Analysis of Image Tranforms for Sketch-based Retrieval

Diploma Thesis

Felix Stürmer

Technische Universität Berlin
Fakultät IV - Elektrotechnik und Informatik
Computer Graphics

02.11.2012

Outline

Motivation

foo

Challenges of CBIR

The Semantic Gap

*"The semantic gap is the **lack of coincidence** between the information that one can extract from the **visual data** and the **interpretation** that the same data have for a user in a given situation." – Smeulders et al.*

The Sensory Gap

*"The sensory gap is the gap between the **object in the world** and the information in a (computational) description derived from a **recording of that scene**." – Smeulders et al.*

Prior Work on Human Recognition

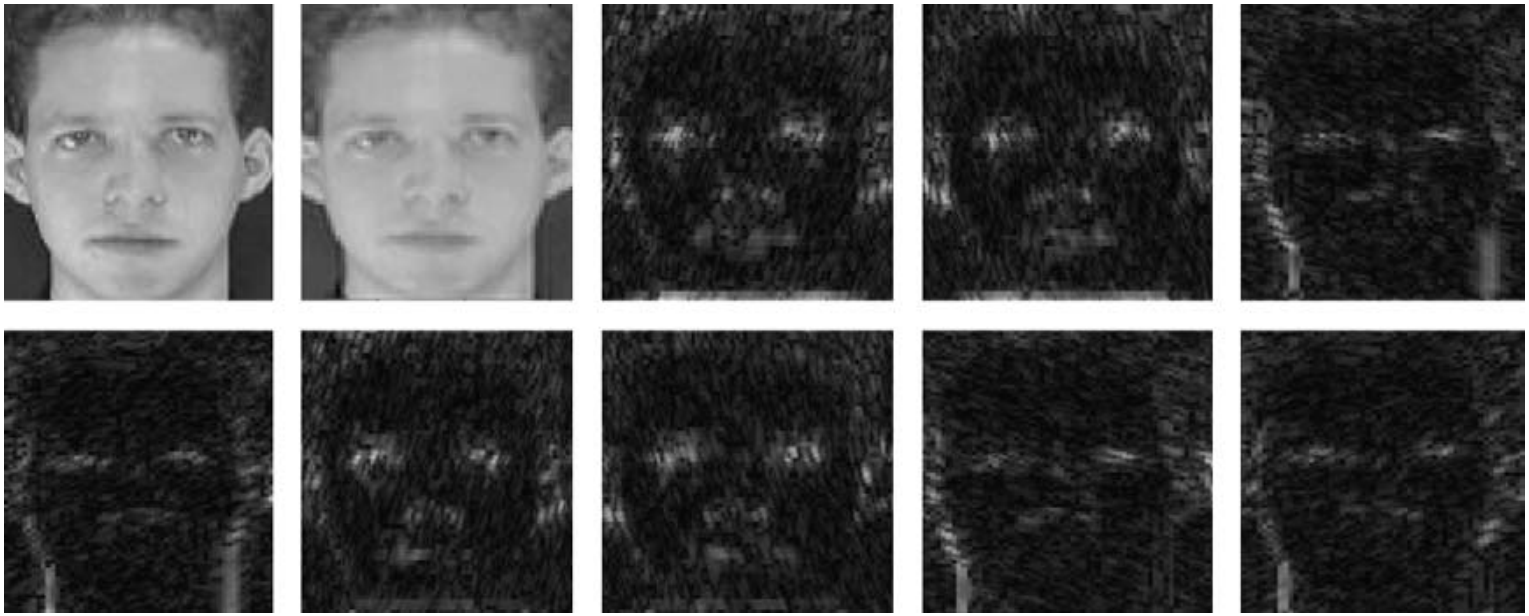


Figure: "Face recognition using curvelet based PCA.", T. Mandal and Q. M.J Wu, ICPR 2008

Prior Work on Human Recognition

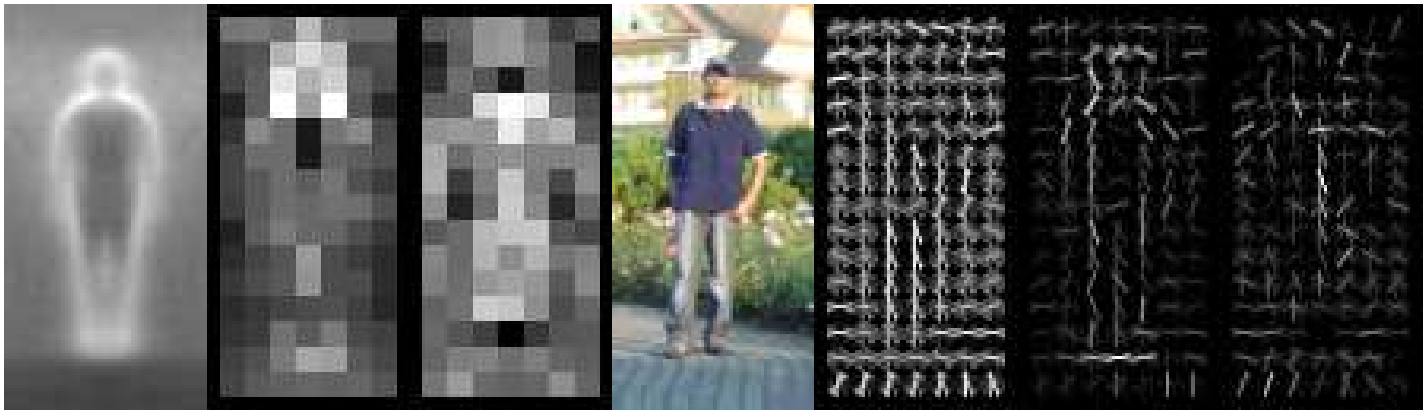
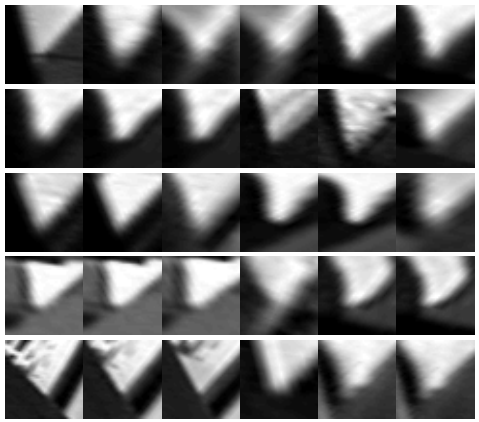


Figure: "Histograms of oriented gradients for human detection", Dalal and Triggs, CVPR 2005

Prior Work on Visual Codebooks



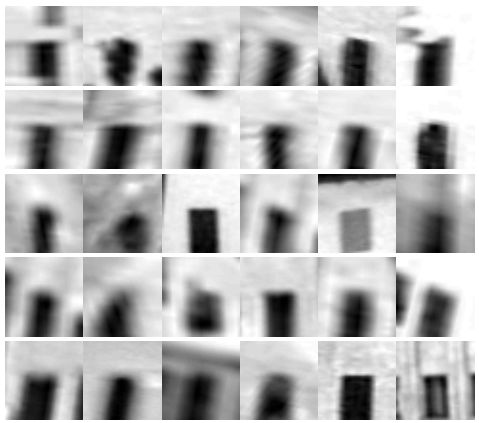


Figure: "Video Google: A text retrieval approach to object matching in videos", Sivic and Zisserman, ICCV 2003

Prior Work on Scene Classification

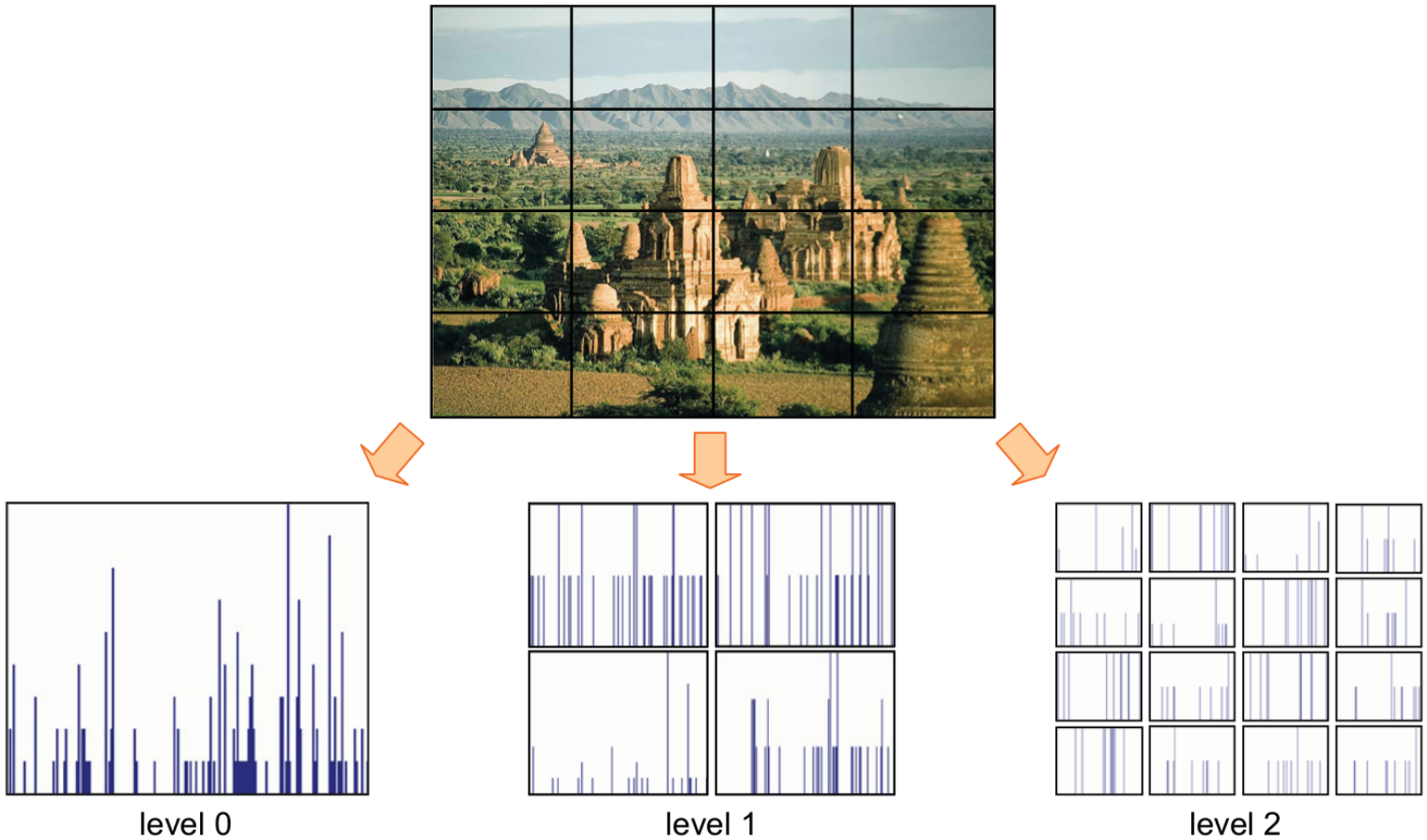


Figure: "Spatial pyramid matching", Lazebnik et al., 2009

Anatomy of a CBIR System

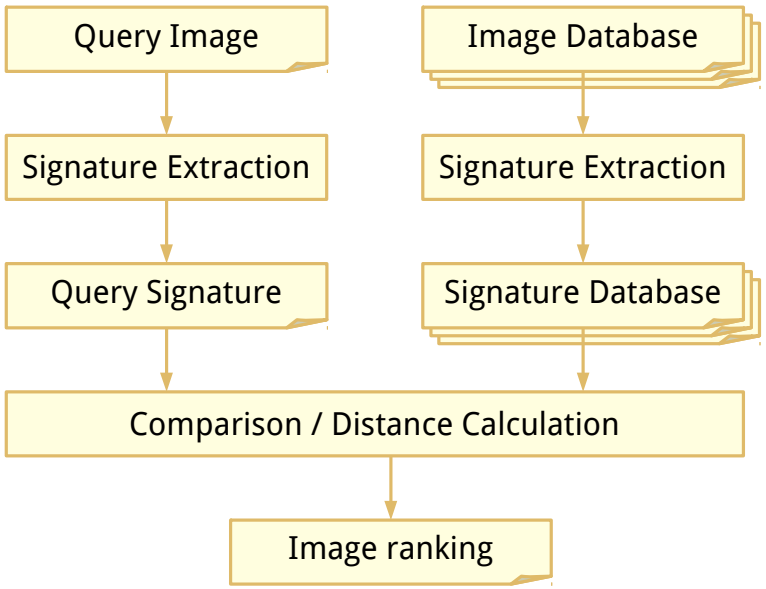


Figure: Global Descriptors

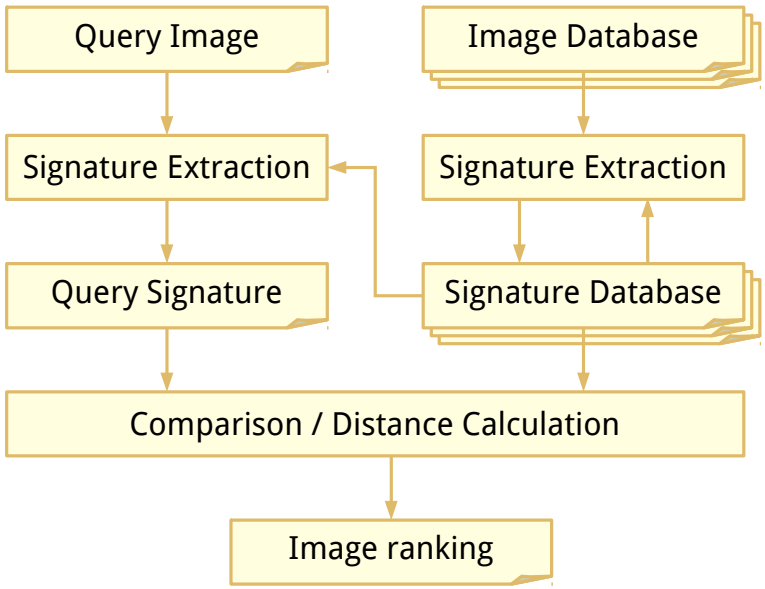


Figure: Local Descriptors

Proposed Retrieval Pipelines (Global)

Proposed Retrieval Pipelines (Local)

Acquisition

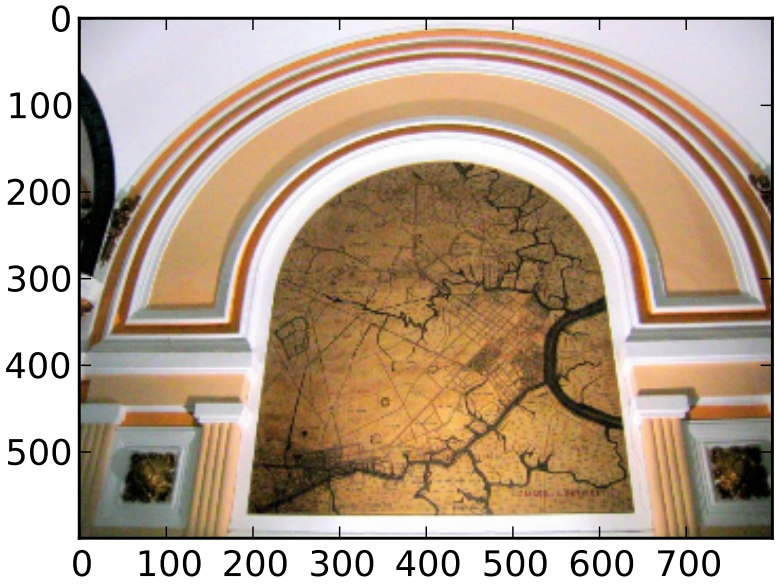


Figure: Original Image

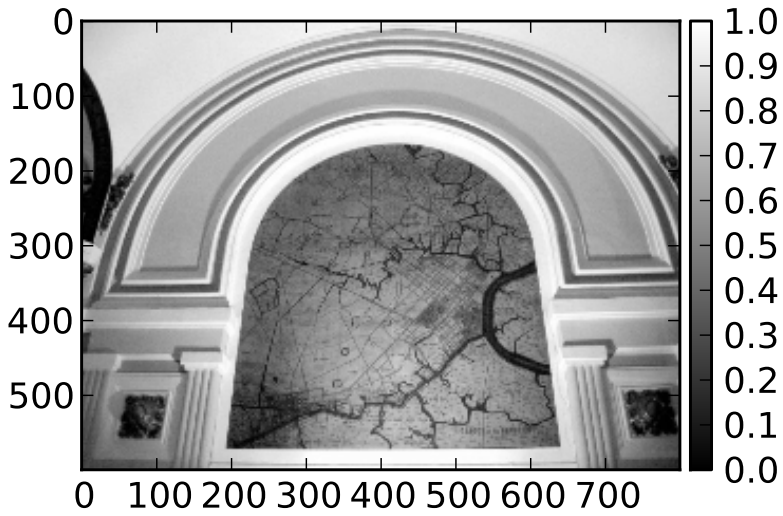


Figure: Luma Conversion

Acquisition

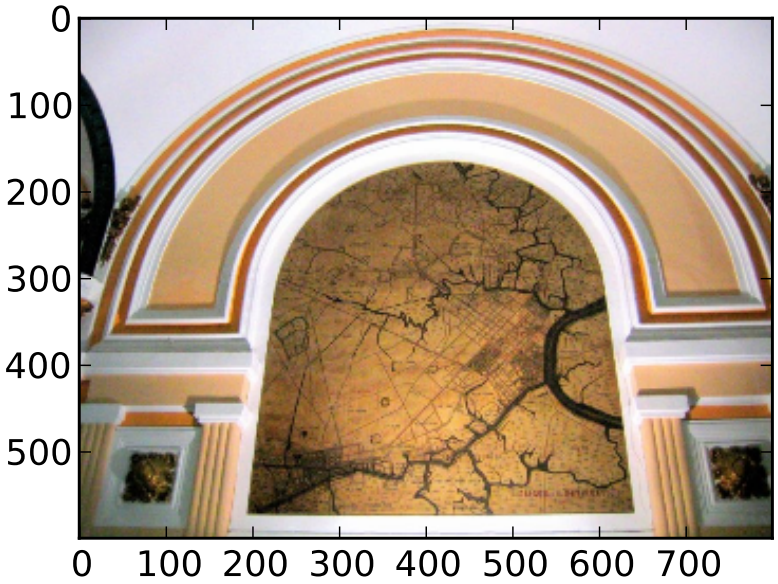


Figure: Original Image

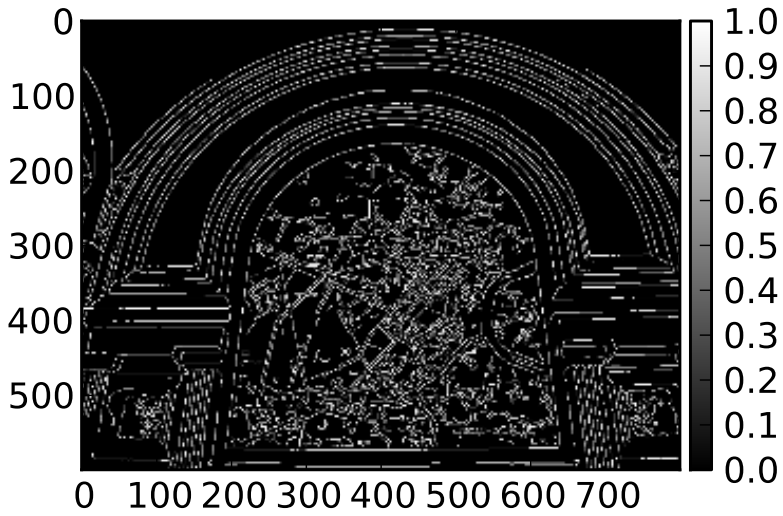


Figure: Canny Operator

Acquisition

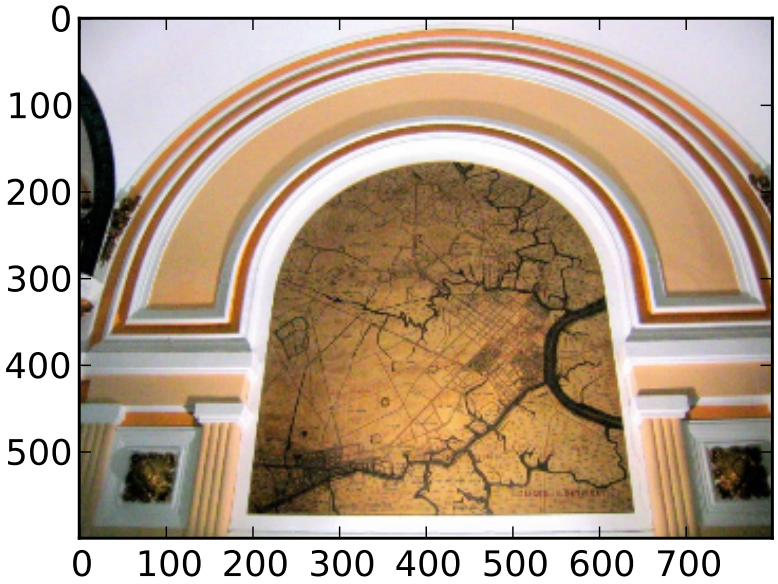


Figure: Original Image

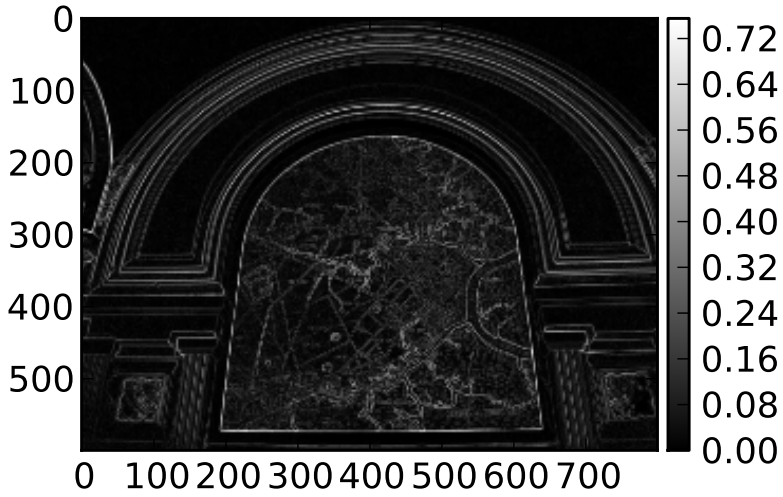


Figure: Sobel Operator

Acquisition

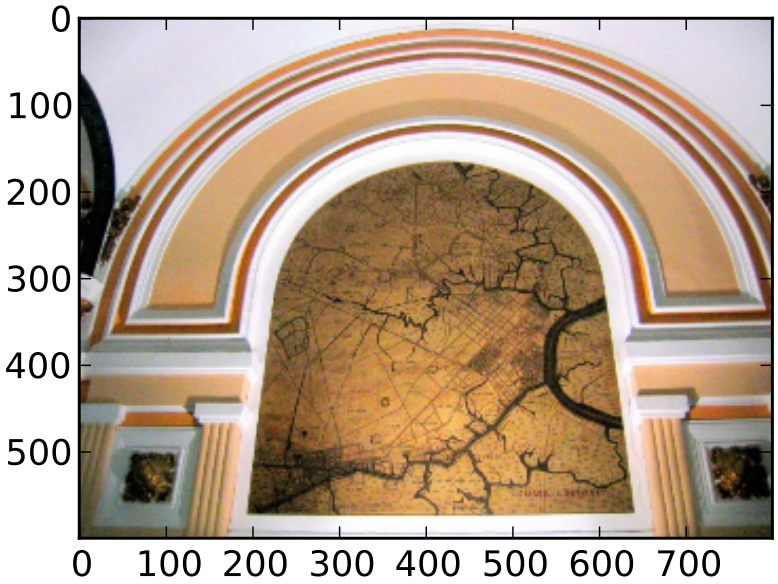


Figure: Original Image

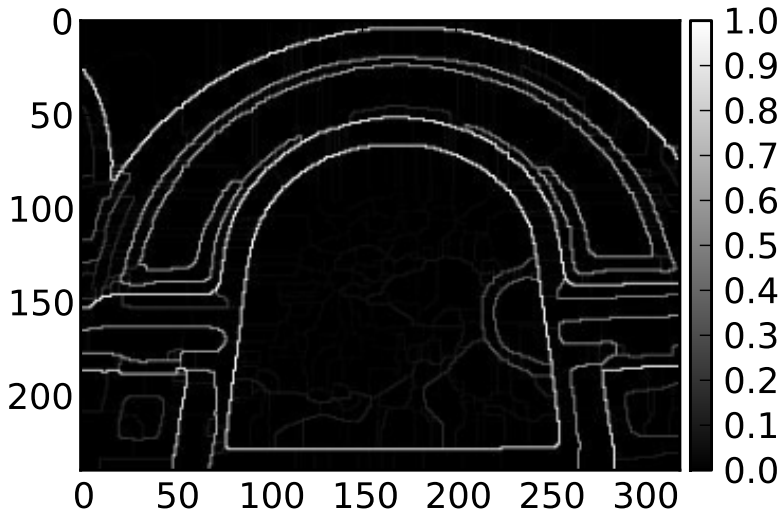


Figure: gPb-owt-ucm Transform

Properties of the Curvelet Transform

- An extension of the wavelet transform
- Localized in *position*, *scale* and *orientation*
- Curvelets obey parabolic scaling: $width \approx length^2$
- Approximation error along edges using m largest coefficients decays with $\frac{log(m)^3}{m^2}$ (compare $\frac{1}{m}$ for wavelets)
- Defined in frequency domain using

Constructing the Curvelets

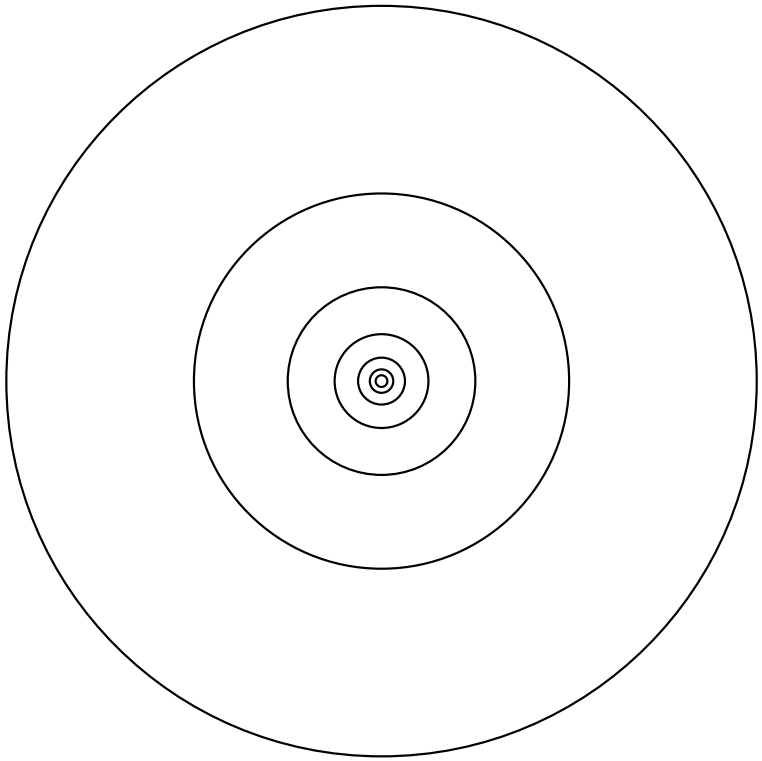


Figure: Frequency Domain

Figure: Spatial Domain

Constructing the Curvelets

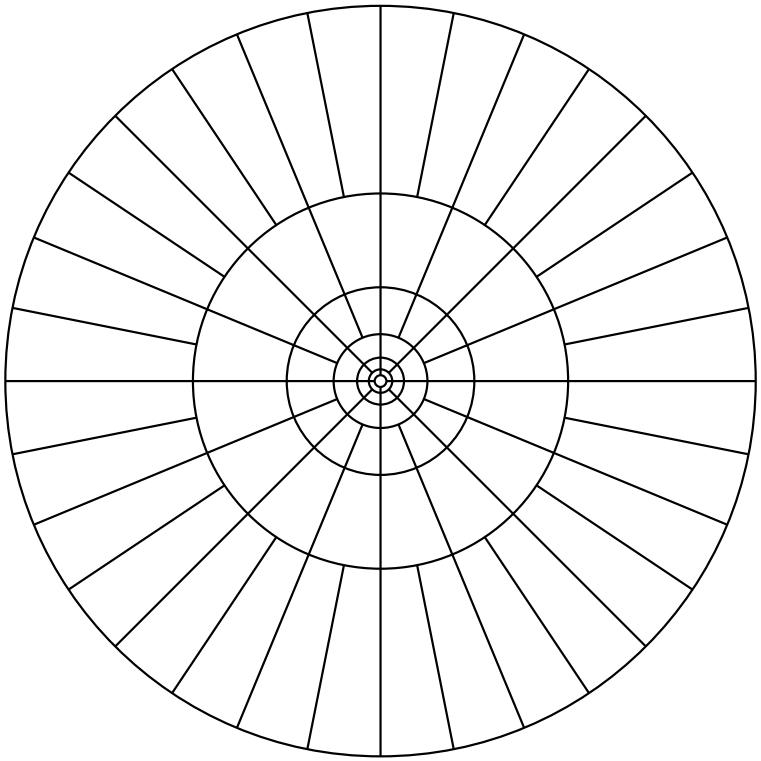


Figure: Frequency Domain

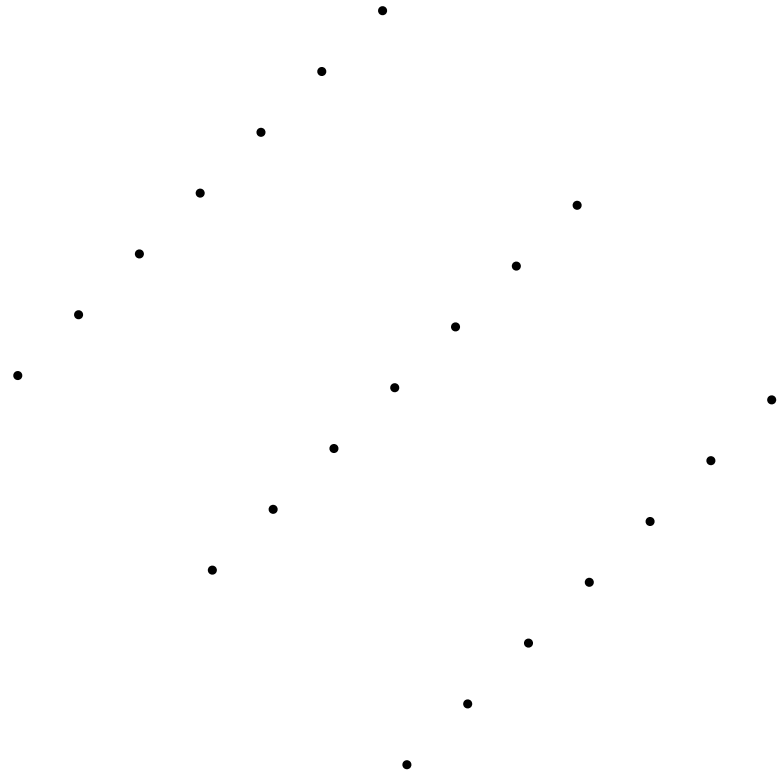


Figure: Spatial Domain

Constructing the Curvelets

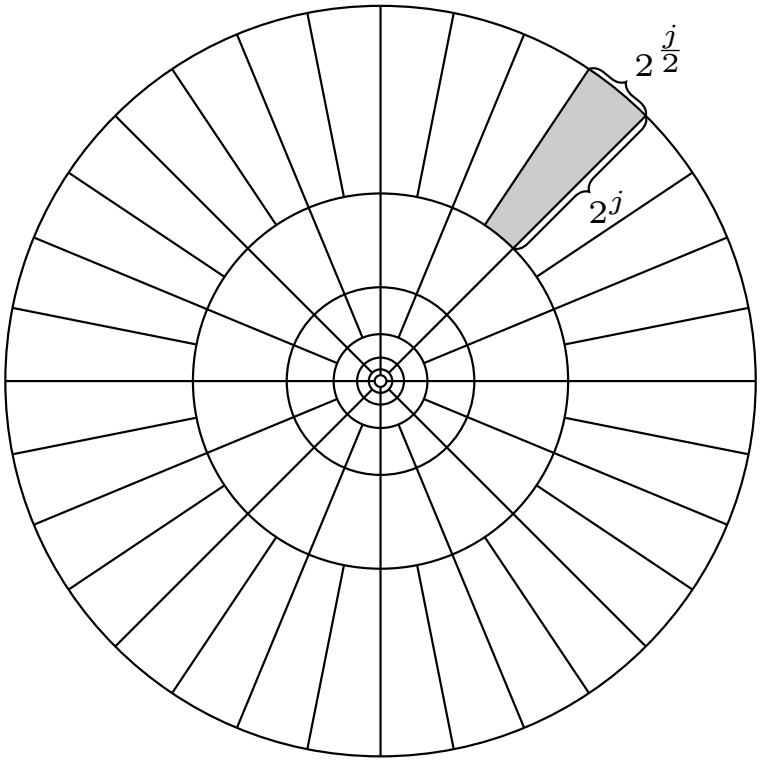


Figure: Frequency Domain

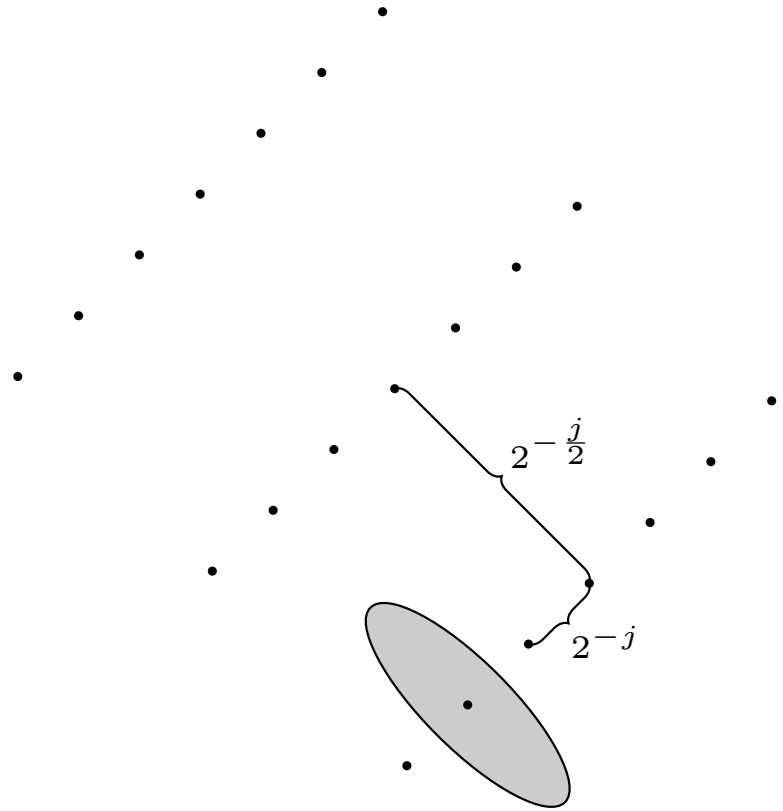


Figure: Spatial Domain

Constructing the Curvelets

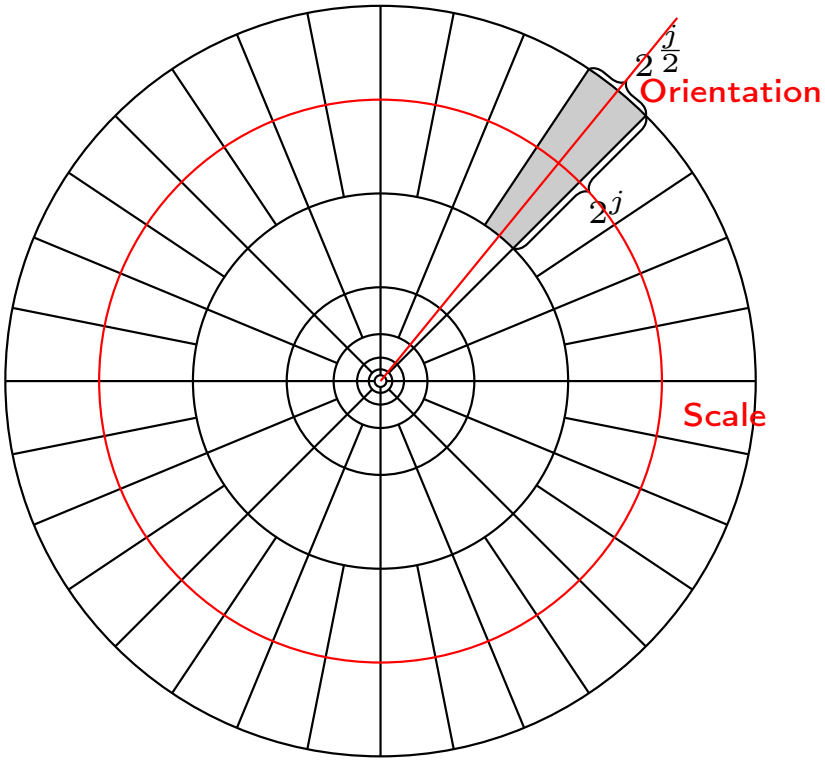


Figure: Frequency Domain

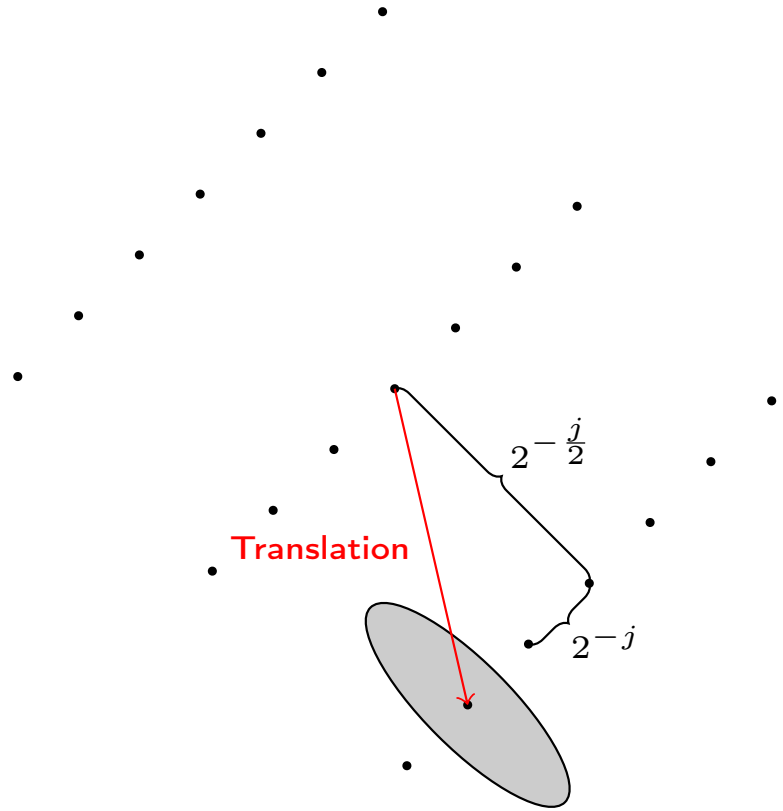


Figure: Spatial Domain

Example Curvelets

Figure: Frequency Domain

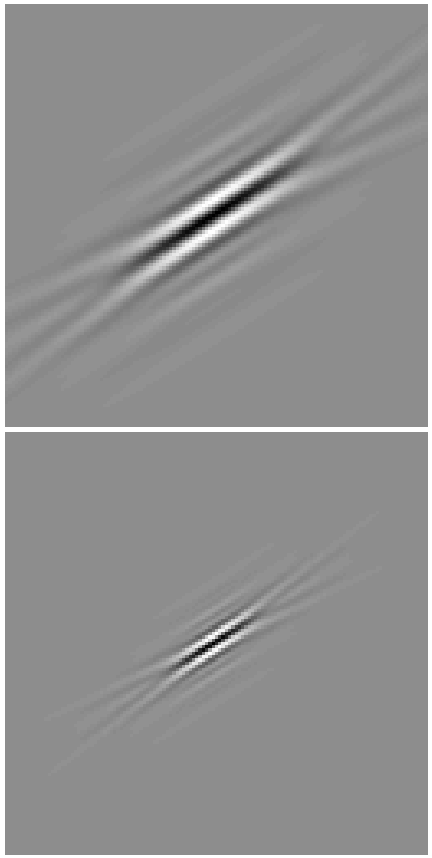


Figure: Spatial Domain

The Fast Discrete Curvelet Transform

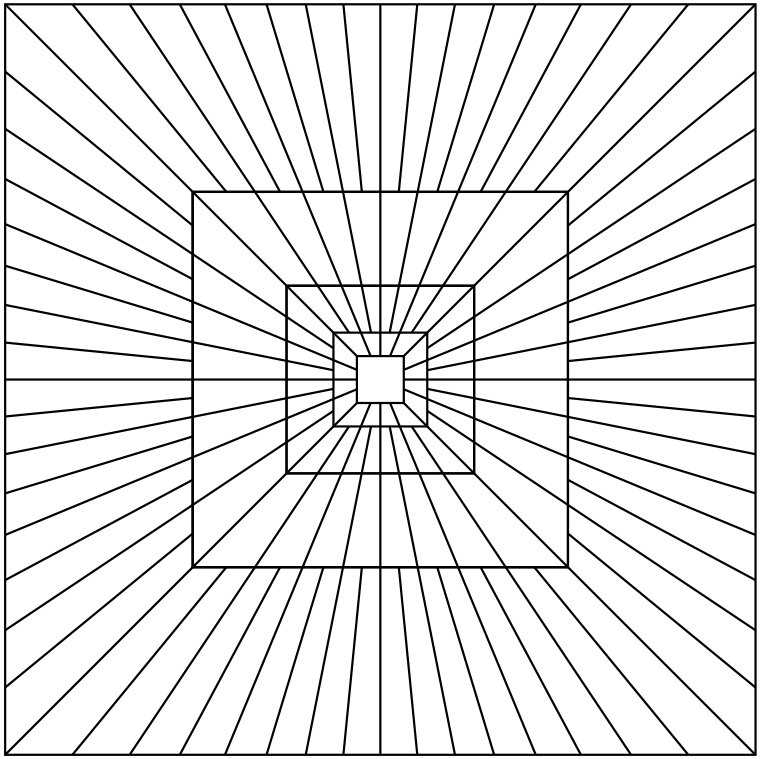


Figure: Frequency Domain

Figure: Parallelogram Support

The Fast Discrete Curvelet Transform

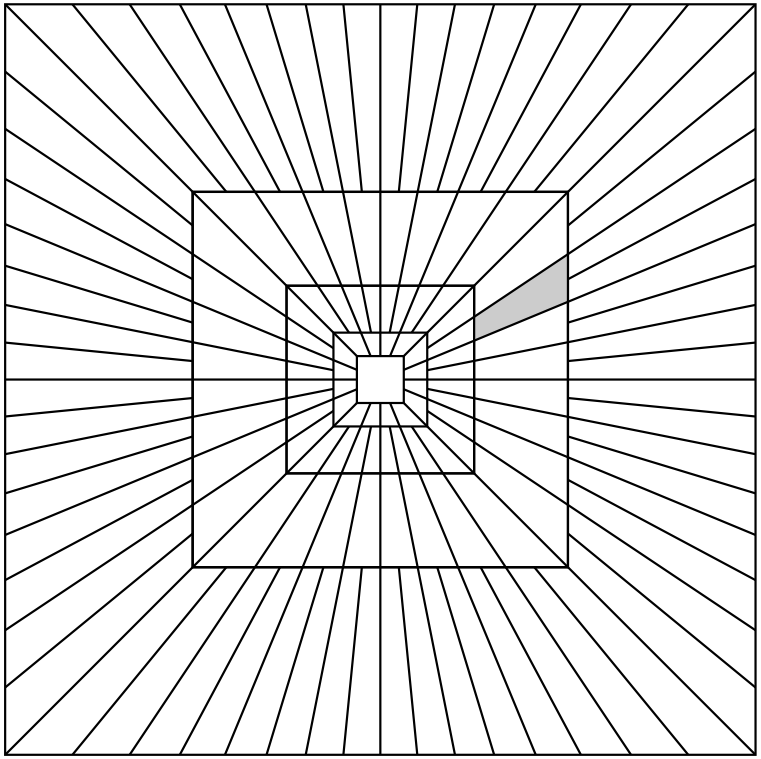


Figure: Frequency Domain

Figure: Parallelogram Support

The Fast Discrete Curvelet Transform

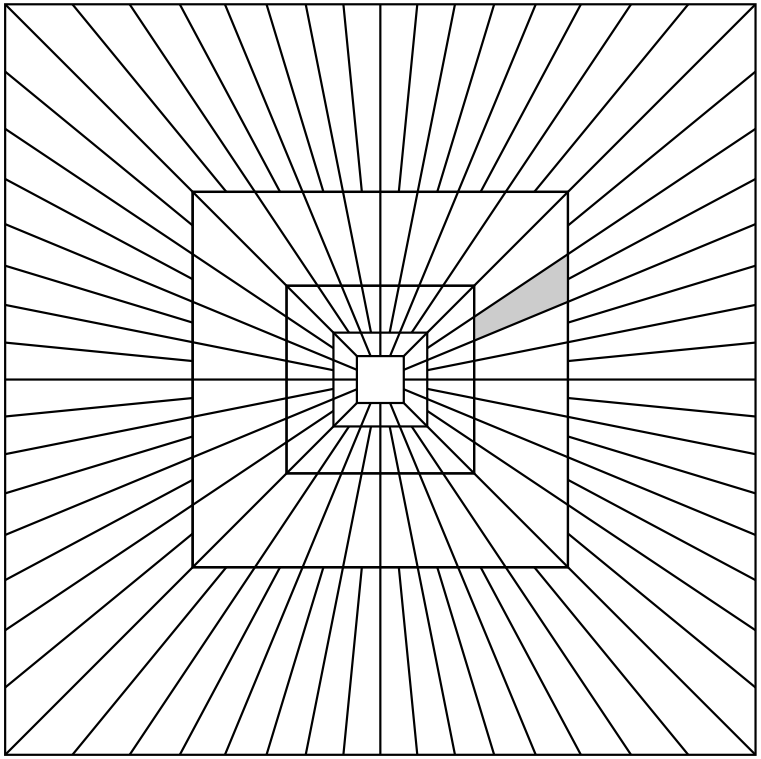


Figure: Frequency Domain

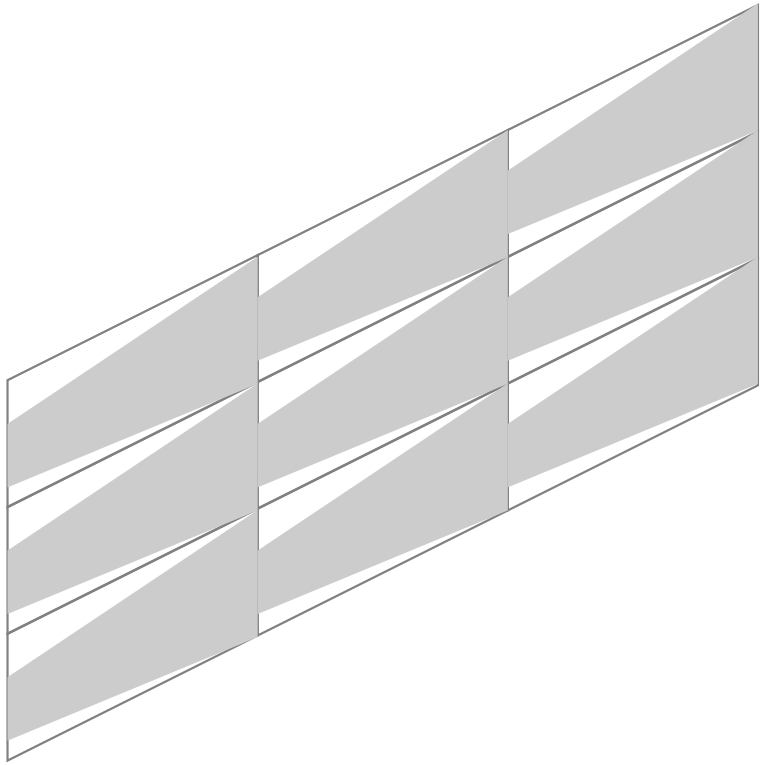


Figure: Parallelogram Support

The Fast Discrete Curvelet Transform

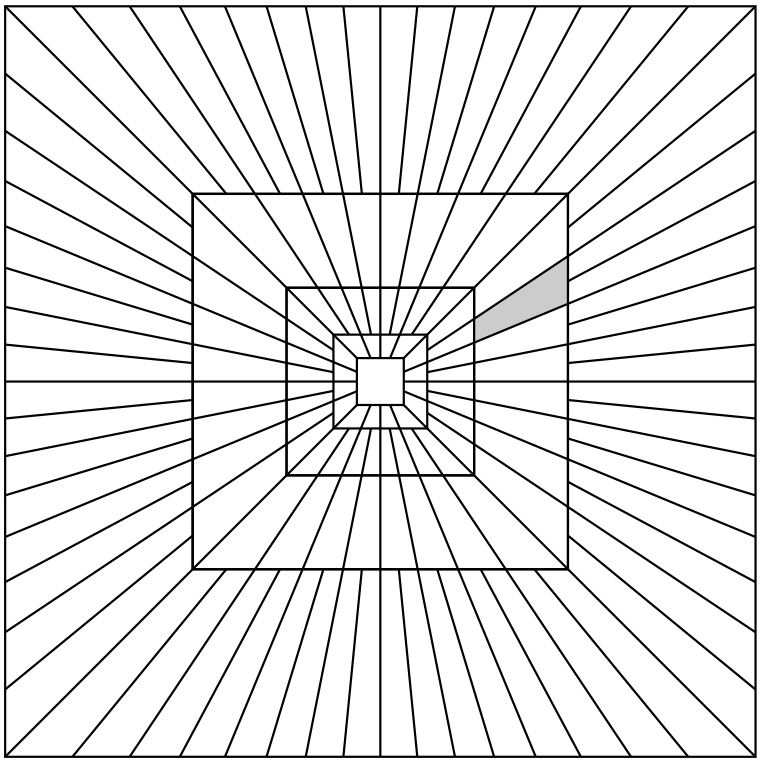


Figure: Frequency Domain

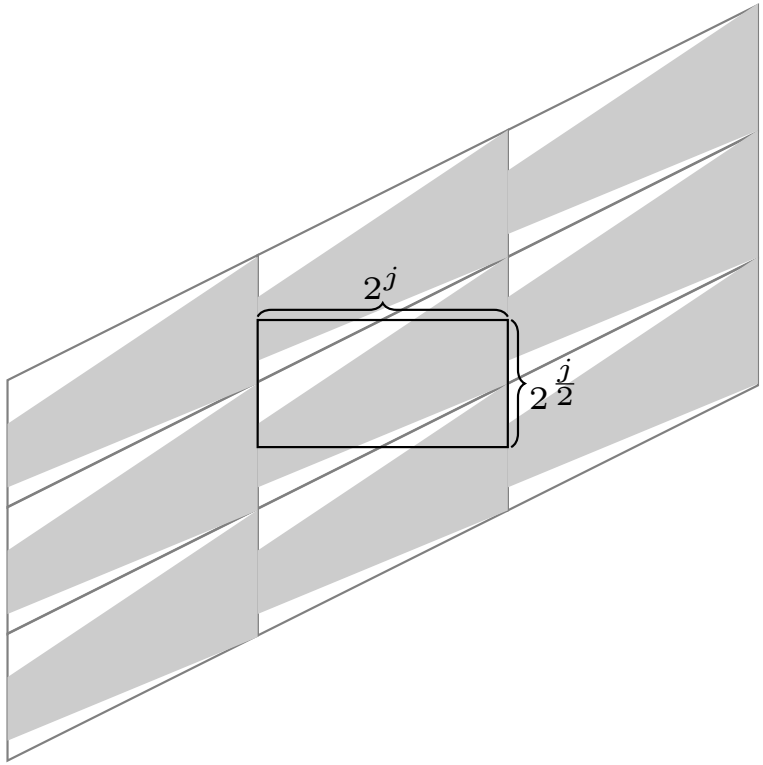


Figure: Parallelogram Support

Global Feature Extraction

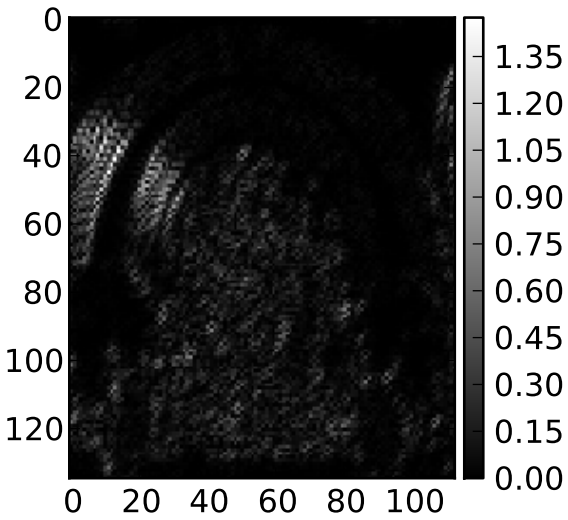


Figure: Curvelet coefficients at a specific scale and angle

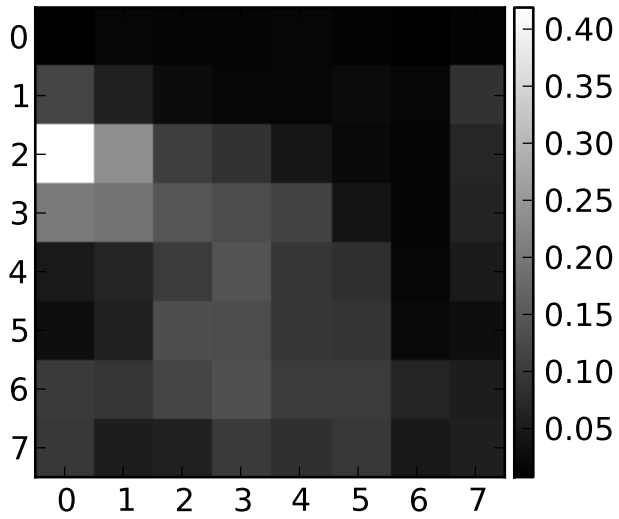


Figure: Mean values on an 8×8 grid

Local Feature Extraction (Sampling)

PMEAN Collect $(n - m + 1)^2$ sample vectors of length $N_s \cdot N_{\theta_s} \cdot m^2$ by concatenating across scales and angles

PMEAN2 Collect $N_s \cdot (n - m + 1)^2$ sample vectors of length $N_{\theta_s} \cdot m^2$ by concatenating across angles

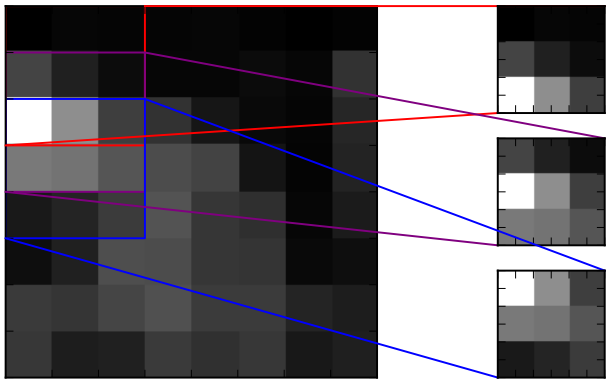


Figure: 8×8 mean coefficient grid sampled using 3×3 window

n image width and height

m window width and height

N_s Number of scales

N_{θ_s} Number of angles at scale s

Local Feature Extraction (Clustering)

- k-means clustering
- Codebook size $k = 1000$
- Each sample vector is assigned to the closest cluster S_i, $i = 1, \ldots, k$
- Image signature is the number of occurences of each "visual word" in the image:

$$\tilde{I} = [|S_1|, |S_2|, \ldots, |S_k|]$$

Ranking

foo

Benchmarking Method

foo

Cross-Domain Results

foo

Intra-Domain Results

foo

Conclusions

foo