# 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 ...

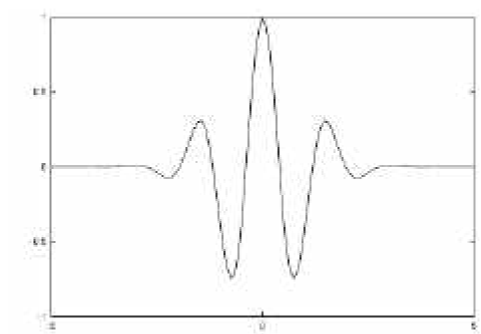


# Overview

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

# Introduction

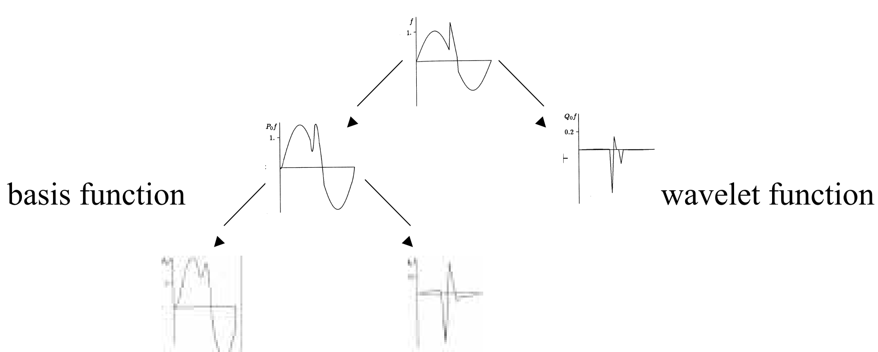
- Notion *Wavelet* comes from seismology
- Wavelets describe of a wave spreading an impulse





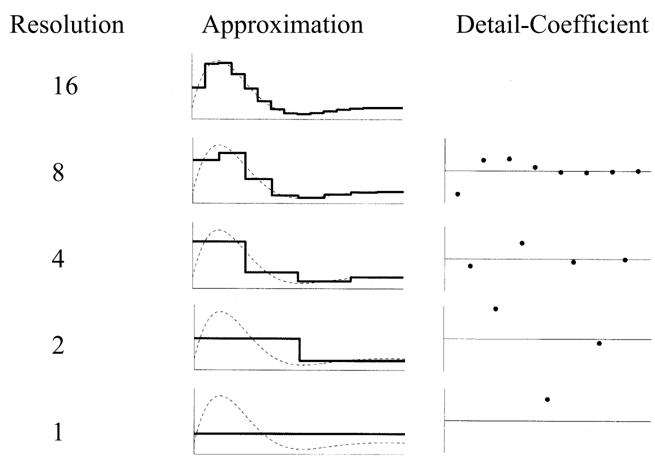
# Introduction

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



# Introduction

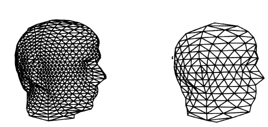
- *Example*



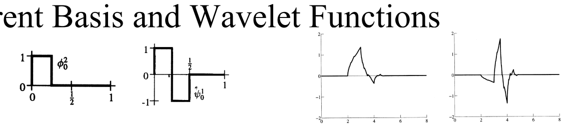



# Advantages of Wavelet Transformations

- Space and Time Efficiency
  –> Low Complexity of DWT



- Multiresolution Properties
  –> Hierarchical Representation & Manipulation
- Generality & Adaptability
  –> Different Basis and Wavelet Functions



# 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

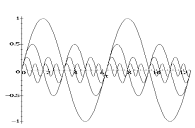


# Wavelets versus Fourier

**Wavelet and Fourier**

- signal decomposition in the frequency domain
- efficiency of FFT and DWT

**Fourier**



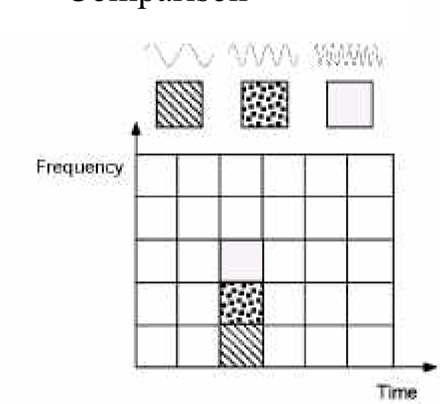
- unlimited support
- sinus/cosinus functions (different frequencies)
- same resolution

**Wavelet**

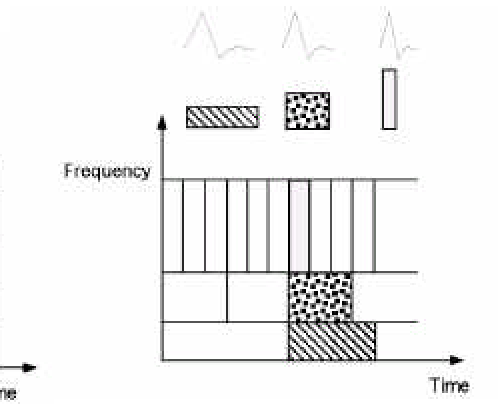
- compact support (–> locality, discontinuities)
- different basis functions (scaling & translation)
- multiresolution (higher frequencies => higher resolution)

# Wavelets versus Fourier

Comparison




Fourier

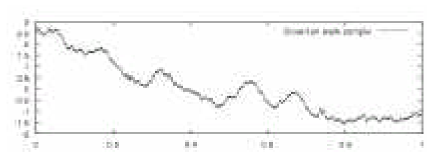



Wavelet

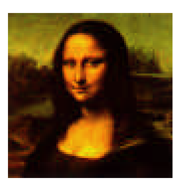


# Applications in Computer Science

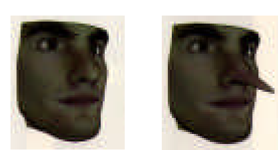
- Signal Processing
  - Time Series Analysis
  - Noise Reduction



- Data Compression
  - Image
  - Video



- Computer Graphics
  - Multiresolution Data Representation
  - Multiresolution Rendering
  - Multiresolution Data Manipulation



# Overview



# 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

$$[0,1/2^j), \; \ldots, \; [(2^j-1)/2^j, 1)$$

# Basics of Wavelet Transformations
(example: Haar Wavelet Transformation)

***Example:***

- 1 data value $\Leftrightarrow$ piecewise constant function over [0, 1)
  in vector space $V^0$
- 2 data values $\Leftrightarrow$ function constant over [0, ½) and [½,1)
  in vector space $V^1$
- $2^j$ data values $\Leftrightarrow$ function constant over $2^j$ subintervals
  of [0,1) in vector space $V^j$
- Observation: $V^0 \subset V^1 \subset V^2 \subset V^3 \subset \ldots$

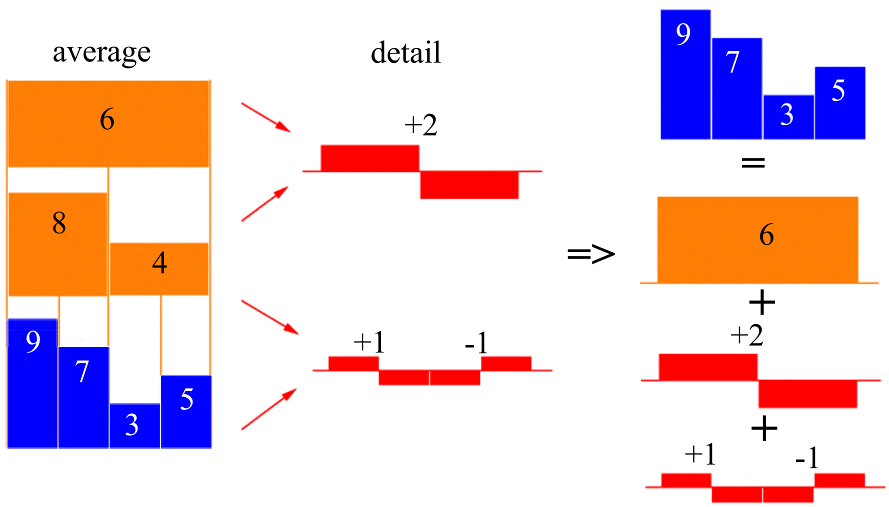


# Wavelet decomposition
Simple Example

- 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





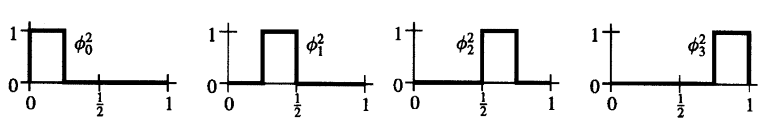

# Basics of Wavelet Transformations
## (example: Haar Wavelet Transformation)

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

$$f_i^j(x) := f(2^j x - i) \qquad i = 0, \ldots, 2^j - 1$$

$$\text{with} \quad f(x) := \begin{cases} 1 & \text{if } 0 \le x < 1 \\ 0 & \text{otherwise} \end{cases}$$

- <u>Example:</u> Haar basis for $V^2$



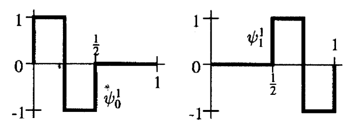
# Basics of Wavelet Transformations
(example: Haar Wavelet Transformation)

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

$$y_i^j(x) := y\,(2^j x - i) \qquad i = 0, \ldots, 2^j - 1$$

$$with \quad y\,(x) := \begin{cases} 1 & if\ \ 0 \le x < 0.5 \\ -1 & if\ \ 0.5 \le x < 1 \\ 0 & otherwise \end{cases}$$

- <u>Example:</u> Haar wavelet for $W^1$

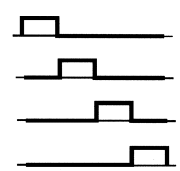





# Example: ( 9 7 3 5 )

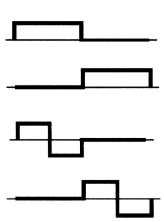
$$\mathcal{I}(x) = c_0^2\,\phi_0^2(x) + c_1^2\,\phi_1^2(x) + c_2^2\,\phi_2^2(x) + c_3^2\,\phi_3^2(x)$$

$\mathcal{I}(x) = 9\times$, $+7\times$, $+3\times$, $+5\times$



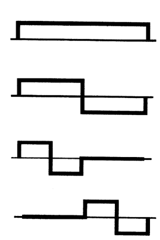
$$\mathcal{I}(x) = c_0^1\,\phi_0^1(x) + c_1^1\,\phi_1^1(x) + d_0^1\,\psi_0^1(x) + d_1^1\,\psi_1^1(x)$$

$= 8\times$, $+4\times$, $+1\times$, $+-1\times$



$$\mathcal{I}(x) = c_0^0\,\phi_0^0(x) + d_0^0\,\psi_0^0(x) + d_0^1\,\psi_0^1(x) + d_1^1\,\psi_1^1(x)$$

$= 6\times$, $+2\times$, $+1\times$, $+-1\times$



# Orthogonal Complement

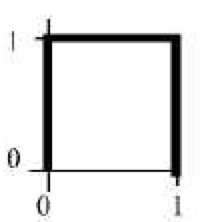
- $W^j$ is orthogonal complement of $V^j$ in $V^{j+1}$
  $\Rightarrow$ all $f \in W^j$ are orthogonal to all $g \in V^j$

- linear independent $y_i^j$ spanning $W^j$
  are the ***wavelets***

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

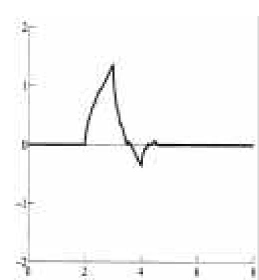


# Scaling Functions



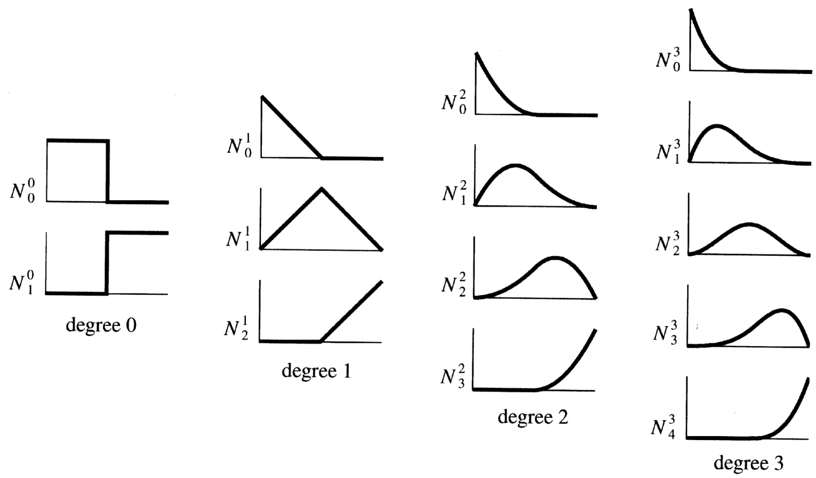

Haar basis

cubic B-spline basis



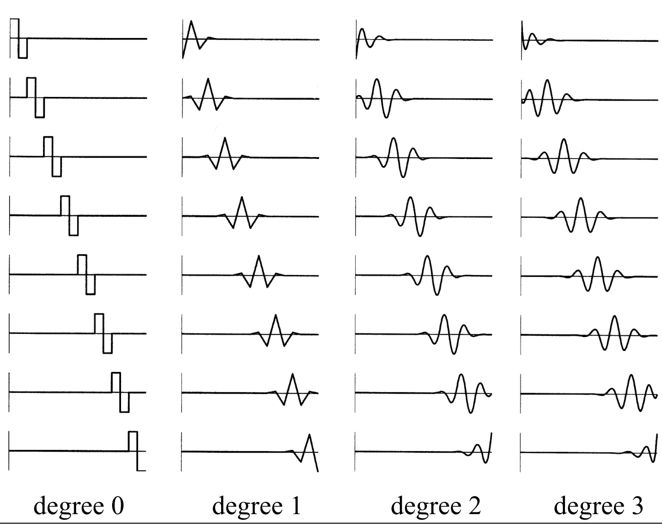
Daubechies basis

# B-Spline Scaling Functions



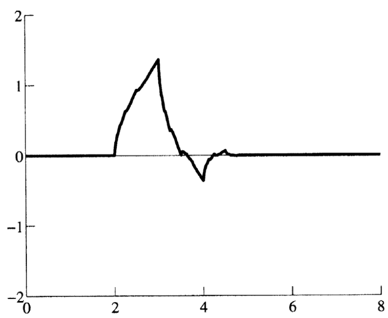

2. Foundations of Wavelet Theory - 2.1 Basics of the Wavelet Transformation

# B-spline Wavelet Functions

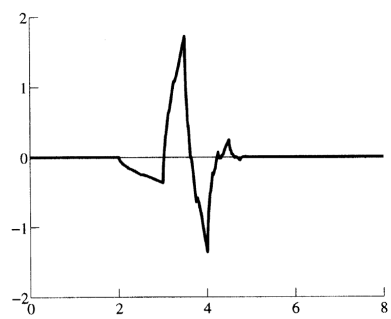



2. Foundations of Wavelet Theory - 2.1 Basics of the Wavelet Transformation

## Daubechies Wavelets



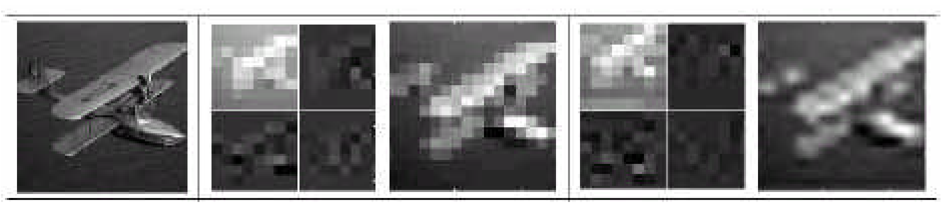

Daubechies Basis



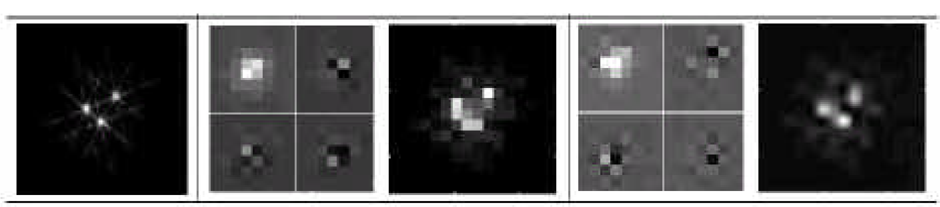

Daubechies Wavelet



## Comparison Haar - Daubechies



Original | Haar | Image Reconstr. | Daubechies-8 | Image Reconstr.



Original | Haar | Image Reconstr. | Daubechies-8 | Image Reconstr.

# Properties

- Orthogonality
  - wavelets $y$ and basis $f$ are orthogonal if

    $$\forall i, j, l: \left\langle f_k^j \,|\, y_l^j \right\rangle = 0$$

    $$\forall j, k, l: \left\langle f_k^j \,|\, f_l^j \right\rangle = \begin{cases} c & if\ \ k = l \\ 0 & otherwise \end{cases} \quad \wedge \quad \left\langle y_k^j \,|\, y_l^j \right\rangle = \begin{cases} c & if\ \ k = l \\ 0 & otherwise \end{cases}$$

  - wavelets $y$ are semi-orthogonal if

    $$\forall j, k, l: \left\langle f_k^j \,|\, y_l^j \right\rangle = 0$$

  - wavelets $y$ are bi-orthogonal if (~ indicates dual basis)

    $$\forall i, j, l: \left\langle f_k^j \,|\, \tilde{y}_l^j \right\rangle = 0 \quad \wedge \quad \left\langle y_k^j \,|\, \tilde{f}_l^j \right\rangle = 0$$



# Properties

- Normalization
  - vector $u$ is normalized if $\| u \| = 1$
- Orthonormality
  - basis $u_1$, $u_2$, ... is orthonormal if

    $$\forall i, j: \left\langle u_i \,|\, u_j \right\rangle = d_{i,j} \quad with\ \ d_{i,j} = 1\ \ if\ \ i = j\ \ and\ \ 0\ otherwise$$

    (orthonormality = orthogonality + normalization)

# Properties

- **Symmetry** of Scaling and Wavelet function (about their center)
- **Compact Support**
- **Smoothness / Differentiability** of the Scaling and Wavelet Functions

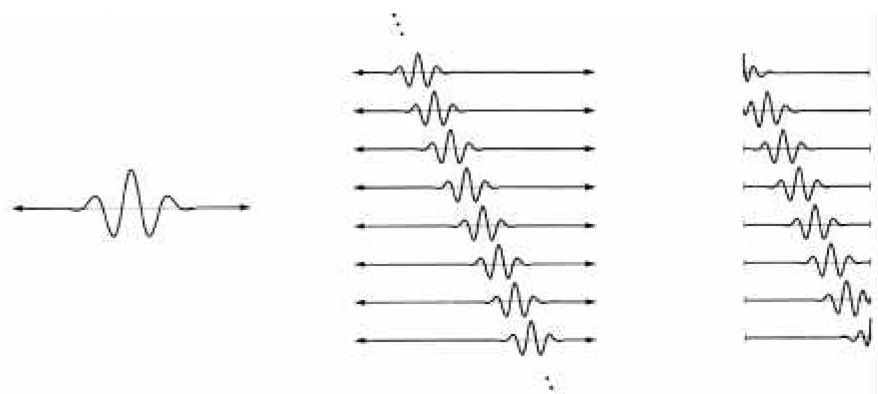
  –> compact support and smoothness are conflicting goals

- **Continuous** Wavelet Transformation (CWT) versus **Discrete** Wavelet Transformation (DWT)



# Properties

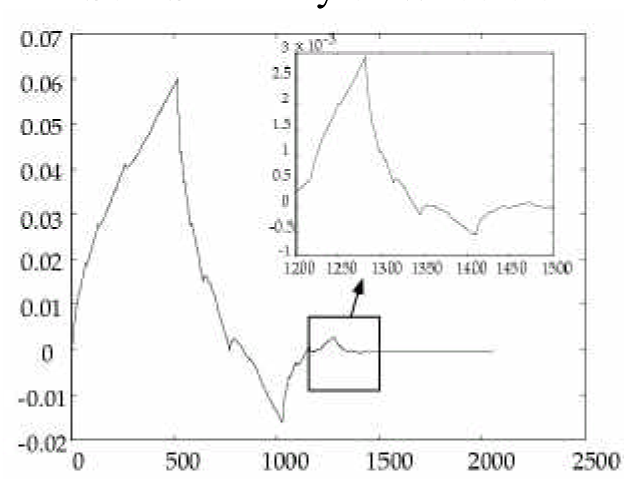
Continuous versus Endpoint-Interpolating Wavelets



mother wavelet | shift-invariant wavelets | shift-variant wavelets

# Properties

Self-Similarity of Wavelets






# Extensions to Higher Dimensions

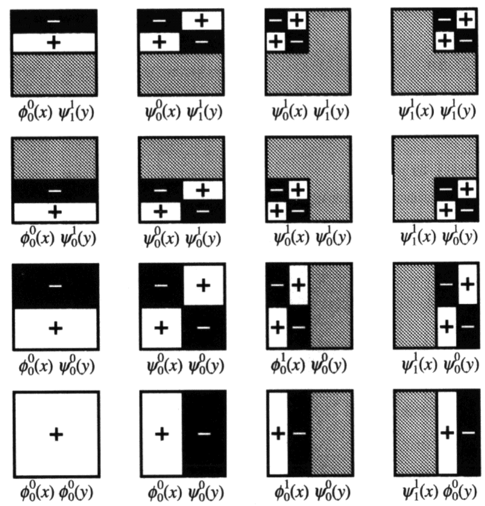
Construction of the basis function

– **Standard**
  corresponds to a transformation of first the rows and then the columns (basis functions corresponds to the Tensor product of $f_0^0 y_0^0 y_0^1 y_1^1$ )

– **Non-Standard**
  corresponds to a mutual transformation of the rows and the columns

# Standard Construction

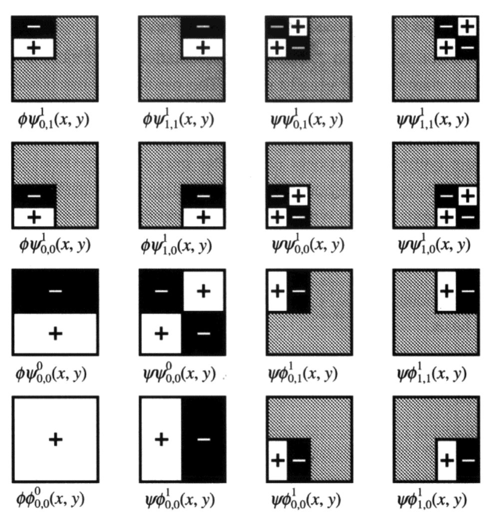
## (2-dim. Haar wavelet)





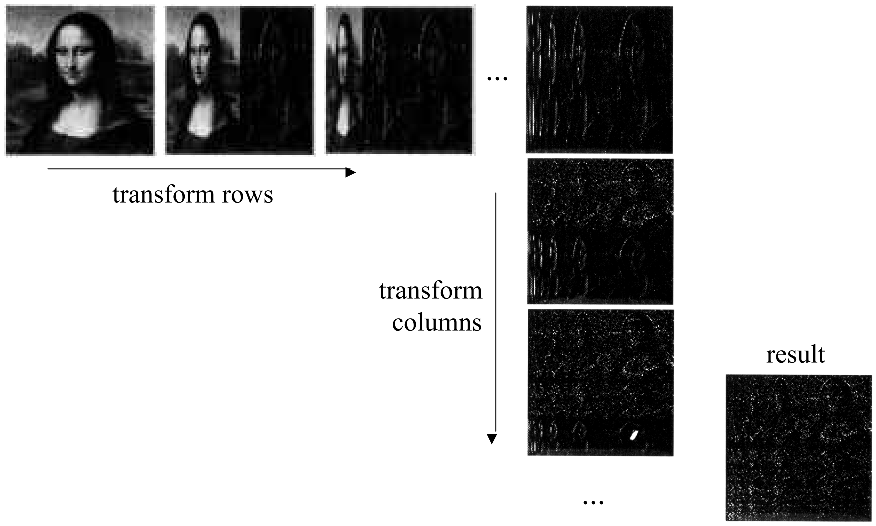

# Non-Standard Construction

## (2-dim. Haar wavelet)




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

## Standard Decomposition






## Non-Standard Decomposition

## 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

<u>*in the following:*</u>
wavelets on bounded domain $\Rightarrow V^j$ has a finite basis



## Multiresolution Analysis

- 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

$$\Phi^j = \left[\phi_0^j \cdots \phi_{v(j)-1}^j\right] \qquad \Psi^j = \left[\psi_0^j \cdots \psi_{w(j)-1}^j\right]$$

- nested function spaces

  $\Rightarrow$ Scaling and Wavelet Functions are refinable

$$\Phi^{j-1}(x) = \Phi^j(x) \cdot P^j \qquad \Psi^{j-1}(x) = \Phi^j(x) \cdot Q^j$$

- two-scale relation for scaling functions and wavelets:

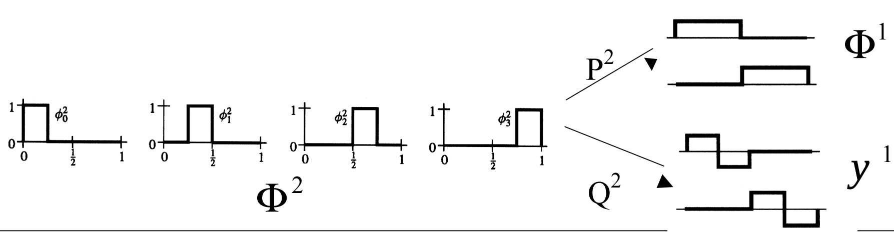
$$\left[\Phi^{j-1} \mid \Psi^{j-1}\right] = \Phi^j \left[P^j \mid Q^j\right]$$

2. Foundations of Wavelet Theory - 2.2 Multiresolution Analysis

# Two-Scale Relation for Haar Basis

$$\left[\Phi^1 \mid \Psi^1\right] = \Phi^2 \left[P^2 \mid Q^2\right]$$

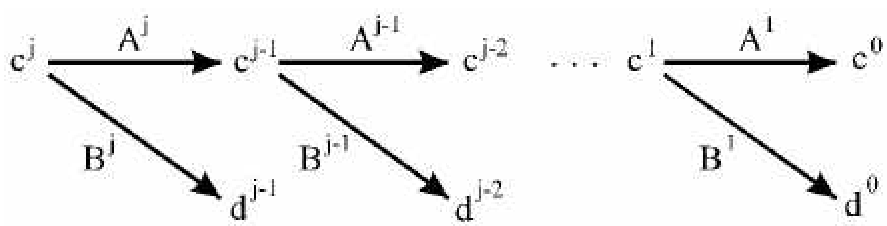
$$\Rightarrow \quad \left[\phi_0^1 \phi_1^1 \psi_0^1 \psi_1^1\right] = \left[\phi_0^2 \phi_1^2 \phi_2^2 \phi_3^2\right] \cdot \begin{bmatrix} 1 & 0 & 1 & 0 \\ 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & -1 \end{bmatrix}$$




2. Foundations of Wavelet Theory - 2.2 Multiresolution Analysis

# 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 $c^{j-1}$






# Analysis and Synthesis Filters

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

satisfying: $$\left[\Phi^{j-1} \mid \Psi^{j-1}\right] \cdot \left[\frac{A^j}{B^j}\right] = \Phi^j$$

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

$$\left[\Phi^{j-1} \mid \Psi^{j-1}\right] = \Phi^j \left[P^j \mid Q^j\right] \quad \Rightarrow \quad \left[\frac{A^j}{B^j}\right] = \left[P^j \mid Q^j\right]^{-1}$$

## Analysis Filter for Haar Basis

Analysis Filter: $A^2 = \frac{1}{2} \cdot \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}$

$$B^2 = \frac{1}{2} \cdot \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix}$$

<u>*Example:*</u>

$$[9 \quad 7 \quad 3 \quad 5]^T \xrightarrow{A^2} \begin{bmatrix} 6 \\ 4 \end{bmatrix}$$

$$[9 \quad 7 \quad 3 \quad 5]^T \xrightarrow{B^2} \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$



## 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

<u>Definitions:</u>

- wavelets $y$ are semi-orthogonal if

$$\forall j,k,l : \left\langle f_k^j \mid y_l^j \right\rangle = 0$$

- wavelets $y$ are bi-orthogonal if (~ indicates dual basis)

$$\forall i,j,l : \left\langle f_k^j \mid \tilde{y}_l^j \right\rangle = 0 \quad \wedge \quad \left\langle y_k^j \mid \tilde{f}_l^j \right\rangle = 0$$

## 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



## Overview

# 3. Standard Applications

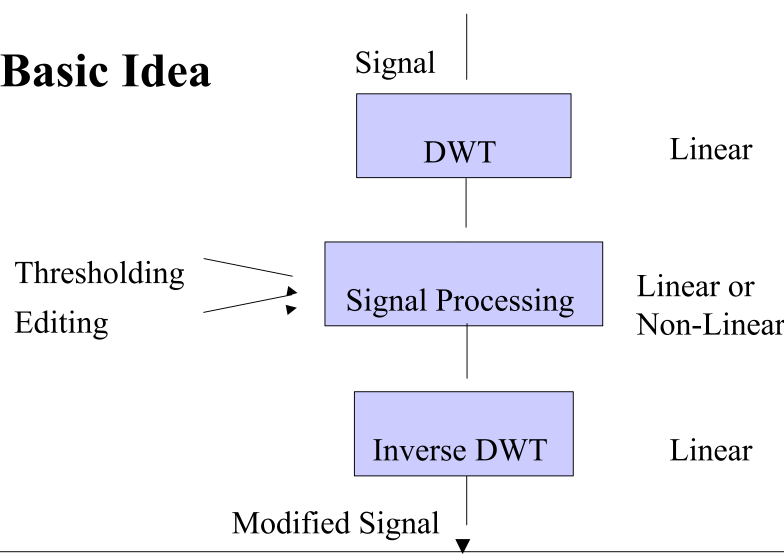
## 3.1 Signal Processing

- Signal Filtering (Linear / Non-Linear)
  - Noise-Reduction (De-Noising)
  - Speckle-Reduction (De-Speckling)
- Edge-Detection
- Multi-Resolution Editing

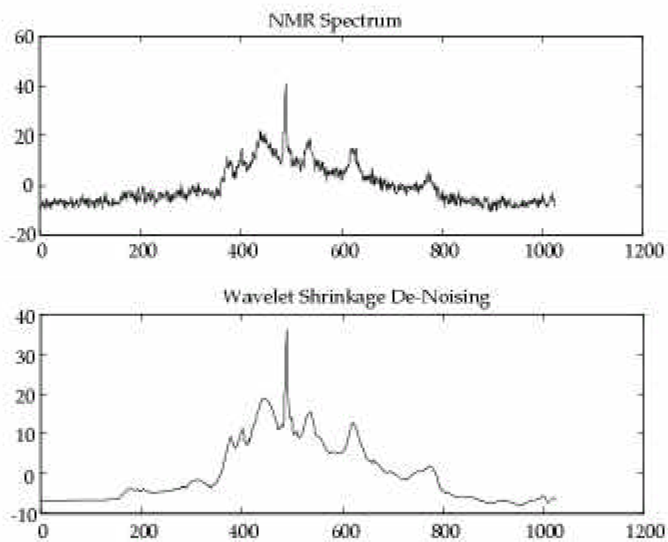


# 3.1 Signal Processing

## Basic Idea

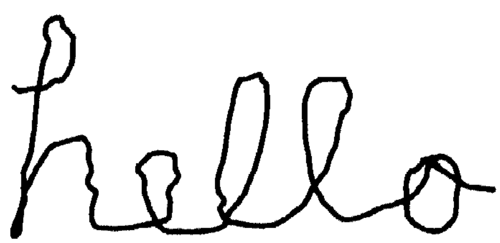


## Noise-Reduction

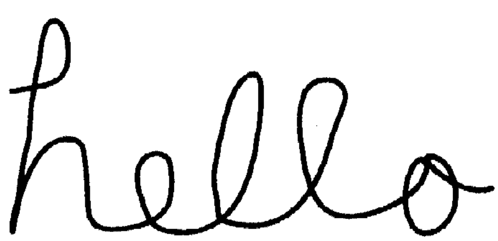





## Noise-Reduction

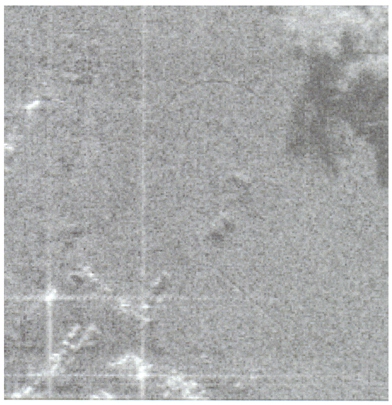


before
noise-reduction

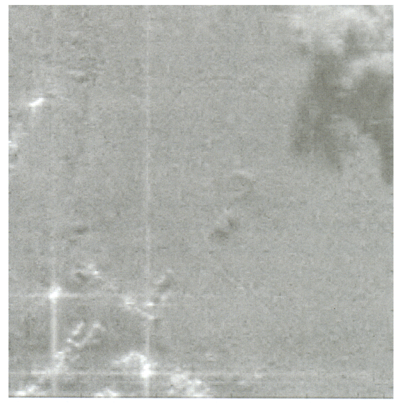


after
noise-reduction

## Speckle-Reduction



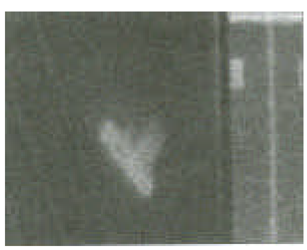
before reduction



after reduction



## Edge-Detection

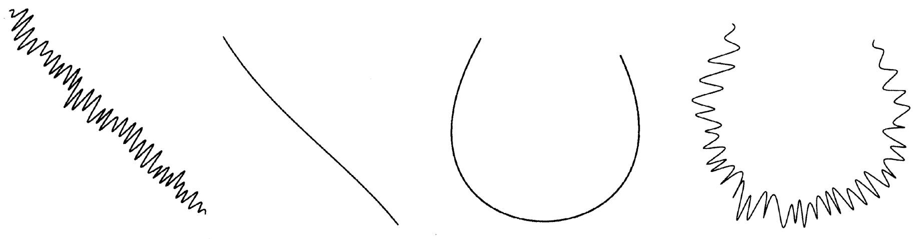


without wavelet transformation

with wavelet transformation

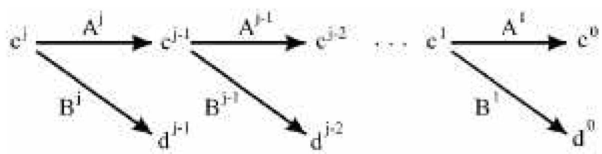
3. Standard Applications - 3.1 Signal Processing

## Multiresolution Editing



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

Principle Idea






## 3.2 Data Compression

**Basic Idea**

Data

| Step | |
|---|---|
| Transformation | Lossy or Loss-less |
| Quantization | Lossy |
| Run-length Coding | Loss-less |
| Entropy Coding | Loss-less |

Compressed Data

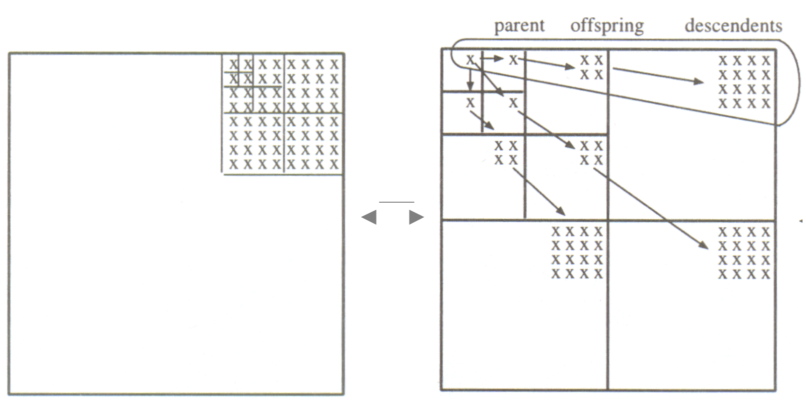
# Data Compression

- 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



# Data Compression

- Data Storage



# Data Compression

- Data Storage



# **Example:**
## Wavelet Image Compression Algorithm

**procedure** *Compress*($\boldsymbol{c}$: **array** $[1\,..\,m]$ **of reals**; $\varepsilon$: **real**)

$\tau_{\min} \leftarrow \min\{\,|\boldsymbol{c}[i]|\,\}$
$\tau_{\max} \leftarrow \max\{\,|\boldsymbol{c}[i]|\,\}$
**do**
  $\tau \leftarrow (\tau_{\min} + \tau_{\max})/2$
  $s \leftarrow 0$
  **for** $i \leftarrow 1$ **to** $m$ **do**
    **if** $|\boldsymbol{c}[i]| < \tau$ **then** $s \leftarrow s + |\boldsymbol{c}[i]|^2$
  **end for**
  **if** $s < \varepsilon^2$ **then** $\tau_{\min} \leftarrow \tau$ **else** $\tau_{\max} \leftarrow \tau$
**until** $\tau_{\min} \approx \tau_{\max}$
**for** $i \leftarrow 1$ **to** $m$ **do**
  **if** $|\boldsymbol{c}[i]| < \tau$ **then** $\boldsymbol{c}[i] \leftarrow 0$
**end for**
**end procedure**

*c*[i]: coefficients of Wavelet Transformation

t : threshold with error e

## Wavelet Image Compression

Original

1:50

1:100

1:200



## Wavelet Image Compression

Original

1:6 Compression

## Wavelet Image Compression

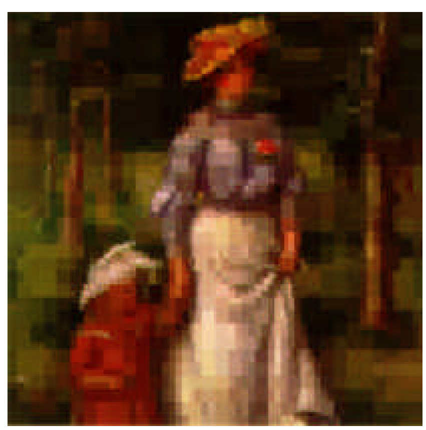
Original

1:42 Compression



## Wavelet Image Compression

Original



1:222 Compression

3. Standard Applications - 3.2 Data Compression

## Example:
## FBI Fingerprint Compression

Original | 1:26 Compression



## Wavelet Image Compression

Wavelet Compression | Fourier 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!)

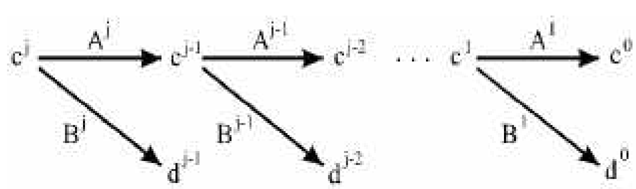
Example



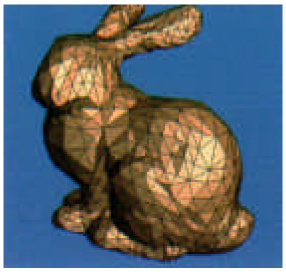
# 3.3 Computer Graphics

- Approximation of Curves and Surfaces
- Multiresolution Data Representation
- Multiresolution Manipulation of Images, Curves, and Surfaces

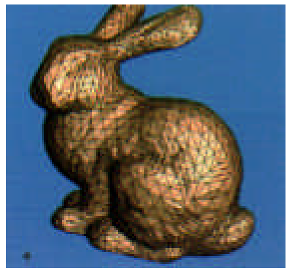
Principle Idea



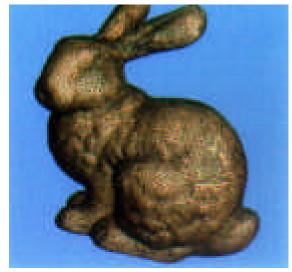
## Approximation of Surfaces



952 Wavelet Coefficients



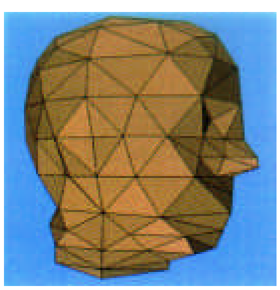
2268 Wavelet Coefficients



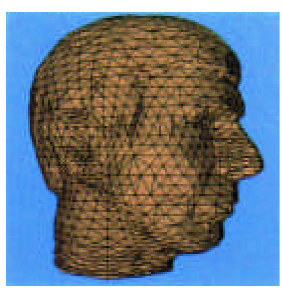
18636 Wavelet Coefficients



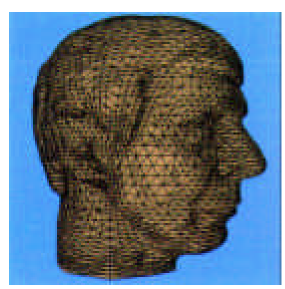
## Multiresolution Data Representation



952 Wavelet Coefficients



2268 Wavelet Coefficients

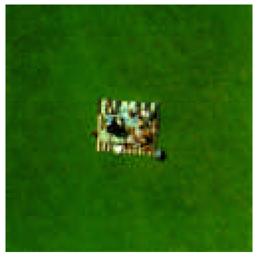


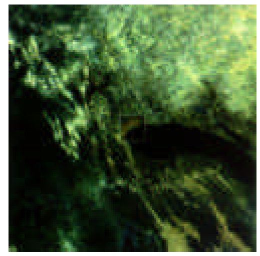
18636 Wavelet Coefficients

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

## Multiresolution Image Manipulation

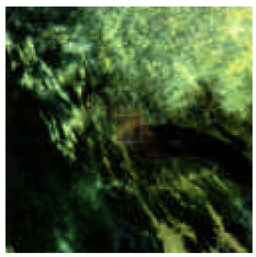
Original

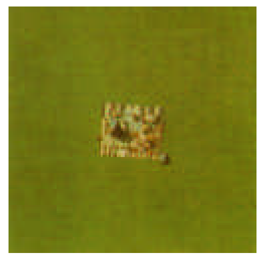




Zoom out by a Factor of 100,000

Adding Smog





Changed Original



## Multiresolution Curve Manipulation

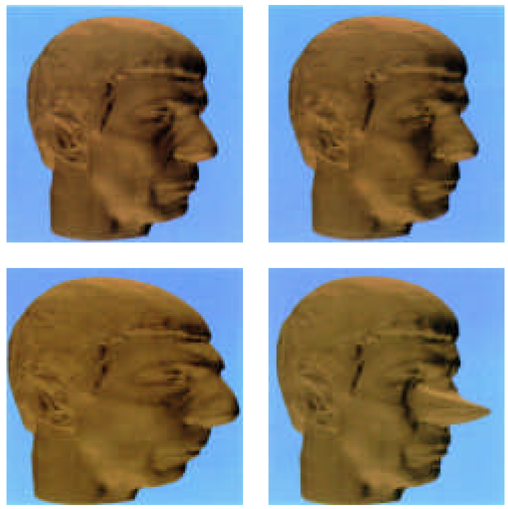
Original Spline Surface

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]

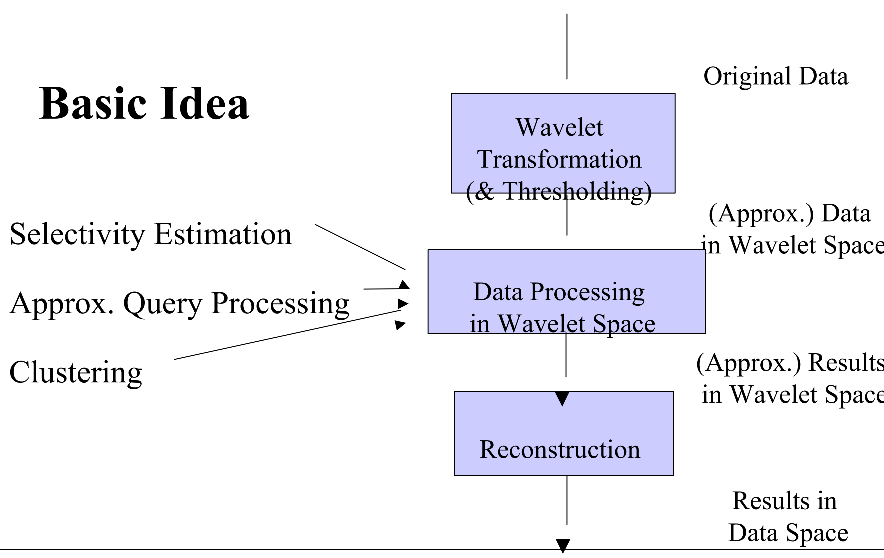
– Approximate Query Processing
  - Data Cubes [VWI 98, VW99]
  - Relational Databases [CGRS00]

– Clustering Techniques
  - WaveCluster [SCZ98, SCZ00]



# 4.1 Efficient Data Processing

## Basic Idea



# Selectivity Estimation:
## Wavelet-based Histograms [MVW98, MVW00]

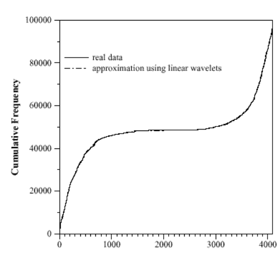
– **Goal**
  - Compact histograms
  - Accurate estimation of the data distribution

– **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

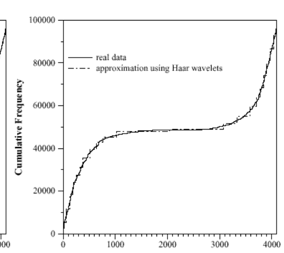


# Selectivity Estimation:
## Wavelet-based Histograms [MVW98, MVW00]

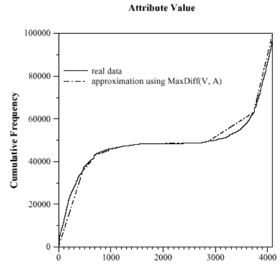
linear wavelet



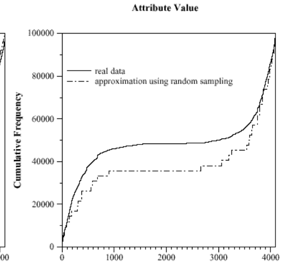

Haar wavelet




MaxDiff




random sampling

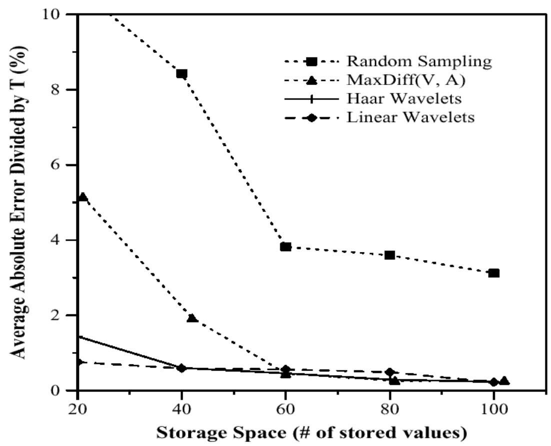



4. Applications in Databases - 4.1 Efficient Data Processing

# 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]

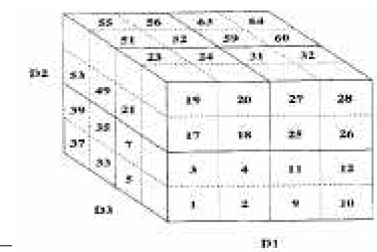
**Ideas**

– partial sum cube

<u>Motivation:</u> 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]

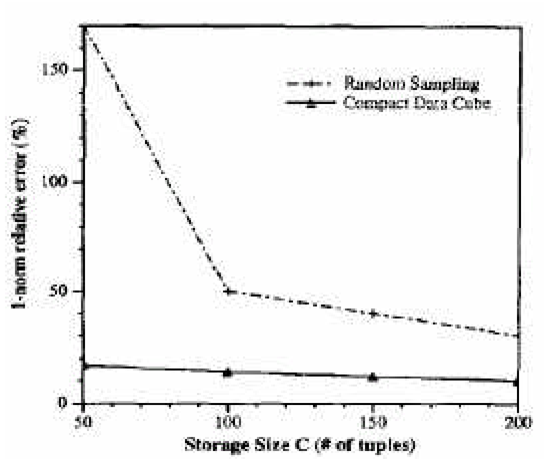
– **Haar Wavelet Decomposition**
  - based directly on the sparse repres. of the original cube
  - result: C´ coefficients (~ non-zero entries of the cube)

– **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)





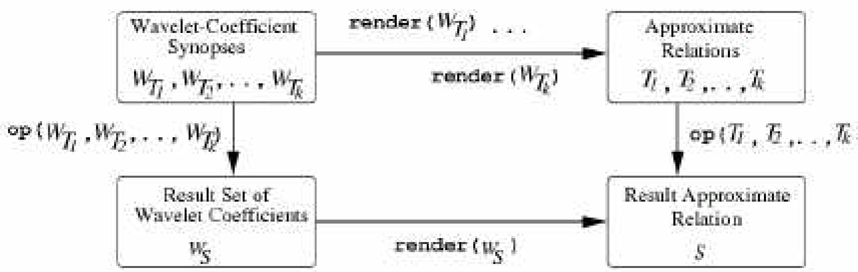

## Approximate Query Processing:
### Relational Databases [CGRS00]

- **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

# Approximate Query Processing

## Relational Databases [CGRS00]

Schema of Wavelet-based Approximate Query Processing



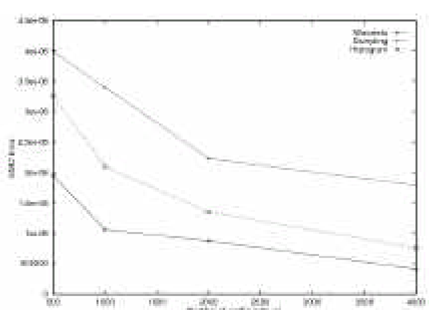



# Approximate Query Processing

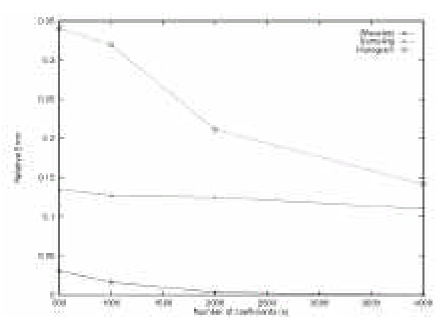
## Relational Databases [CGRS00]

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

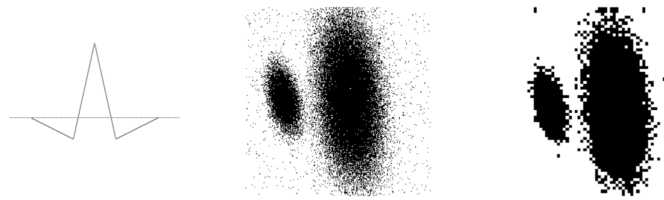
Algorithm:

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

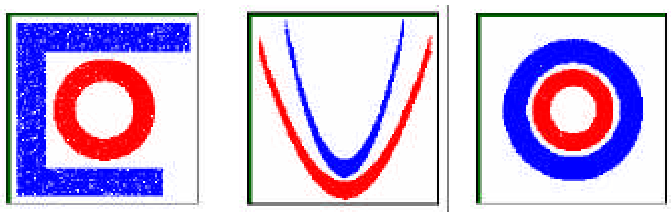
1. Quantize feature space, then assign objects to the units.
2. Apply wavelet transform on the feature space.
3. Find the connected components (clusters) in the subbands of transformed feature space, at different levels.
4. Assign label to the units.
5. Make the lookup table.
6. Map the objects to the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

## Clustering Techniques:
### WaveCluster [SCZ98, SCZ00]

- Effect of Wavelet Transformation



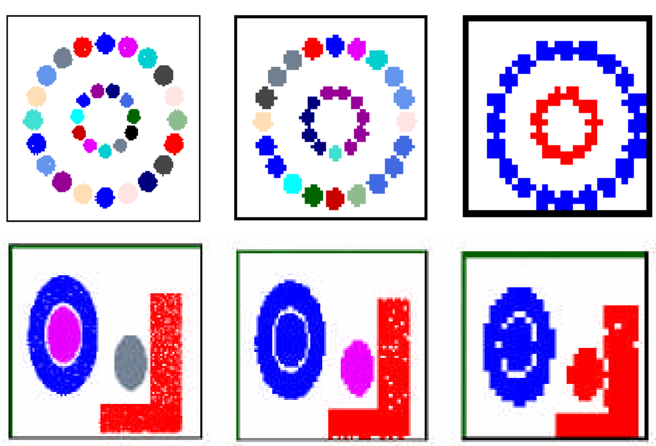
- Arbitrary shaped clusters found by WaveCluster





## Clustering Techniques:
### WaveCluster [SCZ98, SCZ00]

Results From Clustering on different Resolution Levels



## 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

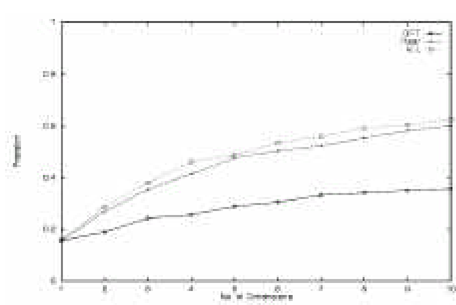
# Similarity Search in Time Series Databases [CF99]

- **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

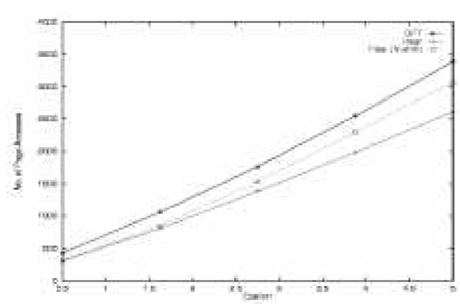


# Similarity Search in Time Series Databases [CF99]

## Results



Accuracy



Page Accesses

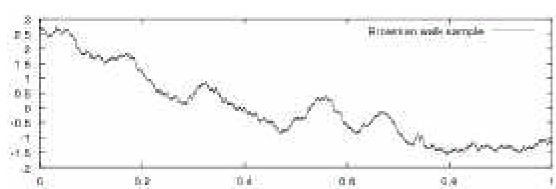
## Similarity Search in Time Series Databases [SS99a,b]

- **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

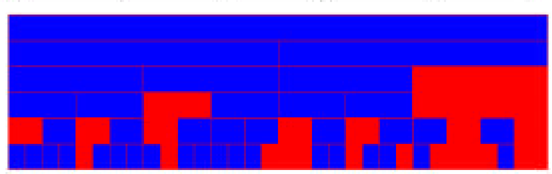


## Similarity Search in Time Series Databases [CF99]

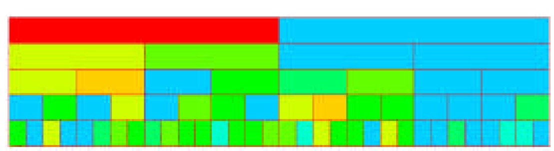
Time Series Data



Sign Information
(6 levels)



Hölder Decomposition
(5-levels)



# Similarity Search in Image Databases

## Content-based Similarity Search

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



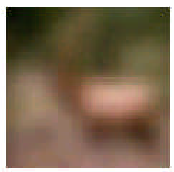
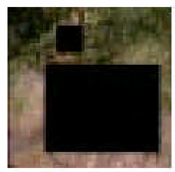
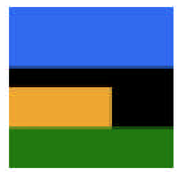
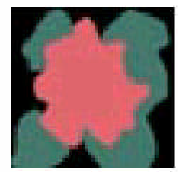
# Similarity Search in Image Databases

- **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

## Similarity Search in Image Databases
## WBIIS [WWFW97a,b]

- **Goal**
  - Improved Content-based Retrieval:

    Partial Image Retrieval & Sketch Retrieval

| | | | |
|---|---|---|---|
|  |  |  |  |
| low resolution | partial image | block sketch | hand-drawn sketch |



## Similarity Search in Image Databases
## WBIIS [WWFW97a,b]

**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



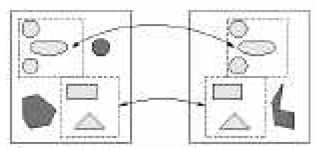
## Similarity Search in Image Databases
## WBIIS [WWFW 97a,b]

- 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



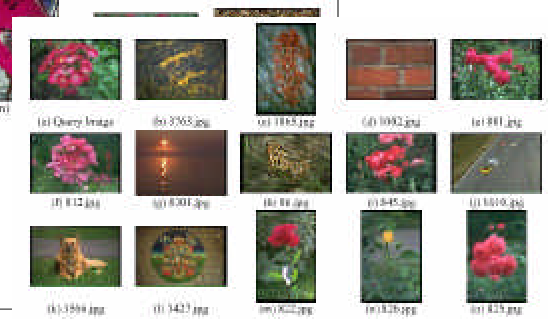
- **Approach**
  - Haar Wavelet Transformation of sliding window of varying size
  - Clustering of Signatures in Wavelet Space (BIRCH)

    => variable Number of Signatures per Image
  - Storage of Centroids of Clusters in Index Structure (R*-Tree)
  - Similarity Search: Matching Pairs of Signatures (largest overlap)



# Similarity Search in Image Databases
## WALRUS [NRS99]

WALRUS Results

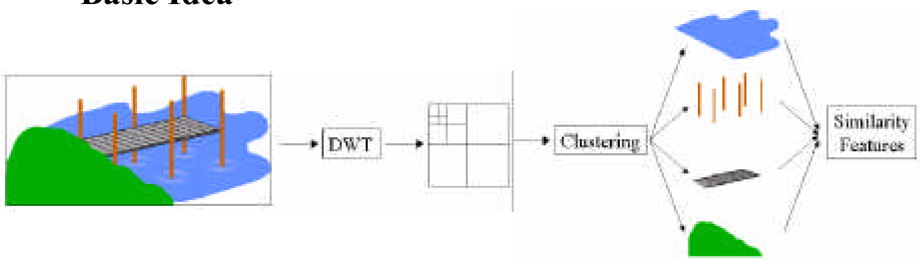


WBIIS Results

## Similarity Search in Image Databases
## Windsurf [ABP99]

- **Goal**
  - Content-based Similarity Search in Large Image Databases: Partial Similarity based on Image Regions

- **Basic Idea**

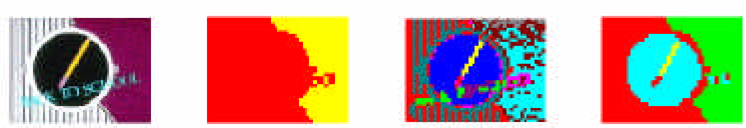





## Similarity Search in Image Databases
## Windsurf [ABP99]

**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)

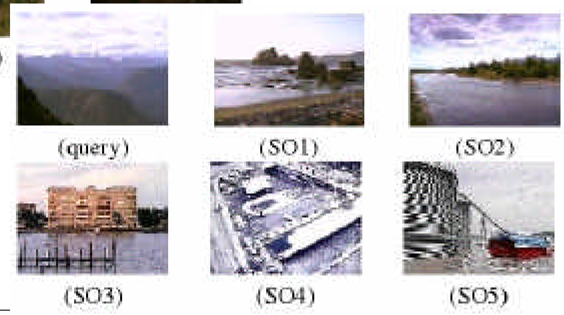


Original k-means k=2 k=10 k=4

- Feature Vectors correspond to Regions found in Clustering Step:
  (Size, Centroids, Covariance Matrix of Pixels in Region)
- Similarity Retrieval based on Matching Regions

## Similarity Search in Image Databases Windsurf [ABP99]

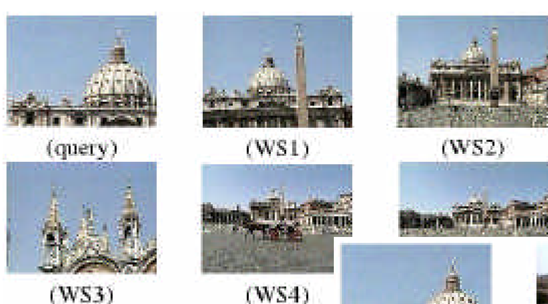

Windsurf Results




Results of comparative System

4. Applications in Databases - 4.2 Similarity Search

## Similarity Search in Image Databases Windsurf [ABP99]




Windsurf Results


Results of comparative System

4. Applications in Databases - 4.2 Similarity Search

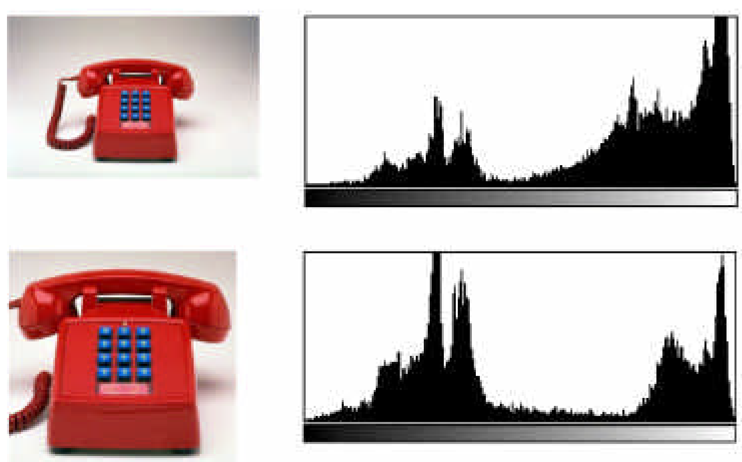
## Similarity Search in Image Databases
## Multiresolution Search [Hec99]

- **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



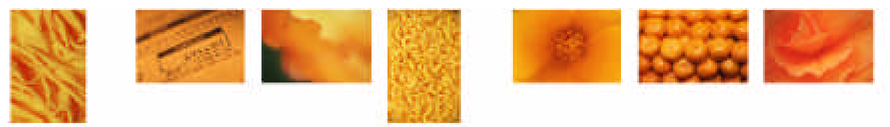
## Similarity Search in Image Databases
## Multiresolution Search [Hec99]

Color Histogram Example



# Similarity Search in Image Databases
## Multiresolution Search [Hec99]

Search on different level Wavelet Coefficients



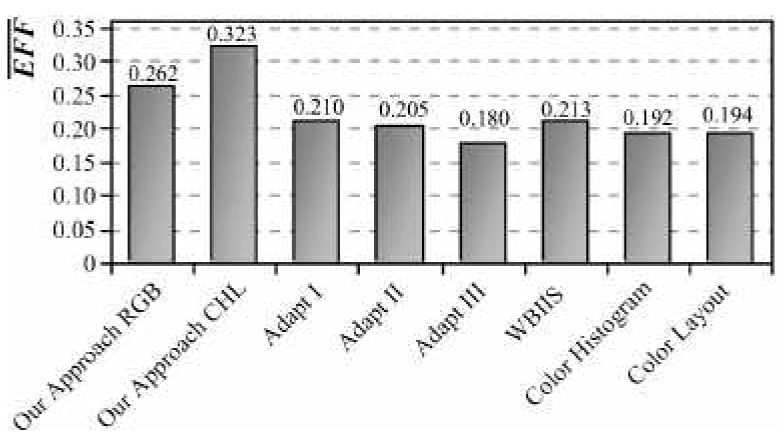
Query Image Results of Search on Low-Frequency Coefficients

Query Image Results of Search on High-Frequency Coefficients



# Similarity Search in Image Databases
## Multiresolution Search [Hec99]

Comparison to other Approaches: Effectiveness Results




4. Applications in Databases - 4.2 Similarity Search

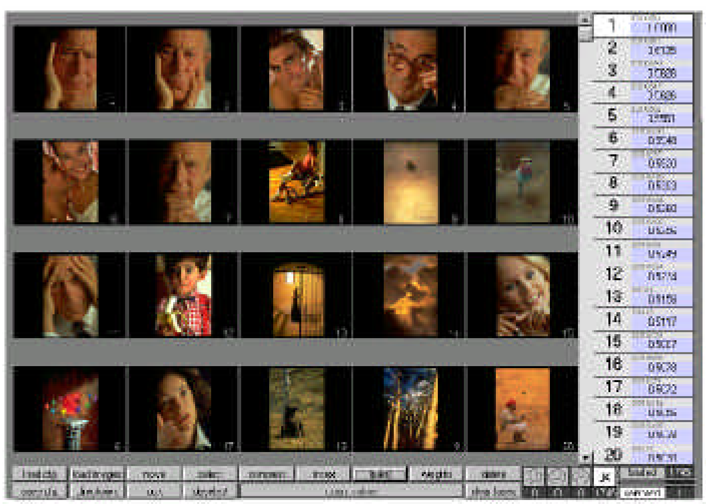
## Similarity Search in Image Databases
## Similarity Measure Learning [BVGS99]

- **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



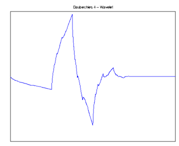
## Similarity Search in Image Databases
## Similarity Measure Learning [BVGS99]

Example Results



## Similarity Search in Image Databases
## WIPE [WWF97]

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

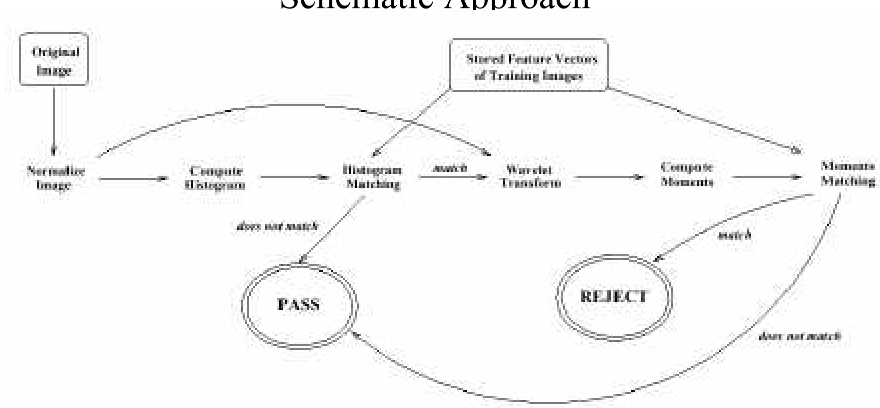


- **Approach**
  - Normalization of Images to Standard Size
  - Wavelet Transformation using Daubechies-3 Wavelets
  - Edge Detection in Different Subbands of Wavelet Transformation
  - Feature Vectors used for Similarity Matching:
    Central Moments & Invariant Moments & Color Histograms
  - Filtering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



## Similarity Search in Image Databases
## WIPE [WWF97]

### Schematic Approach




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

- 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

# 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

5. Summary and Conclusion

# Bibliography

[ABP99] Stefania Ardizzoni, Ilaria Bartolini, and Marco Patella. Windsurf: Region-based image retrieval using wavelets. In Proceedings of 10th International Workshop on Database & Expert Systems Applications, DEXA 1999, pages 167-173. IEEE Computer Society, 1999.

[BVGS99] C. Brambilla, A. Della Ventura, I. Gagliardi, and R. Schettini. Multiresolution wavelet transform and supervised learning for content-based image retrieval. In Proceedings of 1999 IEEE International Conference on Multimedia Computing and Systems, ICMCS, volume 1. IEEE Computer Society, 1999.

[CF99] Kin-Pong Chan and Ada Wai-Chee Fu. Ecient time series matching by wavelets. In Proceedings of the 15th International Conference on Data Engineering, pages 126-133. IEEE Computer Society, 1999.

[CGRS00] Kaushik Chakrabarti, Minos N. Garofalakis, Rajeev Rastogi, and Kyuseok Shim. Approximate query processing using wavelets. In VLDB'2000, Proceedings of 26nd International Conference on Very Large Data Bases. Morgan Kaufmann, 2000.

[FBF+94] C. Faloutsos, R. Barber, M. Flickner, J. Hafner, W. Niblack, D.Petkovic, and W. Equitz. Ecient and eective querying by image content. Journal of Intelligent Information Systems, 3:231-262, 1994.

[GG97] Ramesh A. Gopinath and Haitao Guo. Introduction to Wavelets and Wavelet Transforms: A Primer. Prentice Hall, 1997.

[GZT99] C. Garcia, G. Zikos, and G. Tziritas. Face detection in color images using wavelet packet analysis. In Proceedings of 1999 IEEE International Conference on Multimedia Computing and Systems, ICMCS, volume 1. IEEE Computer Society, 1999.

Bibliography

# Bibliography


[Haa10] Alfred Haar. Zur Theorie der orthogonalen Funktionen-Systeme. Mathematische Analen, 69:331-371, 1910.

[Hec99] Martin Heczko. Waveletbasierte Ahnlichkeitssuche in Multimedia-Datenbanken. Master's thesis, University Halle-Wittenberg, 1999.

[HJS94] Michael L. Hilton, Bjorn D. Jawerth, and Ayan Sengupta. Compressing still and moving images with wavelets. Multimedia Systems, 2(5):218-227, December 1994.

[LW99] Ian K. Levy and Roland Wilson. Three dimensional wavelet transform video compression. In Proceedings of 1999 IEEE International Conference on Multimedia Computing and Systems, ICMCS, volume 2, pages 924{928. IEEE Computer Society, 1999.

[Mal89] Stephane Mallat. A theory for multiresolution signal decomposition: The wavelet representation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 11(7):674-693, July 1989.

[Mal99] Stephane G. Mallat. A Wavelet Tour of Signal Processing. Academic Press, 1999.

[Mey93] Yves Meyer. Wavelets: Algorithms and Applications. SIAM, Philadelphia, 1993.

[MVW98] Yossi Matias, Jerey Scott Vitter, and Min Wang. Wavelet-based histograms for selectivity estimation. In Laura M. Haas and Ashutosh Tiwary, editors, SIGMOD 1998, Proceedings ACM SIGMOD International Conference on Management of Data, pages 448-459. ACM Press, 1998.

[MVW00] Yossi Matias, Je Scott Vitter, and Min Wang. Dynamic maintenance of wavelet-based histograms. In VLDB'2000, Proceedings of 26nd International Conference on Very Large Data Bases. Morgan Kaufmann, 2000.




# Bibliography


[NRS99] Apostol Natsev, Rajeev Rastogi, and Kyuseok Shim. WALRUS: A similarity retrieval algorithm for image databases. In Alex Delis, Christos Faloutsos, and Shahram Ghandeharizadeh, editors, SIGMOD 1999, Proceedings ACM SIGMOD International Conference on Management of Data, pages 395-406. ACM Press, 1999.

[RB98] Raghuveer M. Rao and Ajit S. Bopardikar. Wavelet Transforms: Introduction to Theory and Applications. Addison-Wesley, 1998.

[RBC+92] Mary Beth Ruskai, Gregory Beylkin, Ronald Coifman, Ingrid Daubechies, Stephane Mallat, Yves Meyer, and Louise Raphael. Wavelets and Their Applications. Jones and Bartlett Publishers, Boston, MA, 1992.

[SCZ98] Gholamhosein Sheikholeslami, Surojit Chatterjee, and Aidong Zhang. Wavecluster: A multi-resolution clustering approach for very large spatial databases. In Ashish Gupta, Oded Shmueli, and Jennifer Widom, editors, VLDB'98, Proceedings of 24rd International Conference on Very Large Data Bases, pages 428-439. Morgan Kaufmann, 1998.

[SCZ00] Gholamhosein Sheikholeslami, Surojit Chatterjee, and Aidong Zhang. Wavecluster: A wavelet based clustering approach for spatial data in very large databases. VLDB Journal, 8(3-4):289-304, 2000.

[SDS96] Eric J. Stollnitz, Tony D. DeRose, and David H. Salesin. Wavelets for Computer Graphics, Theory and Applications. Morgan Kaufman Publishers, Inc., San Francisco, California, 1996.

[Sha92] J. M. Shapiro. An embedded wavelet hierarchical image coder. In Proceedings of 1992 IEEE International Conference on Acoustics, Speech, and Signal Processing, volume 4, pages 657-660. IEEE Computer Society, 1992.

# Bibliography

[SS99a] Zbigniew R. Struzik and Arno Siebes. The haar wavelet transform in the time series similarity paradigm. In Jan M. Zytkow and Jan Rauch, editors, Lecture Notes in Computer Science: Proceedings of the PKDD '99, Principles of Data Mining and Knowledge Discovery, volume 1704, pages 12-22. Springer, 1999.

[SS99b] Zbigniew R. Struzik and Arno Siebes. Measuring time series' similarity through large singular features revealed with wavelet transformation. In Proceedings of 10th International Workshop on Database & Expert Systems Applications, DEXA 1999, pages 162-166. IEEE Computer Society, 1999.

[SZ97] G. Sheikholeslami and A. Zhang. An approach to clustering large visual databases using wavelet transform. In Proceedings of the SPIE Conference on Visual Data Exploration and Analysis IV, 1997.

[Top98] Pankaj N. Topiwala. Wavelet Image and Video Compression. Kluwer Academic Publisher, 1998.

[VW99] Jerey Scott Vitter and Min Wang. Approximate computation of multidimensional aggregates of sparse data using wavelets. In Alex Delis, Christos Faloutsos, and Shahram Ghandeharizadeh, editors, SIGMOD 1999, Proceedings ACM SIGMOD International Conference on Management of Data, pages 193-204. ACM Press, 1999.

[VWI98] Jerey Scott Vitter, Min Wang, and Balakrishna R. Iyer. Data cube approximation and histograms via wavelets. In Georges Gardarin, James C. French, Niki Pissinou, Kia Makki, and Luc Bouganim, editors, Proceedings of the 1998 ACM CIKM, pages 96-104. ACM Press, 1998.

[Wei95] Karl Weierstraß. Mathematische Werke II. Mayer & Müller, Berlin, 1895.

Bibliography

# Bibliography

[WWF97] James Ze Wang, Gio Wiederhold, and Oscar Firschein. System for screening objectionable images using daubechies' wavelets and color histograms. In Ralf Steinmetz and Lars C. Wolf, editors, Lecture Notes in Computer Science, volume 1309, pages 20-30. Springer, 1997.

[WWFW97a] James Ze Wang, Gio Wiederhold, Oscar Firschein, and Sha Xin Wei. Content-based image indexing and searching using daubechies' wavelets. In International Journal on Digital Libraries, volume 1(4), pages 311-328, 1997.

[WWFW97b] James Ze Wang, Gio Wiederhold, Oscar Firschein, and Sha Xin Wei. Wavelet-based image indexing techniques with partial sketch retrieval capability. In Proceedings of the 4th Forum on Research and Technology Advances in Digital Libraries (ADL'97), pages 13-24, Washington D.C., May 1997.

Bibliography