWORK

by

Dongsheng Yao

Bachelor of Science
Nankai University, Tianjin
1993

A thesis submitted in partial fulfillment
of the requirements for the

Master of Science Degree
Department of Electrical and Computer Engineering
Howard R. Hughes College of Engineering

Graduate College
University of Nevada, Las Vegas
May 2001
The Thesis prepared by

Dongsheng Yao

Entitled

Image Segmentation in the Wavelet Domain Using N-Cut Framework

is approved in partial fulfillment of the requirements for the degree of

Master of Science

Examination Committee Chair

Dean of the Graduate College

Examination Committee Member

Examination Committee Member

Graduate College Faculty Representative
ABSTRACT

Image Segmentation in the Wavelet Domain Using N-Cut Framework

by

Dongsheng Yao

Dr. Shahram Latifi, Examination Committee Chair
Professor of Electrical and Computer Engineering
University of Nevada, Las Vegas

We introduce a wavelet domain image segmentation algorithm based on Normalized Cut (NCut) framework in this thesis. By employing the NCut algorithm we solve the perceptual grouping problem of image segmentation which aims at the extraction of the global impression of an image. We capitalize on the reduced set of data to be processed and statistical features derived from the wavelet-transformed images to solve graph partitioning more efficiently than before. Five orientation histograms are computed to evaluate similarity/dissimilarity measure of local structure. We use properties of the wavelet transform filtering to capture edge information in vertical, horizontal and diagonal orientations. This approach allows for direct processing of compressed data and results in faster implementation of NCut framework than that in the spatial domain and also decent quality of segmentation of natural scene images.
# TABLE OF CONTENTS

ABSTRACT ............................................................................................................................... iii  

TABLE OF CONTENTS ....................................................................................................... iv  

LIST OF FIGURES ................................................................................................................ v  

ACKNOWLEDGMENTS ....................................................................................................... vi  

CHAPTER 1 INTRODUCTION ............................................................................................ 1  

CHAPTER 2 METHODOLOGIES OF THE WAVELET TRANSFORM ............... 9  
  Wavelet Overview ............................................................................................................. 9  
  Wavelet and Other Time-Frequency Transformation Methods ..................................... 10  
  Wavelet Analysis ............................................................................................................. 16  
  Multiresolution Analysis Using Wavelet Transform ..................................................... 19  
  Different Scaling Functions and Wavelets ................................................................. 23  
  Properties of the Wavelet Transform and Multiresolution analysis ....................... 29  
  Wavelet-Based Signal Processing and Applications .................................................. 34  

CHAPTER 3 NORMALIZED CUTS FRAMEWORK ............................................... 36  
  Orientation Energy Weighting ...................................................................................... 43  

CHAPTER 4 ALGORITHM DESCRIPTION AND RESULTS ...................... 47  
  Feature Extraction in Our Algorithm .......................................................................... 47  
  Weighting Matrix and Results ....................................................................................... 51  

CHAPTER 5 CONCLUSION ............................................................................................ 59  

BIBLIOGRAPHY ................................................................................................................... 61  

VITA ......................................................................................................................................... 67
LIST OF FIGURES

Figure 1  A hierarchy of digital image processing ........................................... 3
Figure 2  Time-frequency space for the STFT .................................................... 11
Figure 3  Short-Time Fourier Transform ....................................................... 12
Figure 4  Resolution comparison for the STFT and Wavelet Transform .......... 15
Figure 5  Nest Spaces for the Scaling Function and Wavelets ....................... 22
Figure 6  Haar Scaling Function and Haar Wavelet ....................................... 25
Figure 7  Wavelet-Based Signal Processing ................................................... 29
Figure 8  Diagram of the Wavelet Transform ............................................... 30
Figure 9  Daubechies2 Wavelet Transform of an image ............................... 31
Figure 10 Five Masks Used to Compute Orientation Histogram .................... 32
Figure 11 Three Texture Blocks ..................................................................... 32
Figure 12 A Case Where Minimum Cut Fails .............................................. 37
Figure 13 A Point Set .................................................................................... 38
Figure 14 Eigenvalue of the Point Set Case .................................................. 39
Figure 15 Eigenvector Associated With the Smallest Eigenvalue of the Point Set Case ................................................................. 40
Figure 16 Eigenvector Associated With the 2nd Smallest Eigenvalue of the Point Set Case ......................................................................................... 41
Figure 17 Eigenvector Associated With the 3rd Smallest Eigenvalue of the Point Set Case ......................................................................................... 42
Figure 18 Eigenvector Associated With the 4th Smallest Eigenvalue of the Point Set Case ......................................................................................... 43
Figure 19 Eigenvector Associated With the 5th Smallest Eigenvalue of the Point Set Case ......................................................................................... 44
Figure 20 Eigenvector Associated With the 6th Smallest Eigenvalue of the Point Set Case ......................................................................................... 45
Figure 21 Segmentation 1 for the Point Set ...................................................... 48
Figure 22 Segmentation 2 for the Point Set ...................................................... 49
Figure 23 Orientation Histograms of the 1st Level Wavelet Transform .......... 50
Figure 24 Orientation Histograms of the 2st Level Wavelet Transform ........... 51
Figure 25 Segmentation of the Brodatz Image Using Non-orientation Histogram 52
Figure 26 Segmentation Using Haar Wavelet ............................................... 53
Figure 27 Segmentation Using Bior2.2 Wavelet ............................................. 54
Figure 28 Segmentation of the Sofa Image .................................................... 55
Figure 29 Segmentation of a Brodatz Image ............................................... 55
Figure 30 Segmentation of People Image ....................................................... 56
Figure 31 Segmentation of a Scene Image ..................................................... 57
Figure 32 Segmentation of Another Scene Image .......................................... 58
ACKNOWLEDGEMENTS

I want to thank Dr. Latifi for his guidance on my research and study at the University of Nevada, Las Vegas. Without his precious advice and encouragement, I could not have reached this stage in my academic pursuits.

I would also like to thank my Examination Committee: Dr. Henry Selvaraj, Dr. Sahjendra Singh, and Dr. Zhiyong Wang; for their valuable suggestions and comments on this thesis. I would like to give special thanks to Dr. Emma Regentova, who spent much time helping me on the research, and gave me many suggestions on the implementation of our algorithm and the research direction of the thesis. Thanks must also go to Shulan Deng and Shaojie Wu for their discussions on this research.

I wish to express my great appreciation for my wife Yuqiu’s confidence in me and for all her assistance during my study in Las Vegas. Finally, I would like to share this thesis with my parents, Buting Yao and Taolan Miao, for their constant support and understanding during my study here.

This research was supported by NSF Research Grant IRI-9616206 and NASA Research grant NAG5-3994.
CHAPTER 1

INTRODUCTION

One of the major roles played by visual observation is to guide science applications. For instance, we hope that a computer can distinguish a bird from the sky and clouds, in the same way we do by visual observation. This demand has caused image processing to be extensively researched by many scientists using many different approaches. Our goal is to get a better view and understanding of object of interest in the image. In other words, because we need to distinguish the objects of interest from the background and other objects, the operation of image segmentation is needed. In most of the natural images, textures are the main differences between different parts of interest of the human being. This property causes the area of texture segmentation to undergo tremendous growth in recent years. In [1], the segmentation is pointed out to be the operation at the threshold between low-level image processing and image analysis. There are many steps for the whole procedure of image processing; segmentation is only one part of the problem. Figure 1 shows the diagram of the hierarchy of the image processing operations. From the figure we can see image processing begins with the capture of an image with a suitable acquisition system. Once the image is sampled, it must be converted into a form or structure that can be treated with computers. This process is called digitization. The first steps of digital processing
may include a number of different operations and are known as image preprocessing. At these steps, we have to correct the non-linear characteristics of the sensors such as CCD, perform coordinate transformations, and correct radiometric and geometric errors. After this, to analyze and identify objects, we need to apply adequate filtering procedures. It is very important that feature images are extracted from the original images. The basic operations for the feature image generation are averaging, edge detection and the analysis of simple neighborhoods and complex texture patterns. After this the object has to be separated from the background; the segmentation resides right here.

Image segmentation is defined as the process of partitioning a digital image into disjoint (non-overlapping) regions [2]. From three different philosophical perspectives, image segmentation can be approached either by assigning pixels to a particular region (region merging), by attempting to locate boundaries between regions, or by identifying edge pixels, and then, linking them together to form boundaries. That is, these algorithms in some respect simulate the visual perception mechanism, which in fact combines merging and splitting based on the low cues such as color, brightness, texture etc. Also, some prior knowledge is used by human beings for image partitioning. To make computer interpretation more probable, either low- or high-level features should be utilized, such as image models. Obviously, it is not a simple problem to obtain such a knowledge.

Generally, there are several basic concepts for segmentation. Pixel-based methods only use the gray values of the individual pixels. Region-based methods analyze the
Figure 1: A Hierarchy of Digital Image Processing
gray values in larger areas. Finally, edge-based methods detect edges and then try to follow them. Pixel-based methods are considered the simplest approach because they only consider the difference between pixel's intensities; thus they are only suitable for the image containing no textures. Edge-based techniques are based on the fact that the position of an edge is given by an extreme of the first-order derivative or a zero-crossing in the second-order derivative. Thus these approaches are able to deal with the objects with gradually changed intensities and graded backgrounds, while the pixel-based approaches will have problems. Region-based methods pay more attention to the statistical characteristics of the neighborhood of objects. By examining the features within a small window around a pixel, this method computes similarities between them by their texture relationships. Thus these techniques work very well on the textured images. While most images we have to deal with contain textures, we need to give texture segmentation more attention. Although a wide variety of methodologies have been applied to this problem, most of the research focuses on feature-based approaches and on the appropriate texture features.

The crucial importance of the features for the texture representation [3] causes many researchers to try many different ways to find the proper features. The feature representing methods are categorized into two parts [4]: statistical approaches and spatial/spatial-frequency approaches.

In the statistical approaches, researchers use co-occurrence matrices [5], second-order statistics [6], local linear transformations [7], local extrema [8], and Markov random fields [9][10][11][12]. In the spatial/spatial-frequency approaches, researchers
use features obtained by computing Fourier transform domain energy [13], local orientation and frequency [14], multi-channel Gabor filters [15][16][17][18][19], and wavelet.

The spatial-frequency domain approaches, especially the wavelet approaches, are studied extensively for their efficiency in analyzing the information content of images. Recently, some researchers explored approaches which can take advantage of more than one method used before. For instance, we can see from [20], the authors tried to explore the edge information, i.e. the contour continuity in the region-based approaches. By applying an adaptive contour factor, which is related to the orientation energy, the region-based methods can be weighted according to the region or non-region information of different parts of the images. The authors above process images in the spatial domain, thus they have to construct several orientation filters and apply them to the images, and then use a complex algorithm to extract the contour, i.e. edge information. We can see from here the huge potential of using wavelet approaches.

Among the many segmentation algorithms, Normalized Cut framework is introduced in [21]. This algorithm is in the category of clustering methods derived from the graph theory, such as the minimum spanning tree and the directed tree. The algorithms in this family are capable of detecting clusters of various shapes, at least for the case in which they are well separated. In [21] the authors modeled the approach as a form of solving the perceptual grouping problem. Their approach is in fact based on modeling of perceptual grouping. Image segmentation here is treated as a graph partitioning problem, and a normalized cuts criterion is suggested to measure both
the total similarity within groups and total dissimilarity between regions. If region merging techniques are well suited for textured images, boundary and edge detection based techniques work well for smooth images containing extended objects. Real world images, that is images of natural scenes, usually comprise regions of different local structure. Thus, the algorithm which works efficiently for images of different structural content is preferable. In this respect, the NCuts algorithm based on the region-based split and merge approach is suitable. In the grouping of the NCuts algorithm, we need to partition the set of vertices into mutually exclusive subgraphs by the similarity among the vertices in a subgraph. By computing the weighted matrices among each pair of the vertices and solving the eigensystem, the authors get the optimal bipartition of the graph. Here the problem arises: the computation of the eigensystem solver is greedy and is of $O(n)$, where $n$ is the number of the nodes, i.e. the number of the pixels for images. This is the main drawback of graph partitioning techniques which constitute the base for the NCuts framework. With the wavelet decomposition, we can reduce the data set dimension extremely, thus the computation will be much faster. Also we can employ different weight methods for the weighting of the links between each pair of the nodes to get different similarity of representations.

Wavelet transform and multiresolution decomposition have been studied intensively for their applications in various image analysis tasks. There is considerable experience accumulated on edge detection, texture discrimination, shape feature computation and image similarity estimation implemented by means of transform domain analysis[25][28][29][30][31][35]. We have studied low level structural features which
can be computed by data multiresolution analysis[42]. Edge points locations and edge density are computed as feature vectors for content-based image querying. Using energy analysis across the scales, texture and smooth images are distinguished. Different levels are employed for edge detection and texture feature computation. The energies of the wavelet coefficients in three detailed subbands are computed, and their ratios to the total energy of the scale are estimated. By comparing this ratio to a given threshold, the decision is made on whether the structure of an image or image region possesses dominant directionality or not; and then images are indexed according to their structural properties. The effective scheme of image querying coupled with the reduced set of data to be processed yield fast database retrieval.

In [41], we have shown how to employ wavelet transform for document segmentation. Document images are treated as textured, where texts, drawings and pictures within the document are treated as different classes of textures. Images are processed in the wavelet domain, and such features as variation and energy of coefficients within blocks are derived. Based on these features, k-means classifier solves grouping task.

In this thesis, we will explore texture feature representation, extraction and the application of the Normalized Cut framework in the wavelet domain. Our goal is to find the effect of the wavelet-based feature associated with the Normalized Cut framework, and to improve the efficiency of the computation while keeping a comparable segmentation quality. We will show how to carry computation over the large dataset to the smaller one, and to obtain features required to compute pairwise similarity/dissimilarity measures, i.e. weights required for graph partitioning algorithm.
Five orientation histograms are computed. Three of them are derived directly from the high-pass subbands and two others are computed by low-pass subbands processing. In Chapter 2, we will give a basic explanation about the wavelet decomposition. In Chapter 3, we will show the detail about the Normalized Cut framework. In Chapter 4, we will show our algorithm and the results. The conclusions will be given in Chapter 5.
CHAPTER 2

METHODOLOGIES OF THE WAVELET TRANSFORM

In this chapter, we will give a basic explanation about the wavelet decomposition. Since our experiments are based on the wavelet transformation, it is very important to understand clearly how the wavelet decomposition works and what the important properties of the wavelet transformation are.

2.1 Wavelet Overview

Wavelets are mathematical functions that cut data into different frequency components, and then study each component with a resolution matched to its scale[22]. As pointed by Yves Meyer, the objectives of signal processing are to analyze accurately, code efficiently, transmit rapidly, and then to reconstruct carefully at the receiver the delicate oscillations or fluctuations of this function of time[37]. For the 2-D case, the signal becomes the image, and will be represented by different numerical values in matrices. The goal of most modern wavelet research is to create a set of basis functions or general expansion functions and transforms that will give an informative, efficient, and useful description of a function of signal. The multiresolution decomposition separates components of a signal in a way that is superior to most other methods for analysis and processing. We will show it in Section 2.4. The discrete
wavelet transform has such properties to decompose a signal at different independent scales and to do it in a very flexible way. This gives the wavelet transform a huge potential as a tool for signal and image processing.

2.2 Wavelet and Other Time-Frequency Transformation Methods

Fourier analysis has been used for many years for its precise physical interpretations, such as time-frequency analysis of signals. Both Fourier transforms and Fourier series are Fourier analysis. For a continuous time-domain signal, i.e. an analog signal $x$, the Fourier transform $X(f)$ of the signal $x$ describes the spectral behavior of $x$ in terms of frequency in the frequency domain. While the Fourier series transform the signal into a summation of bi-infinite sequences with $2\pi$-period, the Fourier transform is an integral of the signal and $e^{-2\pi jft}$:

$$X(f) = \int_{-\infty}^{+\infty} x(t)e^{-2\pi jft} dt$$  \hspace{1cm} (1)

As shown in Equation 1, the coefficients $X(f)$ is computed as inner products of the signal with sine-wave basis functions of infinite duration. As a result, Fourier analysis works well if $x(t)$ is composed of a few stationary components. We can imagine that in the Fourier analysis, a signal can be viewed as the simple addition of waves with different periods. The wave here is defined as an oscillating function of time or space, with different phases, amplitudes, such as a sine or a cosine wave. It is well known that for a periodic, time-invariant signal, it's components in the frequency domain keep at fixed places. Due to the characteristics of the Fourier analysis, for a transient, non-
stationary, or time-varying signal, its frequency components may expand the whole frequency to infinity. Also, we are not able to locate the precise time when the signal changes. Therefore, an analysis which can adapt itself according to the frequency components of the signal is required for the non-stationary signal analysis.

There are two ways to satisfy the above need: one is to choose a window function, by which the input signal $x(t)$ is chopped up into sections, and each section is analyzed for its frequency content separately. The windowed-section is processed as a small stationary signal. Another method is the wavelet transform. The former method is called the Short-Time Fourier Transform (STFT) or Gabor transform. In the STFT, once the window is chosen, its size is fixed. In other words, its resolution for different
frequency components is the same. The STFT is the classical method for the time-frequency representation of signals. It is actually a two-dimensional time-frequency representation of the signal $x(t)$. The representation $S(\tau, f)$ of the STFT consists of spectral characteristics associated with time and local frequency. Gabor defined the first $S(\tau, f)$ as:

$$S(\tau, f) = \int x(t)g^*(t - \tau)e^{-2\pi if t} \, dt$$  \hspace{1cm} (2)$$

This approach assumes that the signal $x(t)$ is stationary when it is seen through a window $g(t)$ of limited extent, centered at a time location $\tau$. The signal is windowed by $g(t)$ to $x(t)g^*(t - \tau)$.

![Figure 3: Short-Time Fourier Transform](image)

Figure 3 shows the Short-Time Fourier Transform.

Figure 2 shows the time-frequency space of the Short-Time Fourier Transform.

Figure 3 shows the Short-Time Fourier Transform. From equation 2 we can see that
the spectrum of the window function produces an estimation of the fundamental frequencies of \( x(t) \) around time \( \tau \). There are two problems of the STFT: first, due to the properties of the window function, there are unavoidable artificial frequency components introduced in the interested signals. The only window function without this affect is the delta function \( \delta(t) \), which is defined as:

\[
\delta(t) = \begin{cases} 
+\infty & \text{for } t=0 \\
0 & \text{otherwise}
\end{cases}
\]  

such that

\[
\int_{-\infty}^{+\infty} \delta(t) \, dt = 1
\]  

But we can deduce that with such a short window, the gated signal \( f(t)\delta(t) \) cannot contain useful frequency components, because the time range is too short. According to the uncertainty principle in the information theory, the amount of information we can extract is limited by relating the window size to the frequency resolution. This can be expressed as:

\[
Time \times Bandwidth = \Delta t \Delta f \geq \frac{1}{4\pi}
\]  

This equation is also called the Heisenberg inequality. Gabor uses the Gaussian function as the window function \( g(t) \), which is believed to be the best optimally concentrated function in both time and frequency and meeting the bound with equality. The second problem of the STFT, also the main problem of the STFT, is its fixed
length of the window function. With the fixed window length, the time-frequency resolution for the STFT is also fixed over the entire time-frequency space. Thus, if the input signal composes of small bursts associated with long quasi-stationary components, then by carefully choosing the window function, we can analyze time or frequency components with good resolution. But we cannot do the analysis to both time and frequency components with a good resolution at the same time.

To solve the problems associated with the STFT, the wavelet transform is introduced. Wavelet transforms provide length-variable windows. They are able to obtain both detailed frequency components and extract sharp bursts of the input signal. By using varied short high-frequency basis functions and long low-frequency basis functions, wavelet transforms achieve this requirement and overcome the STFT for the non-stationary signal analysis and processing. Figure 4 shows different time-frequency resolutions of the STFT and the wavelet transform. For both low- and high-frequency components, the STFT has the same resolution, while the wavelet transform has varied resolution by which it can capture both time and frequency position precisely. For the low frequency components, the wavelet transforms have a higher frequency resolution but a lower time resolution; for the high frequency components, it has a lower frequency resolution but a higher time resolution. Since most of the signals we encounter in practice need high resolution for low frequency components, the wavelet transforms give us better solutions.
Figure 4: Resolution Comparison for the STFT and Wavelet Transform
2.3 Wavelet Analysis

We have seen the forte of the wavelet transform over the STFT approaches in the analysis of the real signals. In this section, we will show how to construct wavelet transforms.

Like the Fourier transform, there are two kinds of wavelet transforms: continuous and discrete wavelet transforms (DWT). Actually, the discrete wavelet transform is the discretizing version of the continuous wavelet transform. For the continuous signals, the continuous wavelet transform (CWT) uses two continuous variables to shift (or translate) and scale the basic wavelet, it is defined as:

\[ F(a, b) = a^{-1/2} \int f(t) \psi \left( \frac{t-a}{b} \right) dt \]

and its inverse transform is:

\[ f(t) = K \int \int \frac{1}{a^2} F(a, b) \psi \left( \frac{t-a}{b} \right) dadb \]

where \( \psi(t) \) is the basic wavelet and \( a, b \in \mathbb{R} \) are real continuous variables, and \( K \) is given by:

\[ K = \int \frac{|W(\omega)|^2}{|\omega|} d\omega \]

with \( W(\omega) \) being the Fourier transform of the wavelet \( \psi(t) \). \( K \) is used to normalize the wavelet transform. Equation 7 to be satisfied requires \( K < \infty \). In most cases, this requires that \( W(0) = 0 \) and \( W(\omega) \) go to zero fast enough[23]. Note that the Fourier transform \( W(\omega) \) of the wavelet \( \psi(t) \) is dilated as the scale increases. The large scale
corresponds to contracted signals while the small scale corresponds to dilated signals. The shifting and scaling factors $a$ and $b$ in the CWT are continuous. That means the signal $f(t)$ is transformed by taking the inner product of the signal and the basic wavelet shifting and scaling at any possible continuous real values.

In the discretization of the wavelet transform, we set $a$ to be positive and $a, b$ to be discrete values only. Let's consider the dilation factor of the wavelet transform: $a$ is at values of $a_0^m$, where $m \in Z$, and $a_0 \neq 1$ is constant and positive. With each different value of $m$, we get a different dilation value. We can define the continuous wavelet transform family as:

$$\psi_{a,b}(t) = |a|^{-1/2} \psi\left(\frac{t-b}{a}\right)$$

thus, for each value of $m$, the width of $a_0^{-m/2} \psi(a_0^{-m} t)$ is $a_0^m$ time the width of $\psi(t)$.

For $m = 0$, we can discretize $b$ by taking the integer multiples of a constant $b_0$, where $b_0$ is chosen so that the $\psi(t - nb_0)$ covers the whole line [24]. Consider both cases, we choose $a = a_0^m, b = nb_0a_0^m$, where $m, n$ both range over $Z$. Thus the corresponding wavelet changes to:

$$\psi_{m,n}(t) = a_0^{-m/2} \psi\left(\frac{t-nb_0a_0^m}{a_0^m}\right) = a_0^{-m/2} \psi(a_0^{-m} t - nb_0)$$

For computational efficiency, we use $a_0 = 2$ and $b_0 = 1$. These data will partition the frequencies into consecutive "octaves", in other words, adjacent frequency bands with doubling bandwidth. This function $2^{-m/2} \psi(2^{-m} t - n)$ is obtained from the single
wavelet function $\psi(t)$ by a binary dilation (i.e. dilation by $2^m$ and a dyadic translation of $n/2^m$). With this wavelet basis, we can write the series representation of $f(t)$ as

$$f(t) = \sum_{j,k=-\infty}^{\infty} c_{j,k} \psi_{j,k}(t) \quad (11)$$

where the wavelet coefficients $c_{j,k}$ are given by

$$c_{j,k} = \langle f, \psi_{j,k} \rangle \quad (12)$$

We have defined the integral transform or inner production $W_\psi$ as

$$(W_\psi f)(b, a) := |a|^{-1/2} \int_{-\infty}^{\infty} f(t) \psi(\frac{t - b}{a}) dt \quad (13)$$

and $a = 2, b = 1$, then the wavelet coefficients in Equation 12 become

$$c_{j,k} = (W_\psi f)(\frac{k}{2^j}, \frac{1}{2^j}) \quad (14)$$

Thus, the $(j, k)^{th}$ wavelet coefficient of $f$ is given by the integral wavelet transformation of $f$ evaluated at the dyadic position $b = k/2^j$ with binary dilation $a = 2^{-j}$, and the same wavelet $\psi$ is used to generate the wavelet series in Equation 11. For $j > 0$, i.e. $2^j > 1$, the $\psi_{j,k}(t) = \psi_k(2^j t)$ is narrower and is translated in smaller steps. Thus, it can represent finer detail. For $j < 0$, i.e. $2^j < 1$, $\psi_{j,k}(t)$ is wider and is translated in larger steps, so these wider scaling functions can represent only coarse information, and the space they span is smaller.
2.4 Multiresolution Analysis Using Wavelet Transform

Multiresolution representations are very effective for analyzing the information content of images. We can use the wavelet transform to approximate a signal at a given resolution. Under such conditions, the difference of information between the approximation of a signal at the resolutions $2^{j+1}$ and $2^j$ can be extracted by decomposing this signal on a wavelet orthonormal basis of $L^2(R^n)$. We already know that in $L^2(R)$, a family of wavelet basis is formulated by $2^{j/2}\psi(2^j t - k)$. This decomposition defines an orthogonal multiresolution representation called a wavelet representation[25]. Given a sequence of increasing resolutions $\{r_j, j \in Z\}$, the details of an image at the resolution $r_j$ are defined as the difference of information between its approximation at the resolution $r_j$ and its approximation at the lower resolution $r_{j-1}$. At different resolutions, the details of an image generally characterize different physical structures of an image. At a coarse resolution, these details correspond to the larger structures; at a finer resolution, these details correspond to the smaller blocks of the image. The coarse-to-fine strategy has already been widely studied. In the discrete wavelet cases, the difference of information between two approximations at the resolutions $2^{j+1}$ and $2^j$ is extracted by decomposing the function in a wavelet orthonormal basis. The basic requirement of multiresolution analysis is nesting spanned spaces such that:

$$\ldots \subset V_{-2} \subset V_{-1} \subset V_0 \subset V_1 \subset V_2 \ldots \subset L^2$$ (15)
or

\[ V_j \subset V_{j+1} \text{ for all } j \in \mathbb{Z} \]  \hspace{1cm} (16)

with

\[ V_\infty = 0, V_0 = L^2. \] \hspace{1cm} (17)

The space containing higher resolution signals will contain the space of lower resolution [23]. The condition

\[ f(t) \in V_j \iff f(2t) \in V_{j+1} \] \hspace{1cm} (18)

does not necessarily mean that the elements in a space are simply scaled versions of the elements in the next space. The approximation of a signal at a resolution \( 2^{j+1} \) contains all the necessary information to compute the same signal at a smaller resolution \( 2^j \). When computing an approximation of \( f(t) \) at resolution \( 2^j \), some information about \( f(t) \) is lost. However, as the resolution increases to \( +\infty \) the approximated signal should converge to the original signal. The nesting of the spans of \( \phi(2^j t - k) \), i.e. \( V_j \) is achieved by requiring that \( \phi(t) \in V_1 \); that is, if \( \phi(t) \) is in \( V_0 \), it is also in the space spanned by \( \phi(2t) \), \( V_1 \). This means \( \phi(t) \) can be expressed in terms of a weighted sum of shifted \( \phi(2t) \) as

\[ \phi(t) = \sum_{n} h(n) \sqrt{2} \phi(2t - n), n \in \mathbb{Z} \] \hspace{1cm} (19)

where the coefficients \( h(n) \) are a sequence of real or complex numbers called the scaling function coefficients (or the scaling filter or the scaling vector), and the \( \sqrt{2} \) maintains
the norm of the scaling function with the scale of two. This equation is called the multiresolution analysis equation, or the dilation equation. Now let us consider the orthogonality of the scaling functions and wavelets. Orthogonal basis functions allow simple calculation of expansion coefficients; and they have the Parseval's theorem to keep the energy in the wavelet transform domain as the same in the spatial domain.

We denote the orthogonal complement of $V_j$ in $V_{j+1}$ as $W_j$ and shown in Figure 5. In the Figure 5, $V_1 = V_0 \oplus W_0$, and $V_2 = V_0 \oplus W_0 \oplus W_1$. The relationship of the $V_j$ is $V_3 \supset V_2 \supset V_1 \supset V_0$. The whole space is defined as

$$L^2 = V_0 \oplus W_0 \oplus W_1 \oplus \ldots$$

(20)

where $V_0$ is the initial space spanned by the scaling function $\phi(t - k)$. Thus all members of $V_j$ are orthogonal to all members of $W_j$. In other words

$$\langle \phi_{j,k}(t), \psi_{j,l}(t) \rangle = \int \phi_{j,k}(t) \psi_{j,l}(t) dt = 0$$

(21)

for all appropriate $j, k, l \in \mathbb{Z}$.

We know that $W_0 \subset V_1$, that is the wavelets reside in the space spanned by the next narrower scaling function. Thus the wavelets can be expressed by a weighted sum of shifted scaling function $\phi(2t)$ defined in Equation 19 by

$$\psi(t) = \sum_n h_1(n) \sqrt{2} \phi(2t - n), n \in \mathbb{Z}$$

(22)
Figure 5: Nest Spaces for the Scaling Function and Wavelets

for some set of coefficients $h_1(n)$. The wavelets span the difference or orthogonal complement spaces of the scaling functions. It is shown in [23] that

$$h_1(n) = (-1)^n h(1 - n)$$

(23)

in order to satisfy the orthogonality of integer translations of the wavelet or scaling function. From the Equation 22 we can derive that

$$\psi_{j,k}(t) = 2^{j/2} \psi(2^j t - k)$$

(24)

where $2^j$ is the scaling of $t$. Using these set of scaling functions $\phi_k(t)$ and wavelets $\psi_{j,k}(t)$, any given function $f(t) \in L^2(R)$ can be represented as

$$f(t) = \sum_{k=-\infty}^{\infty} c(k) \phi_k(t) + \sum_{j=0}^{\infty} \sum_{k=-\infty}^{\infty} d(j,k) \psi_{j,k}(t)$$

(25)
this is actually a series expansion of the scaling functions and wavelets. In the Equation 25, the first term gives the low resolution or coarse approximation of $f(t)$. The second term gives the increasing detail by adding up the higher or finer resolution functions. The coefficients $c(k)$ and $d(j, k)$ are calculated by

$$c(k) = c_0(k) = \langle f(t), \phi_k(t) \rangle = \int f(t) \phi_k(t) dt \tag{26}$$

and

$$d(j, k) = d_j(k) = \langle f(t), \psi_{j,k}(t) \rangle = \int f(t) \psi_{j,k}(t) dt. \tag{27}$$

Using Equation 19 and 22, Equation 25 can be represented as

$$f(t) = \sum_k c_{j_0}(k) 2^{j_0/2} \phi(2^{j_0} t - k) + \sum_k \sum_{j=2}^\infty d_j(k) 2^{j/2} \psi(2^j t - k) \tag{28}$$

where the $j_0$ is a factor setting the coarsest scale space spanned by $\phi_{j_0,k}(t)$. This is the discrete wavelet transform. There are many different bases we can choose for the scaling functions and the wavelets.

### 2.5 Different Scaling Functions and Wavelets

From the Equation 19, the scaling function $\phi(t)$ satisfies the multiresolution formulation

$$\phi(t) = \sum_n h(n) \sqrt{2} \phi(2t - n). \tag{29}$$

If we denote the Fourier transformation of $\phi(t)$ and $h(n)$ as $\Phi(\omega)$ and $H(\omega)$ respec-
tively, then the Equation 29 is equivalent to

$$\Phi = \prod_{k=1}^{\infty} \left\{ \frac{1}{\sqrt{2}} H \left( \frac{\omega}{2^k} \right) \right\} \Phi(0) \tag{30}$$

where

$$\sum_n h(n) = \sqrt{2} \tag{31}$$

and $\Phi(0)$ is well defined. If the $\phi(t)$ is orthogonal as defined by

$$\int \phi(t)\phi(t-k)dt = E\delta(k) = \begin{cases} E & \text{if } k=0 \\ 0 & \text{otherwise} \end{cases} \tag{32}$$

and

$$\sum_n h(n)h(n-2k) = \delta(k) = \begin{cases} 1 & \text{if } k=0 \\ 0 & \text{otherwise} \end{cases} \tag{33}$$

then we can construct the orthogonal wavelet[23]. Now we consider several wavelets.

The Haar wavelet can be constructed by

$$h(n) = \left\{ \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}} \right\} \tag{34}$$

and the scaling function is

$$\phi(t) = \begin{cases} 1 & \text{for } 0 < t < 1 \\ 0 & \text{otherwise} \end{cases} \tag{35}$$
The wavelet is
\[
\psi(x) = \begin{cases} 
1 & \text{for } 0 < t < 1/2 \\
-1 & \text{for } 1/2 < t < 1 \\
0 & \text{otherwise}
\end{cases}
\] (36)

Figure 6 displays the Haar scaling function and the Haar wavelet.

\[\phi(t) \quad 1 \begin{array}{c}
0 \\
-1
\end{array} \quad 1
\]

\[\psi(t) \quad 1 \begin{array}{c}
0 \\
-1
\end{array} \quad 0.5 \quad 1\]

Figure 6: Haar Scaling Function and Haar Wavelet

This orthonormal base \(\psi_0(t), \psi_1(t), \ldots, \psi_n(t), \ldots\) defined on \([0,1]\) makes the series

\[< f, \psi_0 > \psi_0(t) + < f, \psi_1 > \psi_1(t) + \ldots + < f, \psi_n > \psi_n(t) + \ldots\]

converges to \(f(t)\) uniformly on \([0,1]\) for any continuous function \(f(t)\) on \([0,1]\).

If function \(f(t)\) is continuous and has a continuous derivative, then its approximation by step functions is not appropriate. The Haar function is not continuous and its Fourier transformation decays only like \(|\omega|^{-1}\). This corresponds to the bad

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
frequency localization. Thus, the Haar function is even worse than some windowed Fourier bases. For the analysis of smoother functions, the discontinuous Haar basis is not suitable.

In the DWT, the compact support of the $\phi(t)$ and $h(n)$ aids in the time localization of the DWT and reduces the computation. Thus choosing the compact supported $\phi(t)$ and $h(n)$ is very important. With the simple FIR filters, we can implement the compact supported $h(n)$. The length of the sequence $h(n)$ together with the linear constraints and bilinear constraints described in Equation 31 and 33 provide a $\frac{N}{2} - 1$ degrees of freedom in choosing the $h(n)$. For a $h(n)$ of length $N = 2$, there are no degrees of freedom for us to choose the $h(n)$. This leads to the Haar scaling function coefficients $h_{Haar} = \{\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\}$, which are also the length-2 Daubechies coefficients. For the length-4 coefficient sequence, there is one degree of freedom for us to choose the $h(n)$, we can set the parameter as $\alpha$, and $h(n)$ satisfies

$$h(0) + h(1) + h(2) + h(3) = \sqrt{2} \quad (37)$$

$$h^2(0) + h^2(1) + h^2(2) + h^2(3) = 1 \quad (38)$$

and

$$h(0)h(2) + h(1)h(3) = 0 \quad (39)$$

We can get

$$h(0) = (1 - \cos(\alpha) + \sin(\alpha))/(2\sqrt{2}) \quad (40)$$
\begin{align*}
h(1) &= \frac{(1 + \cos(\alpha) + \sin(\alpha))}{(2\sqrt{2})} \\
h(2) &= \frac{(1 + \cos(\alpha) - \sin(\alpha))}{(2\sqrt{2})} \\
h(3) &= \frac{(1 - \cos(\alpha) - \sin(\alpha))}{(2\sqrt{2})}
\end{align*}

where if we set \( \alpha = \pi/3 \), we can get the Daubechies-4 coefficients

\[ h_{\text{Daub}} = \left\{ \frac{1 + \sqrt{3}}{4\sqrt{2}}, \frac{3 + \sqrt{3}}{4\sqrt{2}}, \frac{3 - \sqrt{3}}{4\sqrt{2}}, \frac{1 - \sqrt{3}}{4\sqrt{2}} \right\} \] (44)

For the length-6 coefficient vector or longer vectors, by simply setting several parameters associated with the degrees of freedom and the values of these parameters, we can get several different wavelet coefficient vectors.

Daubechies showed methods for constructing orthonormal wavelets with the compact support and the maximum number of vanishing moments[23]. The first step is choosing the length \( N \) for the \( h(n) \), and then setting factors of \( |H(\omega)|^2 \), which is the Fourier transformation of \( h(n) \). The calculations are carried out using the z-transform of the transfer function and using convolution in the time domain rather than multiplication in the frequency domain. Daubechies showed that it is not possible for a completely symmetric real \( h(n) \) to have compact support and support orthogonal wavelets at the same time; but it is possible for complex \( h(n) \), biorthogonal systems, infinitely long \( h(n) \), and multiwavelets. By choosing different set of the \( N - 1 \) roots of the \( H(z)H(1/z) \), we can get some variations in Daubechies’ wavelets. This zero moment design approach assures that the resulting wavelet system is an orthonor-
mal basis. The systems having a maximum number of zero moments of the wavelets also have a high degree of smoothness for the scaling and wavelet functions. The moments of the wavelets provide information about the flatness of $H(\omega)$ and smoothness of $\psi(t)$. The moments of the scaling function $\phi(t)$ and scaling coefficients $h(n)$ are measures of the localization and symmetry characteristics of the wavelet transform.

The orthogonality across both translation and scale gives a clean, robust, and symmetric formulation satisfying the Parseval's theorem. It also places strong limitations on the possibilities of the wavelet system, because the orthogonality needs to use a large number of the degrees of freedom. This results in complicated design of the equations. The complicated design prevents linear phase of the analysis and synthesis filter banks; and prevents asymmetric analysis and synthesis systems. The biorthogonal wavelet system allows greater flexibility at the expense of the energy partitioning property of the Parseval's theorem. In the orthogonal wavelets, the analysis filters and synthesis filters are time reversal of each other, i.e., $\hat{h}(n) = h(-n), \hat{g}(n) = g(-n)$.

For the biorthogonal case, the four filters have to be related as

\[ g(n) = (-1)^n h(1 - n), g(n) = (-1)^n h(1 - n) \]  \hspace{1cm} (45)

and

\[ \sum_n \hat{h}(n) h(n + 2k) = \delta(k) \]  \hspace{1cm} (46)

The $\hat{h}(n)$ is orthogonal to $h(n)$. We know for the orthogonal case, we have $\sum_n h(n)h(n + 2k) = \delta(k)$. This is the reason that it is called biorthogonal. The biorthogonal wavelet
systems generalize the classical orthogonal wavelet systems. They are more flexible and generally easier to design. There are several differences between the orthogonal and the biorthogonal wavelet systems: The orthogonal wavelet filter and scaling filter must be of the same length, and the length must be even, say 2, 4, 6, etc. This restriction is not needed for biorthogonal systems. In the orthogonal wavelet systems, the symmetric compact support wavelets and scaling functions are impossible, but for the biorthogonal wavelet systems, this can be easily achieved. The main disadvantages of the biorthogonal systems is that the Parseval's theorem is no longer satisfied. That is the norm of the wavelet coefficients is not the same as the norm of the signals spanned. At this point, the orthogonal system beats the biorthogonal.

![Figure 7: Wavelet-Based Signal Processing](image.png)

2.6 Properties of the Wavelet Transform and Multiresolution Analysis

We have briefly reviewed wavelets and multiresolution analysis. For more detailed treatments we refer to [24][32][37][40].

Wavelet transform decomposes a signal into the coarser resolution representation containing low frequency approximation information and the higher frequency details.
Wavelet transform employs a group of filtering functions (filter bank) to decompose signals into different frequency components. For each component, there is a scaling function which matches the appropriate level of resolution. Figure 8 shows the dyadic transform, where $h_0$ and $h_1$ are high and low-pass filters respectively. And the down-sampling reduces the resolution by a factor of 2. Thus, at any given scale we have initial two-dimensional data reduced by a factor of 4. By inverting the procedure, the perfect reconstruction of an initial image can be obtained by means of conjugate mirror filters.

We know that the Haar wavelet is not suitable for the continuous signal processing; people turn to the new basis of the wavelet transform. Battle-Lemarie obtained another orthonormal wavelet by orthonormalizing $m$th order cardinal B-splines $N_m(t)$ for $m \geq 2$. However, this leads to an infinitely long filter. We have known that the Daubechies' orthonormal basis (see [37]) has compact support. That is, for each
integer $r \geq 0$, a function $\psi(t)$ of class $C^r$, the Daubechies' orthonormal basis for $L^2(R)$ is defined as

$$\psi_{r,j,k}(t) = 2^{j/2} \psi_r(2^j t - k), j, k \in Z$$  \hspace{1cm} (47)

where the function $\psi_r(t)$ in $L^2(R)$ has the property that $\{\psi_r(t - k)|k \in Z\}$ is an orthonormal sequence in $L^2(R)$.

Then the trend $f_j$ at scale $2^{-j}$, of a function $f \in L^2(R)$ is defined by

$$f_j(t) = \sum_k < f, \psi_{r,j,k} > \psi_{r,j,k}(t).$$  \hspace{1cm} (48)

The fluctuations, or details in the case of an image, are denoted by $d_j(t)$ and defined by

Figure 9: Daubechies2 Wavelet Transform of an Image
To analyze these details further, we define an orthonormal scaling function $\phi_r(t)$ such that $\{\phi_r(t - k), k \in Z\}$ is also an orthonormal sequence. The scaling function has properties similar to those of the wavelet $\psi_r(t)$:

$$2^{j/2}\phi_r(2^j t - k) = \phi_{r,j,k}(t), j, k \in Z \quad (50)$$
The two functions $\phi_r(t)$ and $\psi_r(t)$, are also called the father wavelet and the mother wavelet respectively. Figure 9 shows wavelet transform of an image using Daubechies'2 wavelets. The family of wavelets are generated from $\phi_r(t)$ and $\psi_r(t)$ by changing the scale and the translation in time (or space in image processing).

As mentioned in [35], Daubechies' orthonormal basis has the following properties:

- $\psi_r$ has the compact support interval $[0, 2r+1]$
- $\psi_r$ has about $r/5$ continuous derivatives
- $\int_{-\infty}^{\infty} \psi_r(t)dt = \ldots = \int_{-\infty}^{\infty} x^r \psi_r(t)dt = 0$.

Due to their properties, Daubechies' wavelets give remarkable results in image analysis and synthesis. In fact, a wavelet function with compact support can be easily implemented by finite length filters. This finite length property is important for spatial domain localization. Furthermore, functions with continuous derivatives allow the continuity to be analyzed more efficiently. Also, edge artifacts can be avoided. Since the mother wavelets are used to characterize details of a signal, they should have a zero integral so that the trend information is stored in the coefficients obtained by the father wavelet.

Any given wavelet transformation is in fact a multiresolution edge detector. Splitting an image into low-pass and three high-pass subbands at three orientations yields edge information in vertical, horizontal and off-diagonal directions. At each orientation, the high frequency parts, such as edges and ridges, are captured according to the frequency of a band. That is, if an edge or a ridge is within the support of a wavelet basis function, the corresponding wavelet coefficient is large. Likewise, the
smooth image region has the corresponding small wavelet coefficients across scale. The wavelet transformation of an $N \times N$ discrete image can be efficiently computed using a filter bank or lifting algorithm in $O(N^2)$ computations. The following properties of the wavelet transforms make them attractive for image analysis.

. Locality: Each wavelet coefficient represents the image content localized in spatial location and frequency.

. Multiresolution: The wavelet transform analyzes the image at a nested set of scales.

. Energy Compaction: The wavelet transforms of real-world images tend to be sparse. A wavelet coefficient is large only if singularities are present within the support of the wavelet.

. De-correlation: The wavelet coefficients of real-world images tend to be approximately decorrelated.

2.7 Wavelet-Based Signal Processing and Applications

We can process signals in the wavelet domain as displayed in Figure 7. At the heart of the discrete wavelet transform are a pair of filters $h_0$ and $h_1$: the low-pass and high-pass filters. The input data are first filtered by $h_0$ and $h_1$, then down-sampled. The same building block is further iterated on the low-pass outputs. The computational complexity of the DWT algorithm for a length-N DWT is $O(N)$. For 2D signals, there are 2D wavelet transforms which consist of two steps: performing 1D wavelet transform on each row of the signal $f$, thus producing a new image $f'$;
and on the new image $f'$, performing the same 1D wavelet transform on each columns.
CHAPTER 3

NORMALIZED CUTS FRAMEWORK

Shi and Malik proposed a novel approach solving the perceptual grouping problem: Normalized Cuts [21]. This approach aims at extracting the global impression of an image rather than focusing on local features and their consistencies in the image data. The normalized cut criterion measures both the total dissimilarity between different groups as well as the total similarity within groups.

In the normalized cuts algorithm, visual grouping is formulated as a graph partitioning problem. The nodes of the graph represent pixels which have to be partitioned. The edges that link the nodes are marked by weights which correspond to the strength of the belonging group. These weights can be assigned for color, intensity, texture and some other features. Graph partitioning can be considered as a task of finding the minimum cut, that is the minimum sum of weights of connections across the groups and the maximum sum of them within the groups.

Shi and Malik’s approach considers the set of points in an arbitrary feature space represented as a weighted graph $G = (V, E)$. The nodes of the graph are the points in the feature space, and edges are formed between each two nodes. The weight on each edge, $w(i, j)$, is a function of the similarity between nodes $i$ and $j$. How to set a criterion for a good partition precision and computation efficiency is the main problem.
of the Normalized Cuts Framework. As the authors pointed out, the normalized cut associated with the graph partitioning methods gives a good global segmentation for images.

Let $A$ and $B$ are partitions of the graph $G=(V,E)$, where $V$ are the nodes and $E$ are the edges. $A \cup B = V$, and $A \cap B = \phi$. The

$$
cut(A, B) = \sum_{u \in A, v \in B} w(u, v)
$$

where $w$ is the weight of the edge between nodes $i$ and $j$. The optimal bi-partitioning of a graph is the one that minimizes this cut value. Some researchers proposed clustering method based on this minimum cut criterion. Their methods can be efficiently achieved by recursively finding the minimum cuts. The methods can produce good
global segmentation on some images. But since the cut defined in Equation 51 will increase with the number of edges going across the two partitioned parts, these methods often partition small sets of isolated nodes out rather than make a good segmentation. Figure 12 shows this case: assuming the edge weights are inversely proportional to the distance between the two nodes, these methods may partition nodes $n_1$ and $n_2$ out because they have a very small value. To solve this problem, Shi and Malik proposed the Normalized Cuts. Instead of looking at the value of total edge weight connecting the two partitions, their measure computes the cut cost as a fraction of the total edge connections to all the nodes in the graph.
\[ Ncut(A, B) = \frac{cut(A, B)}{assoc(A, V)} + \frac{cut(A, B)}{assoc(A, V)} \] (52)

where

\[ assoc(A, V) = \sum_{u \in A, v \in V} w(u, v) \]

is the total weight of connections from nodes in partition A to all nodes in the graph. With this equation, the cut that partitions out small isolated node sets will not have small \( Ncut \) value.

Let us denote weight matrix as \( W \), where \( W_{ij} \) are the weights, and \( D \) is the
diagonal matrix such that

\[ D_{ii} = \sum_j W_{ij} \]

The NCuts problem becomes

\[
\min_{\mathbf{x}} n\text{cut}(\mathbf{x}) = \min_{\mathbf{y}} \frac{\mathbf{y}^T (\mathbf{D} - \mathbf{W}) \mathbf{y}}{\mathbf{y}^T \mathbf{D} \mathbf{y}}
\]

(53)

with two constraints on \( \mathbf{y} \): \( y(i) \in \{1, -b\} \) and \( \mathbf{y}^T \mathbf{D} \mathbf{1} = 0 \), where \( \mathbf{1} \) is an \( N \times 1 \) vector of all ones. \( \mathbf{x} \) is an \( N = |V| \) dimensional indicator vector; \( x_i = 1 \) if node \( i \) is in \( \mathbf{A} \), and \( x_i = -1 \) if node \( i \) is not in \( \mathbf{A} \). \( \mathbf{y} = (1 + \mathbf{x}) - b(1 - \mathbf{x}) \), \( b \) is a constant. By solving
the generalized eigenvalue system

\[(D - W)y = \lambda Dy\]  \hspace{1cm} (54)

we can minimize Equation 53 when \(y\) takes on real values.

Shi and Malik's algorithm needs the input weight matrix, that is the similarity between the weight on the edge connecting each pair of nodes in the processed data. Let us consider a point set case first. The point set is show in Figure 13.

In this case, the weight on the graph edge between point \(i\) and \(j\) is given by \(w_{i,j} = e^{d(i,j)/\sigma_{x}}\), where \(d(i,j)\) is the Euclidean distance between the two points, and
$\sigma_x$ is a scaling factor controlling the spatial proximity. With this weight matrix, we can solve the eigenvalue problem. The solution of the eigenvalue is shown in Figure 14 and the eigenvectors corresponding to the 6 smallest eigenvalues are shown in Figure 15 to Figure 20.

The eigenvector with the second smallest eigenvalue is the bi-partitioning indicator vector used to partition the point set. By recursively partitioning the point set using the eigenvector with the second smallest eigenvalue, we get two consequent segmentation results for the point set case, which are shown in Figure 21 and Figure 22.

As proved in [21], the second smallest eigenvector of this system is the real valued
solution to the normalized cut problem. The solutions based on higher eigenvectors become unreliable, as we can see from the Figure 17 to Figure 20. One key problem of the application of the presented algorithm is the computation of weights. The weight on each edge connecting two nodes in the graph $G = (V, E)$ should reflect the likelihood of the two nodes belonging to the same object or different objects. We can simply use the brightness value of the pixels and/or their spatial locations to define the edge weight, or use the texture information and color information.

### 3.1 Orientation Energy Weighting

In [20], Leung and Malik proposed a new weighting method which provides a way
of incorporating curvilinear grouping into region-based image segmentation. By computing the soft contour information obtained through orientation energy, the authors exploited the powerful cue of contour continuity and the Normalized Cut approach. The region-based approaches can hardly deal with the curvilinear continuity, a very powerful constraint for the image segmentation. The contour-based approaches cannot deal with textured regions easily and decisions are made locally. The authors' goal is to incorporate curvilinear grouping in a region-based setting. They proposed a weighting function which takes contours into account softly through orientation energy. The calculation of the weighting function has the following steps: first, applying elongated Gaussian filters and their Hilbert transforms at different orientations on the
images; second, setting values in the weight matrices containing edge or contour information by calculating energy at different orientations.

\[
\mathcal{F}_1(x, y) = \frac{d^2}{dy^2} \left( \frac{1}{C} \exp\left(\frac{y^2}{\sigma^2}\right) \exp\left(\frac{x^2}{\lambda^2\sigma^2}\right) \right)
\]

(55)

\[
\mathcal{F}_2(x, y) = \text{Hilbert}(\mathcal{F}_1(x, y))
\]

(56)

where \(\sigma\) is the scale and \(\lambda\) is the elongation of the filter. The energy is defined as the square of the coefficients derived from the transformation:

\[
\text{Energy} = (I \ast \mathcal{F}_1)^2 + (I \ast \mathcal{F}_2)^2
\]

(57)
where \( I \) is the intensity of the image to be processed. \( \ast \) is convolution. After this, find the maximum energy of different filtered coefficients at the same position. The maximum energy is used as the feature value. Here we can see the potential of using the wavelet transform: the wavelet transform itself is an algorithm with the properties of elongated filtering at several orientations; besides at a specific scale, we can reduce the data size which is important for the solving of the eigensystem for the Normalized Cut framework. Leung’s and Malik’s approach is based on the spatial domain of images. This means that their algorithm has to deal with a larger data size than that in the wavelet domain. This is the idea and motivation behind the algorithm used in this thesis.
CHAPTER 4

ALGORITHM DESCRIPTION AND RESULTS

Texture feature extraction intends to map differences in spatial domain, frequency domain, or both into different groups of vectors with close distances. Many methods have been developed addressing the feature extraction problem. In this thesis, we use the orientation histogram, a statistically based feature. Table 1 shows some choices we can use for the feature set. The orientation histogram is more suitable for the description of the texture feature. Its statistical and orientation properties fit the texture image segmentation very well. Most of the features in Table 1 are only suitable for the non-textured images.

4.1 Feature Extraction in Our Algorithm

Texture cues are important for image segmentation, because most of the images of natural scenes in some respect are textured or contain regions which can be characterized by repetition, (in fact, pseudo-repetition) of the elementary pattern of local structure. Texture is created by edges and by ridges of different intensity variations at different frequencies. In the wavelet analysis, detail subbands possess high transform energy, which is created by high and moderate intensity variations of texture edges and also by noise. Smooth images with the extended objects exhibit energy decrease towards finer scales whereas textured images increase energy at these scales.

47
This property of the wavelet multiresolution decomposition of the images can be used effectively for the computation of the texture features. Due to the decorrelation properties of the wavelet transform, dependencies of the coefficients are much stronger than dependencies of the pixels in spatial domain. The locality property of the wavelet transform yields the large coefficients, and the compaction property gives the concentration of the high amplitudes across scales. Coefficient processing at the appropriate level of resolution reduces the computational cost. Thus, overall data processing becomes more efficient. The considerable previous research background has shown that particularly natural textures can be efficiently characterized by statistical features, such as mean, variation, covariance, and the high-order moments[21].
Orientation histograms and some other histograms, such as histograms of segment length, intersecting segments, and parallel segment length have been used for texture discrimination. To compute similarity/dissimilarity measure and to feed graph partitioning algorithm, we utilize orientation histograms to model textures at three different orientations.

Consider two adjacent blobs, one with horizontal orientation of local structure and the other with the vertical one, which have to be distinguished. Five histograms for each transform scale are computed as follows: For three detail subbands, we compute horizontal, vertical, and two diagonal histograms. For the low-pass subbands, which are of low frequency content, we compute cross horizontal and vertical histograms.
As the orientation information is provided by detailed subbands, we simply apply the orientation masks, (presented in Figure 10) to extract singularity information. For the filtered smooth subband, we compute the histogram of the coefficients within $n \times n$ block. The pairwise difference between two histograms is evaluated as follows:

$$\chi^2(h_1, h_2) = \frac{1}{2} \sum_{m=1}^{N} \frac{[h_1(m) - h_2(m)]^2}{[h_1(m) + h_2(m)]}$$

(58)

Where $N$ is the dimension of the histograms, $h_i$ is the histogram. We divide the $\chi^2$ by a scaling factor $\sigma_c$ to allow for emphasizing the difference between regions.

In Figure 11, three texture blocks are shown. Their orientation histograms com-
Figure 24: Orientation Histograms of the 2nd Level Wavelet Transform

puted at two decomposition levels are presented in Figure 23 and 24 respectively.

4.2 Weighting Matrix and Results

For the gray scale image in spatial domain, the edge weight is defined as the product of a feature similarity and spatial proximity term[21]:

\[
    w_{ij} = e^{-\frac{\|F_i - F_j\|^2}{\sigma_i^2}} \cdot e^{-\frac{\|X_i - X_j\|^2}{\sigma_x^2}}
\]  

(59)

if

\[
    \|X_i - X_j\|_2 < r
\]
where $r$ is a threshold, $F_i$ is the intensity value of node $i$.

In our algorithm, we define the edge weight $w_{i,j}$ between coefficient $i$ and $j$ as the following:

$$w_{i,j} = e^{-\frac{x^2(h_i,h_j)}{\sigma_c}}$$  \hspace{1cm} (60)

Where $\sigma_c$ is a scaling factor introduced before.

First, we decompose images into two levels of decomposition using Daubechies' 2, which is featured by compact support, and it suites well for our task. Then the
histograms are computed; their differences are scaled and the weight matrix is filled. The block size of $7 \times 7$ is used in our experiments and $\sigma_c=0.017$. Finally, the NCuts algorithm is applied. Figures 28 through 32 display the results of segmentation, based on the above described scheme.

Some other segmentation results can be found at http://www.ee.unlv.edu/~dyao/segmentation.html. We also did experiments on the segmentation algorithm proposed by Malik in [34] using the non-orientation histogram feature. The quality of such a segmentation is inferior when compared with the results yielded by orientation histograms (see Figure 25).

We have probed different features for NCuts framework, such as standard variation of energy within blocks and signatures but have not achieved as good results as in case of orientation histograms. The number of edges and their lengths can be considered as possible candidates, but since wavelet transform yields some artificial artifacts for
<table>
<thead>
<tr>
<th>Feature No.</th>
<th>Feature Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>x-Coordinate</td>
</tr>
<tr>
<td>2</td>
<td>y-Coordinate</td>
</tr>
<tr>
<td>3</td>
<td>Brightness</td>
</tr>
<tr>
<td>4</td>
<td>Directionality</td>
</tr>
<tr>
<td>5</td>
<td>Orientation</td>
</tr>
<tr>
<td>6</td>
<td>Area</td>
</tr>
<tr>
<td>7</td>
<td>Shape</td>
</tr>
<tr>
<td>8</td>
<td>Standard Deviation</td>
</tr>
<tr>
<td>9</td>
<td>Angular second moment</td>
</tr>
<tr>
<td>10</td>
<td>Histogram</td>
</tr>
</tbody>
</table>

Table 1: Feature Set

smooth changes, the problem should be solved by multi-scale edge detection. Also, the continuity of edges should be determined. That is, overall processing time will increase dramatically.

Different smooth wavelet functions have been tested. The results based on Haar and biorthogonal 2.2 are presented in Figures 26, 27.

Figure 27: Segmentation Using Bior 2.2 Wavelet

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Figure 28: Segmentation of the Sofa Image

Figure 29: Segmentation of a Brodatz Image
Figure 30: Segmentation of People Image
Figure 31: Segmentation of a Scene Image
Figure 32: Segmentation of Another Scene Image
CHAPTER 5

CONCLUSION

In this thesis, we proposed a wavelet domain approach for the image segmentation using the Normalized Cuts framework. An algorithm was presented that employs the advantages of the normalized cuts method of solving perceptual grouping problem. Images of natural scenes were treated as textured or possessing some local structure which can be described by statistical features. Five orientation histograms and the weights which measure similarity/dissimilarity were computed by processing transform coefficients obtained from a two-level wavelet decomposition. As a result, good segmentation quality was achieved, and the time complexity was substantially reduced. For an image of $256 \times 256$, the time reduction was a factor of about 50, and it varied for images of different sizes and contents. The results of segmentation were presented that showed a decent quality of segmentation for most of the presented images of natural scenes. We also found that the simplest Haar wavelet gave the fastest segmentation; but like the biorthogonal wavelet, it gave us a worse view of the images than the Daubechies wavelet. Besides, with the filter length increasing in the wavelet transform, the segmentation quality at the border of blobs became worse since the filtering averages characteristics of more pixels.

As mentioned before, the key for the wavelet domain image segmentation is the
feature extraction. In our experiments, we used several well-known features for the computation of the similarity and the dissimilarity. The one leading to the best quality was the Chi-square distance of the orientation histogram for the textured images. But using the histogram had drawbacks when it was used to extract the exact boundaries of different texture blobs, although simply using this feature was better than the use of only one other feature. We can derive that a carefully designed feature combining several features together with different weights will lead us a much better result in finding the precise boundaries. We can see the possibility of the use of other techniques for the refinement of the segmentation, such as applying the zero-crossing labeling for the edges in the wavelet domain and then applying the envelop algorithm to extract the regions between the edges. Another potential for the adaptive segmentation, using our method, is to use the energy relationship of the wavelet coefficients across different decomposition levels. Because different textured regions have different cross-level coefficient energy properties, this can be used as the adaptive factor in our algorithm. We can also apply the shape descriptors in the wavelet domain; thus the algorithm can be used in some special applications, such as target recognition, object location. The color information can be added to the weight matrix to exploit the color images.
BIBLIOGRAPHY


Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.


Bayesian Estimation, and Inverse Problems, (1999); 306-320.


[38] Jia Li and et al, “Multiresolution Object-of-Interest Detection for Images with Low Depth of Field”, IEEE Transactions on Pattern Analysis and Machine Intelligence, (2001); 23(1).


VITA

Graduate College
University of Nevada, Las Vegas

Dongsheng Yao

Local Address:
1600 E University Ave. #208
Las Vegas, NV 89119

Home Address:
No. 73, Xinhua Rd., Yangliuqing, Xiqing District
Tianjin 300380, P. R. China

Degrees:
Bachelor of Science, Electrical Engineering, 1993
Nankai University

Publications:

Thesis Title:
Image Segmentation in the Wavelet Domain using NCuts Framework

Thesis Examination Committee:
Chairperson, Dr. Shahram Latifi, Ph.D.
Committee Member, Dr. Sahjendra Singh, Ph.D.
Committee Member, Dr. Henry Selvaraj, Ph.D.
Graduate Faculty Representative, Dr. Zhiyong Wang, Ph.D.