Wavelets and their Applications in Databases

Daniel Keim, Martin Heczko

Computer and Information Science
University of Constance
Introduction

• Wavelets are

„the (re)discovery of the last decade in Computer Graphics“

• Wavelets have
  – firm mathematical basis
  – nice theoretical properties
  – many practical CS applications
    • Data Compression
    • Computer Graphics & Visualization
    • Databases ...

Overview

1. Introduction
2. Foundations of Wavelet Theory
3. Standard Applications
4. Applications in Database Area
5. Summary and Conclusion
Introduction

• Notion *Wavelet* comes from seismology
• Wavelets describe a wave spreading an impulse

---

1. Introduction

---

Introduction

• *Basic Idea:*  
  **Hierarchical Decomposition** of a function into a set of *Basis Functions* and *Wavelet Functions*
Introduction

• **Example**

<table>
<thead>
<tr>
<th>Resolution</th>
<th>Approximation</th>
<th>Detail-Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1. Introduction

Advantages of Wavelet Transformations

• **Space and Time Efficiency**
  → Low Complexity of DWT

• **Multiresolution Properties**
  → Hierarchical Representation & Manipulation

• **Generality & Adaptability**
  → Different Basis and Wavelet Functions

1. Introduction
History

• 1873 - Karl Weierstraß [Wei95]
  Family of scaled overlapping copies of a Basis Function
• 1910 - Alfred Haar [Haa10]
  Orthonormal system of compact functions (Haar-basis)
• 1946 - Dennis Gabor:
  Non-orthogonal Wavelet basis with unlimited support
• 1980s - A. Grossman, J. Morlet and I. Daubechies:
  Signal Analysis with Wavelets
• 1989 - Stephane Mallat [Mal89] & Yves Meyer:
  Multiresolution Analysis
• 1990+ : Rediscovery of Wavelets in Computer Science

1. Introduction

Wavelets versus Fourier

Wavelet and Fourier
  • signal decomposition in the frequency domain
  • efficiency of FFT and DWT

Fourier
  • unlimited support
  • sinus/cosinus functions (different frequencies)
  • same resolution

Wavelet
  • compact support (→ locality, discontinuities)
  • different basis functions (scaling & translation)
  • multiresolution (higher frequencies ⇒ higher resolution)
Wavelets versus Fourier

Comparison

Fourier

Wavelet

Applications in Computer Science

• Signal Processing
  – Time Series Analysis
  – Noise Reduction

• Data Compression
  – Image
  – Video

• Computer Graphics
  – Multiresolution Data Representation
  – Multiresolution Rendering
  – Multiresolution Data Manipulation
Overview

1. Introduction

2. Foundations of Wavelet Theory
   2.1 Basics of Wavelet Transformations
   2.2 Multiresolution Analysis
   2.3 Advantages of Wavelet Transformation

3. Standard Applications

4. Applications in Database Area

5. Summary and Conclusion

Basics of Wavelet Transformations
(example: Haar Wavelet Transformation)

• Data objects can often be described as **piecewise linear functions**

• *Haar Wavelet Transformation* is a hierarchical decomposition based on the vector space of piecewise linear functions on the interval \([0,1)\)

  Let \(V^j\) be defined as the vector space of all \(2^j\) functions over the intervals

  \([0,1/2^j), ..., [(2^j-1)/2^j, 1)\)
Basics of Wavelet Transformations
(example: Haar Wavelet Transformation)

**Example:**

- 1 data value ⇔ piecewise constant function over \([0, 1)\)
in vector space \(V^0\)
- 2 data values ⇔ function constant over \([0, \frac{1}{2})\) and \(\left[\frac{1}{2}, 1\right)\)
in vector space \(V^1\)
- \(2^j\) data values ⇔ function constant over \(2^j\) subintervals of \([0,1)\) in vector space \(V^j\)
- Observation: \(V^0 \subset V^1 \subset V^2 \subset V^3 \subset \ldots\)

2. Foundations of Wavelet Theory  - 2.1 Basics of the Wavelet Transformation

Wavelet decomposition
Simple Example

- **First step:** sequence \((9\ 7\ 3\ 5)\)
  pair-wise average \((8\ 4)\)
  lost detail information \((1\ -1)\)
[\(8+1=9\ 8-1=7\ 4+(-1)=3\ 4-(-1)=5\)]
- **Next step:** sequence \((8\ 4)\)
  average \((6)\)
  detail \((2)\)

→ wavelet transformation \((6\ 2\ 1\ -1)\)
**Wavelet decomposition**

**Simple Example**

```
average
detail
```

```
\[
\begin{array}{c}
6 \\
8 \\
9 \\
\end{array}
\hspace{1cm}
\begin{array}{c}
+2 \\
+1 \\
-1 \\
\end{array}
\hspace{1cm}
\begin{array}{c}
9 \\
7 \\
3 \\
\end{array}
\hspace{1cm}
\begin{array}{c}
= \\
+2 \\
+1 \\
-1 \\
\end{array}
\hspace{1cm}
\begin{array}{c}
6 \\
\end{array}
\end{array}
\]
```

---

**2. Foundations of Wavelet Theory**

**Basics of Wavelet Transformations**

(example: Haar Wavelet Transformation)

- **Haar Basis Function** (for vector space \( V^j \))

\[
\phi_i^j(x) := \phi(2^j x - i) \quad i = 0, \ldots, 2^j - 1
\]

\[\text{with } \phi(x) := \begin{cases} 
1 & \text{if } 0 \leq x < 1 \\
0 & \text{otherwise}
\end{cases} \]

- **Example:** Haar basis for \( V^2 \)

---

2. Foundations of Wavelet Theory - 2.1 Basics of the Wavelet Transforma
Basics of Wavelet Transformations
(example: Haar Wavelet Transformation)

• Haar Wavelet Function (for vector space $W^j$)

$$\psi_i^j(x) := \psi(2^j x - i) \quad i = 0, \ldots, 2^j - 1$$

with

$$\psi(x) := \begin{cases} 
1 & \text{if } 0 \leq x < 0.5 \\
-1 & \text{if } 0.5 \leq x < 1 \\
0 & \text{otherwise}
\end{cases}$$

• Example: Haar wavelet for $W^j$

---

2. Foundations of Wavelet Theory - 2.1 Basics of the Wavelet Transformation

Example: (9 7 3 5)

\[
\mathcal{R}(x) = c_0^0 \phi(x) + c_1^0 \phi(x) + d_0^0 \psi(x) + d_1^0 \psi(x)
\]

\[
\mathcal{R}(x) = c_0^1 \phi(x) + c_1^1 \phi(x) + d_0^1 \psi(x) + d_1^1 \psi(x)
\]

---

2. Foundations of Wavelet Theory - 2.1 Basics of the Wavelet Transformation
Orthogonal Complement

- $W^j$ is orthogonal complement of $V^j$ in $V^{j+1}$
  $\Rightarrow$ all $f \in W^j$ are orthogonal to all $g \in V^j$

- linear independent $\psi^j_i$ spanning $W^j$
  are the wavelets

$\Rightarrow$ - basis of $V^j$ and $W^j$ form a basis for $V^{j+1}$
  - basis functions of $V^j$ and $W^j$ are orthogonal

Scaling Functions

- Haar basis
- cubic B-spline basis
- Daubechies basis
B-Spline Scaling Functions

2. Foundations of Wavelet Theory - 2.1 Basics of the Wavelet Transformation

B-spline Wavelet Functions

2. Foundations of Wavelet Theory - 2.1 Basics of the Wavelet Transformation
Daubechies Wavelets

2. Foundations of Wavelet Theory - 2.1 Basics of the Wavelet Transformation

Comparison Haar - Daubechies

4. Applications in Databases - 4.2 Similarity Search
Properties

- **Orthogonality**
  - wavelets $\psi$ and basis $\phi$ are orthogonal if
    \[
    \forall i, j, l: \left\langle \phi^j_k \mid \psi^l_i \right\rangle = 0
    \]
  - wavelets $\psi$ are semi-orthogonal if
    \[
    \forall j, k, l : \left\langle \phi^j_k \mid \psi^l_i \right\rangle = 0
    \]
  - wavelets $\psi$ are bi-orthogonal if ($\sim$ indicates dual basis)
    \[
    \forall i, j, l : \left\langle \phi^j_k \mid \tilde{\psi}^l_i \right\rangle = 0 \quad \text{and} \quad \left\langle \psi^j_k \mid \tilde{\phi}^l_i \right\rangle = 0
    \]

- **Normalization**
  - vector $u$ is normalized if $\| u \| = 1$

- **Orthonormality**
  - basis $u_1, u_2, \ldots$ is orthonormal if
    \[
    \forall i, j : \left\langle u_i \mid u_j \right\rangle = \delta_{i,j} \quad \text{with} \quad \delta_{i,j} = 1 \text{ if } i = j \text{ and } 0 \text{ otherwise}
    \]
    (orthonormality = orthogonality + normalization)

2. Foundations of Wavelet Theory - 2.1 Basics of the Wavelet Transformation
Properties

- **Symmetry** of Scaling and Wavelet function (about their center)
- **Compact Support**
- **Smoothness / Differentiability** of the Scaling and Wavelet Functions
  -> compact support and smoothness are conflicting goals
- **Continuous** Wavelet Transformation (CWT) versus **Discrete** Wavelet Transformation (DWT)

---

2. Foundations of Wavelet Theory - 2.1 Basics of the Wavelet Transformation

Properties

Continuous versus Endpoint-Interpolating Wavelets

<table>
<thead>
<tr>
<th>mother wavelet</th>
<th>shift-invariant wavelets</th>
<th>shift-variant wavelets</th>
</tr>
</thead>
</table>

---

2. Foundations of Wavelet Theory - 2.1 Basics of the Wavelet Transformation
2. Foundations of Wavelet Theory - 2.1 Basics of the Wavelet Transformation

Self-Similarity of Wavelets

Extensions to Higher Dimensions

Construction of the basis function

- **Standard**
  corresponds to a transformation of first the rows and then the columns (basis functions corresponds to the Tensor product of $\phi^0_0 \psi^0_0 \psi^1_0 \psi^1_1$)

- **Non-Standard**
  corresponds to a mutual transformation of the rows and the columns
2. Foundations of Wavelet Theory - 2.1 Basics of the Wavelet Transformation

Standard Construction
(2-dim. Haar wavelet)

Non-Standard Construction
(2-dim. Haar wavelet)
Standard Decomposition

Transform rows

Transform columns

Result

2. Foundations of Wavelet Theory - 2.1 Basics of the Wavelet Transformation

Non-Standard Decomposition

Transform rows

Transform columns

Result

2. Foundations of Wavelet Theory - 2.1 Basics of the Wavelet Transformation
2.2 Multiresolution Analysis

• Scaling Function and Wavelets can be used to decompose data into components of multiple resolutions

• Transformation is efficient since it can be performed by Matrix Operations (for bounded domain)

• Transformation is reversible

  in the following:
  wavelets on bounded domain \( \Rightarrow V^j \) has a finite basis

\[ \begin{align*}
  V_0 & \subset V_1 \subset V_2 \subset V_3 \subset \ldots \\
\end{align*} \]

• Basis of Multiresolution Analysis:
  nested set of linear function spaces

  \[ V^0 \subset V^1 \subset V^2 \subset V^3 \subset \ldots \]

• Wavelet spaces \( W^j \)
  are the complement of \( V^j \) in \( V^{j+1} \)

  (orthogonality not required)
Multiresolution Analysis

• Matrix Notation of Scaling and Wavelet Functions
  \[ \Phi^j = \begin{bmatrix} \phi_0^j & \cdots & \phi_{v(j)-1}^j \end{bmatrix} \quad \Psi^j = \begin{bmatrix} \psi_0^j & \cdots & \psi_{w(j)-1}^j \end{bmatrix} \]

• nested function spaces
  \[ \Phi^{j-1}(x) = \Phi^j(x) \cdot P^j \quad \Psi^{j-1}(x) = \Phi^j(x) \cdot Q^j \]

• two-scale relation for scaling functions and wavelets:
  \[ [\Phi^{j-1} | \Psi^{j-1}] = \Phi^j \begin{bmatrix} P^j | Q^j \end{bmatrix} \]

2. Foundations of Wavelet Theory - 2.2 Multiresolution Analysis

Two-Scale Relation for Haar Basis

\[ [\Phi^1 | \Psi^1] = \Phi^2 \begin{bmatrix} P^2 | Q^2 \end{bmatrix} \]

\[ [\phi_0 \phi_1 | \psi_0 \psi_1] = \begin{bmatrix} \phi_0^2 \phi_2 \phi_3 \phi_1 \\ \phi_0^2 \phi_1 \phi_2 \phi_3 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 1 & 0 \\ 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & -1 \end{bmatrix} \]

2. Foundations of Wavelet Theory - 2.2 Multiresolution Analysis
2. Foundations of Wavelet Theory - 2.2 Multiresolution Analysis

**Analysis Filter**

- Matrices ($A^j$ and $B^j$) can be used to decompose the data.
- Data $c^j$ can be decomposed by $A^j$ and $B^j$ into low-resolution part $c^{j-1}$ and detail part $d^{j-1}$.
- Decomposition can be applied recursively to $c^{j-1}$.

\[
\begin{align*}
A^j & \quad c^j \\
B^j & \quad d^j
\end{align*}
\]

\[
\begin{align*}
A^{j-1} & \quad c^{j-1} \\
B^{j-1} & \quad d^{j-1}
\end{align*}
\]

\[
\begin{align*}
A^1 & \quad c^1 \\
B^1 & \quad d^0
\end{align*}
\]

2. Foundations of Wavelet Theory - 2.2 Multiresolution Analysis

**Analysis and Synthesis Filters**

- Analysis Filters: $c^{j-1} = A^j c^j$ \quad $d^{j-1} = B^j c^j$

  satisfying: $[\Phi^{j-1} | \Psi^{j-1}] \begin{bmatrix} A^j \\ B^j \end{bmatrix} = \Phi^j$

- Synthesis Filters: $c^j = P^j c^{j-1} + Q^j d^{j-1}$

\[
\begin{bmatrix} A^j \\ B^j \end{bmatrix} \quad = \quad [P^j \mid Q^j]^{-1}
\]

2. Foundations of Wavelet Theory - 2.2 Multiresolution Analysis
Analysis Filter for Haar Basis

Analysis Filter:

\[
A^2 = \frac{1}{2} \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}
\]

\[
B^2 = \frac{1}{2} \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix}
\]

**Example:**

\[
\begin{bmatrix} 9 & 7 & 3 & 5 \end{bmatrix}^T
\]

\[
\begin{array}{c}
A^2 & \rightarrow & \begin{bmatrix} 6 \\ 4 \end{bmatrix} \\
B^2 & \rightarrow & \begin{bmatrix} 1 \\ -1 \end{bmatrix}
\end{array}
\]

Semi- & Bi-orthogonal Wavelets

- *Orthogonal Basis* is no requirement for Multiresolution Analysis

- *Semi-orthogonal* or *Bi-orthogonal Basis* are sufficient, sometimes even better since they can be constructed to be sparse

**Definitions:**

- wavelets \( \psi \) are semi-orthogonal if

\[
\forall j, k, l : \langle \phi_k^j | \psi_l^j \rangle = 0
\]

- wavelets \( \psi \) are bi-orthogonal if (~ indicates dual basis)

\[
\forall i, j, l : \langle \phi_i^j | \tilde{\psi}_l^j \rangle = 0 \quad \land \quad \langle \psi_i^j | \tilde{\phi}_l^j \rangle = 0
\]
2.3 Advantages of the Wavelet Transformation

- Generality of the Transformation
  (Generalization of other Transformations)

- Adaptability of the Transformation
  (Different Basis Functions allow different Properties of the Transformation)

- Transformation is Hierarchical
  (Multiresolution - Properties)

- Transformation is Loss-Free

- Efficiency of the Transformation
  (Linear Time and Space Complexity for Orthogonal Wavelets)

⇒ Advantages translate into specific advantages in the different applications

2. Foundations of Wavelet Theory - 2.3 Advantages of the Wavelet Transformation

Overview

1. Introduction
2. Foundations of Wavelet Theory
3. Standard Applications
   3.1 Signal Processing
   3.2 Data Compression
   3.3 Computer Graphics
4. Applications in Database Area
5. Summary and Conclusion
3. Standard Applications

3.1 Signal Processing

- Signal Filtering (Linear / Non-Linear)
  - Noise-Reduction (De-Noising)
  - Speckle-Reduction (De-Speckling)
- Edge-Detection
- Multi-Resolution Editing
3. Standard Applications - 3.1 Signal Processing

Noise-Reduction

before noise-reduction

after noise-reduction

3. Standard Applications - 3.1 Signal Processing
3. Standard Applications - 3.1 Signal Processing

Speckle-Reduction

before reduction  

after reduction

Edge-Detection

without wavelet transformation  

with wavelet transformation

3. Standard Applications - 3.1 Signal Processing
Multiresolution Editing

3. Standard Applications  -  3.1 Signal Processing

Wavelet Filtering of Low-Res. Signal
Editing of Low-Res. Signal
Re-Application of Detail Signal

Principle Idea

3. Standard Applications  -  3.2 Data Compression

Basic Idea

Data Transformation
Lossy or Loss-less

Quantization
Lossy

Run-length Coding
Loss-less

Entropy Coding
Loss-less

Compressed Data

3. Standard Applications  -  3.2 Data Compression
Data Compression

- Transformation
  - Discrete Cosine Transformation (JPEG)
  - Wavelet Transformation (JPEG2000)

- Lossy Compression
  - Scalar Quantization
  - Vector Quantization

- Loss-less Compression
  - Run-Length Coding
  - Entropy Coding: Huffman or Arithmetic Coding

3. Standard Applications - 3.2 Data Compression

Data Compression

- Data Storage

3. Standard Applications - 3.2 Data Compression
Data Compression

• Data Storage

Example:
Wavelet Image Compression Algorithm

procedure Compress(c: array [1..m] of reals; ε: real)
    \[ \tau_{\min} \leftarrow \min\{|e[i]|\} \]
    \[ \tau_{\max} \leftarrow \max\{|e[i]|\} \]
    do
        \[ \tau \leftarrow (\tau_{\min} + \tau_{\max})/2 \]
        \[ s \leftarrow 0 \]
        for \( i \leftarrow 1 \) to \( m \) do
            if \( |e[i]| < \tau \) then \( s \leftarrow s + |e[i]|^2 \)
        end for
        for \( i \leftarrow 1 \) to \( m \) do
            if \( s < \epsilon^2 \) then \( \tau_{\min} \leftarrow \tau \) else \( \tau_{\max} \leftarrow \tau \)
        end for
    until \( \tau_{\min} = \tau_{\max} \)
    for \( i \leftarrow 1 \) to \( m \) do
        if \( |e[i]| < \tau \) then \( e[i] \leftarrow 0 \)
    end for
end procedure

\( c[i] \): coefficients of Wavelet Transformation
\( \tau \): threshold with error \( \epsilon \)
Wavelet Image Compression

3. Standard Applications - 3.2 Data Compression

Original

1:50

1:100

1:200

3. Standard Applications - 3.2 Data Compression

Wavelet Image Compression

Original

1:6 Compression
3. Standard Applications - 3.2 Data Compression
Example:
FBI Fingerprint Compression

3. Standard Applications - 3.2 Data Compression

Wavelet Image Compression

Wavelet Compression
Fourier Compression

3. Standard Applications - 3.2 Data Compression
Wavelet Image Compression

Advantages

• loss-free (up to 1:5) or lossy (up to 1:200)

• improved image quality (at same compression ratio)

• flexible image format
  (different resolutions, partial password protection)

• fast image preview, successive image loading
  (⇒ important feature for internet applications!)

Example

3. Standard Applications - 3.2 Data Compression

3.3 Computer Graphics

• Approximation of Curves and Surfaces

• Multiresolution Data Representation

• Multiresolution Manipulation of Images, Curves, and Surfaces

Principle Idea
Approximation of Surfaces

In rendering applications, approximations with different resolutions are used depending on the distance from the viewer.

Multiresolution Image Manipulation

Original

Zoom out by a Factor of 100,000

Adding Smog

Changed Original


Multiresolution Curve Manipulation

Original Spline Surface

Changes at narrow Scale

Changes at intermediate Scale

Changes at broad Scale

Multiresolution Surface Manipulation

Original

Compressed to 16%

Changed at Coarse Level

Changes at Fine Level

Overview

1. Introduction
2. Foundations of Wavelet Theory
3. Standard Applications
4. Applications in Database Area
   4.1 Efficient Data Processing
   4.2 Similarity Search
5. Summary and Conclusion
4. Applications in Databases

4.1 Efficient Data Processing

- Selectivity Estimation
  - Wavelet-based Histograms [MVW 98, MVW 00]

- Approximate Query Processing
  - Data Cubes [VWI 98, VW99]
  - Relational Databases [CGRS00]

- Clustering Techniques
  - WaveCluster [SCZ98, SCZ00]
Selectivity Estimation:
Wavelet-based Histograms [MVW98, MVW00]

- **Goal**
  - Compact histograms
  - Accurate estimation of the data distribution

- **Approach**
  - compute extended cumulative data distribution
    (cumulative data distribution with zero values)
  - compute Haar (or linear) wavelet transformation
  - thresholding methods:
    1. Take largest wavelet coefficients
    2. Take wavelet coefficients which lead to a large error reduction
    3. Throw away wavelet coefficients whose deletion lead to the small error increase

4. Applications in Databases - 4.1 Efficient Data Processing
Selectivity Estimation: Wavelet-based Histograms [MVW98, MVW00]

Average Absolute Error Results (depending on storage size)

Approximate Query Processing: Data Cube [VWI98, VW99]

Goal

- compact representation of data cube
- efficient support of partial range-sum queries
- I/O-efficient construction of the wavelet decomposition
  -> construction directly from the sparse representation of orig. cube

Approach

1. Haar wavelet transformation of the data cube
2. thresholding according to storage and accuracy requirements
3. inverse wavelet transformation to determine the results
Approximate Query Processing: Data Cube [VWI98]

Ideas

– partial sum cube
  
  **Motivation:** partial sum cube is monotonously increasing
  ⇒ wavelet decomposition provides better results

– cell-wise logarithmic transformation of the cube

– chunked data processing

Approximate Query Processing: Data Cube [VW99]

– **Haar Wavelet Decomposition**
  
  • based directly on the sparse repres. of the original cube
  • result: C’ coefficients (~ non-zero entries of the cube)

– **Thresholding and Ranking**
  
  • Threshold to reduce coefficients to C << C’
  • Ranking according to importance for answering typical (range) queries by weighting the coefficients

– **Reconstruction**
  
  • Determining the k < C’ coefficients which are most important for answering the aggregation query
  • k determines the accuracy of the answer
Approximate Query Processing: Data Cube [VW99]

Relative Error Results (depending on storage size)

4. Applications in Databases - 4.1 Efficient Data Processing

Approximate Query Processing: Relational Databases [CGRS00]

- **Wavelet Decomposition**
  - Non-Standard Haar-Wavelet Decomposition of the d-dim. array of attribute value combinations (joint frequency distribution)
  - Thresholding to retain highest coefficients

- **Processing of Relational Algebra Operations in the Wavelet-Coefficient Domain**
  - Select-Operation
  - Project-Operation
  - Join-Operation

- **Determining the Approximate Results („Rendering“)**
  - Mapping Results in Wavelet-Domain to Relational Tuples

4. Applications in Databases - 4.1 Efficient Data Processing
Approximate Query Processing
Relational Databases [CGRS00]

4. Applications in Databases - 4.1 Efficient Data Processing

Relative Error Results
(depending on number of coefficients used)

Select Query  Select-Sum Query
Clustering Techniques:
WaveCluster [SCZ98, SCZ00]

Basic Idea
- Partitioning the data space by a grid reduces the number of data objects but induces a small error
- Application of the wavelet-transformation to the reduced feature space provides multiresolution data representation
- Finding the connected components can be performed at different resolutions
- Compression of the grid is crucial for the efficiency
  (→ Does not work in high dimensional space!)

Clustering Techniques:
WaveCluster [SCZ98, SCZ00]

Clustering based on a wavelet approximation of the data

Algorithm:

Input: Multidimensional data objects' feature vectors
Output: clustered objects

1. Quantize feature space, then assign objects to the units.
2. Apply wavelet transform on the feature space.
3. Find the connected components (clusters) in the subbands of transformed feature space, at different levels.
4. Assign label to the units.
5. Make the lookup table.
6. Map the objects to the clusters.
4. Applications in Databases - 4.1 Efficient Data Processing

Clustering Techniques: WaveCluster [SCZ98, SCZ00]

- Effect of Wavelet Transformation

- Arbitrary shaped clusters found by WaveCluster

Results From Clustering on different Resolution Levels

4. Applications in Databases - 4.1 Efficient Data Processing
4.2 Similarity Search

• Similarity Search in Time-Series Databases

  [CF99, SS99a, SS99b]

• Similarity Search in Image Databases

  – WBIIS [WWFW97a, WWFW97b]
  – WALRUS [NRS99]
  – Windsurf [ABP99]
  – Multiresolution Search [Hec99]
  – Similarity Measure Learning [BVGS99]
  – WIPE [WWF97]

---

Similarity Search in Time Series Databases

• **Goal**

  – time series matching and retrieval
  – new similarity measures (→ effectiveness)
  – fast computation of the similarity measure (→ efficiency)

• **General Approach**

  – Haar wavelet transformation
  – Similarity measure defined in Wavelet Space

---
Similarity Search in Time Series Databases [CF99]

• **Specific Goals**
  – efficient support for range and k-NN queries
  – allow for vertical shifts

• **Approach**
  – wavelet coefficients of sliding window are stored in an index
  – Range and k-NN queries are computed based on the index

• **Results**
  – approach outperforms discrete Fourier transform

---

Results

Accuracy

![Graph 1](image1)

Page Accesses

![Graph 2](image2)
Similarity Search in Time Series Databases [SS99a,b]

- **Goal**
  - new similarity measure for data mining applications
  - effectiveness and efficiency

- **Approach**
  - only sign change and maximum information (Hölder exponent) of the wavelet transformed data is stored
  - step-wise hierarchical comparison for correlations
  - time shifts are considered

---

4. Applications in Databases - 4.2 Similarity Search

**Similarity Search in Time Series Databases [CF99]**

- **Time Series Data**
- **Sign Information** (6 levels)
- **Hölder Decomposition** (5-levels)

---

4. Applications in Databases - 4.2 Similarity Search
Content-based Similarity Search
- WBIIS [WWFW97a, WWFW97b]
- WALRUS [NRS99]
- Windsurf [ABP99]
- Multiresolution Search [Hec99]
- Similarity Measure Learning [BVGS99]
- WIPE [WWF97]

• Goal
  - Content-based Similarity Search in Large Image Databases
  - Improved Recall without explicit Object Recognition

• General Approach
  - Wavelet Transformation to extract compact Feature Vectors
    (possibly more than one per image)
  - Post-processing in Feature Space (e.g., Clustering of Feature Vectors)
  - Search on Feature Vectors using Index Structure / Linear Scan
  - Support of different Similarity Measures

4. Applications in Databases - 4.2 Similarity Search
• **Goal**
  - Improved Content-based Retrieval:
    Partial Image Retrieval & Sketch Retrieval

---

**Approach**
- Wavelet transformation for each color component using Daubechies-8 Wavelets
- Low Frequency Wavelet Coefficients and their Variance are stored as Feature Vectors
- 2-step retrieval:
  - pre-selection (filtering) based on variance (→ candidates)
  - similarity computation based on full feature vectors of candidates
- Extension: Two-level Multi-Resolution Similarity Search
Similarity Search in Image Databases
WBIIS [WWFW 97a,b]

Partial Match Retrieval Examples

4. Applications in Databases - 4.2 Similarity Search

Similarity Search in Image Databases
WBIIS [WWFW 97a,b]

• Retrieval by Sketch Example
Similarity Search in Image Databases

WALRUS [NRS99]

• Goal
  – Content-based Similarity Search in Large Image Databases:
    Invariance w.r.t. Translation and Scaling of Regions in Image

• Approach
  – Haar Wavelet Transformation of sliding window of varying size
  – Clustering of Signatures in Wavelet Space (BIRCH)
    => variable Number of Signatures per Image
  – Storage of Centroids of Clusters in Index Structure (R*-Tree)
  – Similarity Search: Matching Pairs of Signatures (largest overlap)

4. Applications in Databases - 4.2 Similarity Search

WALRUS Results

WBIIS Results
• **Goal**
  – Content-based Similarity Search in Large Image Databases:
    Partial Similarity based on Image Regions

• **Basic Idea**

  4. Applications in Databases - 4.2 Similarity Search

---

**Similarity Search in Image Databases**

**Windsurf [ABP99]**

**Approach**

– Haar Wavelet Transformation of each color channel

– Partitioning of the Image based on Clustering the three color coefficients (k-means clustering on 3rd subband wavelet coefficients)

– Feature Vectors correspond to Regions found in Clustering Step:
  (Size, Centroids, Covariance Matrix of Pixels in Region)

– Similarity Retrieval based on Matching Regions

---

4. Applications in Databases - 4.2 Similarity Search
4. Applications in Databases - 4.2 Similarity Search

Similarity Search in Image Databases

Windsurf [ABP99]

Results of comparative System

4. Applications in Databases - 4.2 Similarity Search
• **Goal**
  – Content-based Similarity Search in Large Image Databases:
    Flexible Similarity Search which allows
    – global / detail matching
    – texture / structure matching

• **Approach**
  – Haar Wavelet Transformation of Color Histograms
  – Storage of Wavelet Coefficients on different resolutions as Feature Vectors
  – Search based on arbitrary Combinations of Wavelet Coefficients
4. Applications in Databases - 4.2 Similarity Search

Comparison to other Approaches: Effectiveness Results
4. Applications in Databases - 4.2 Similarity Search

- **Goal**
  - Content-based Similarity Search in Large Image Databases: Interactive Learning of the Similarity Measure

- **Approach**
  - Vector median Filtering (to remove noise)
  - Haar Wavelet Transformation
  - Storage of 128 largest coefficients (quantized to +1 / -1)
  - Supervised learning to find similarity measure to find weighting for feature vector comparison

**Example Results**

4. Applications in Databases - 4.2 Similarity Search
Similarity Search in Image Databases
WIPE [WWF97]

• **Goal**
  – WIPE: Wavelet Image Pornography Elimination
  -> Fast Special Purpose Image Filtering

• **Approach**
  – Normalization of Images to Standard Size
  – Wavelet Transformation using Daubechies-3 Wavelets
  – Edge Detection in Different Subbands of Wavelet Transformation
  – Feature Vectors used for Similarity Matching:
    Central Moments & Invariant Moments & Color Histograms
  – Filtering based on Training the Search for the desired filtering

Results: 95% Correct Images found with 10% Wrong Rejects
5. Summary and Conclusion

• Wavelet Transformations
  – improve the Efficiency of existing Methods (→ faster)
  – improve the Effectiveness of existing Methods (→ better)
  – enable new Applications (→ new)

• Successful Applications
  – Signal Processing (Noise Reduction, Edge Detection, ...)
  – Data Compression (Image & Video Compression)
  – Multi-Resolution Data Representation and Manipulation (Computer Graphics)
5. Summary and Conclusion

- **Database Applications**
  - Wavelets often only used for an efficient pre-processing
  - often only wavelet coefficients of one resolution are used

- **Future Research Directions**
  - Loss-free Database Compression
  - Approximate Query Results
  - Multi-Resolution Data Analysis

- **Potential New Application**
  - New Similarity Search Applications (Video, ...)
  - Fast Approximate Data Mining
  - Fast Approximate Information Retrieval

---

Bibliography


Bibliography


Bibliography


