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

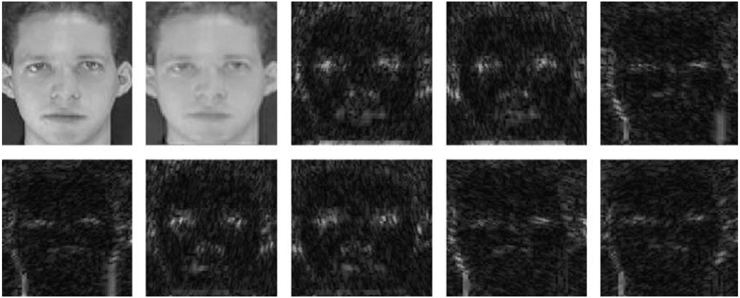


Figure: "Face recognition using curvelet based PCA.", T. Mandal and Q. M.J Wu, ICPR 2008

Prior Work on Human Recognition

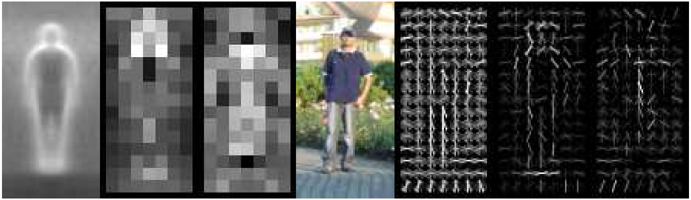
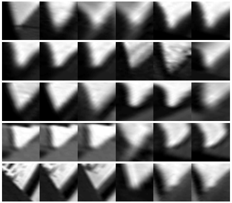


Figure: “Histograms of oriented gradients for human detection”, Dalal and Triggs, CVPR 2005

Prior Work on Visual Codebooks



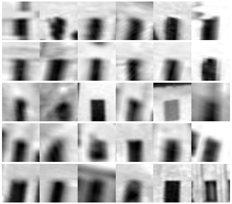


Figure: "Video Google: A text retrieval approach to object matching in videos", Sivic and Zisserman, ICCV 2003

Prior Work on Scene Classification

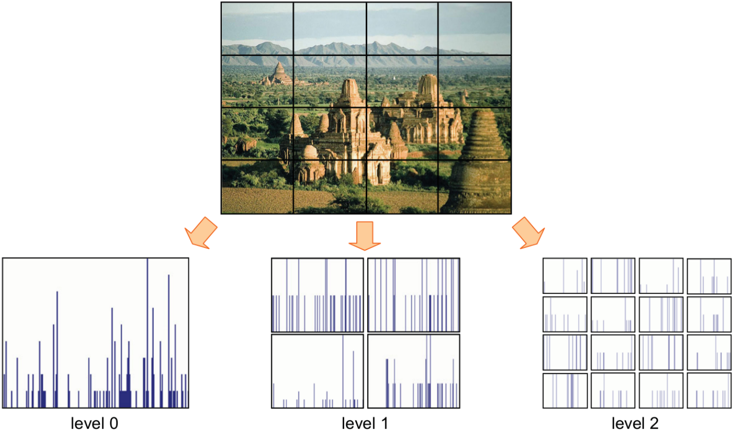


Figure: "Spatial pyramid matching", Lazebnik et al., 2009

Anatomy of a CBIR System

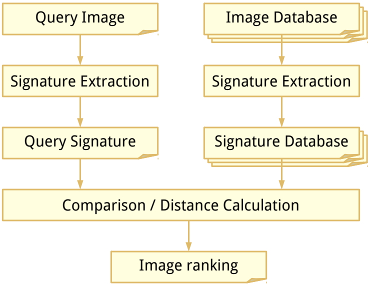


Figure: Global Descriptors

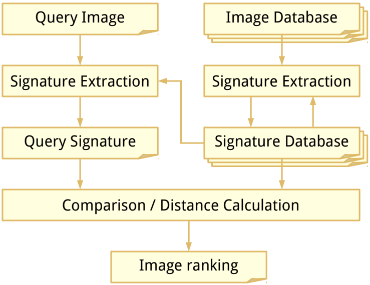
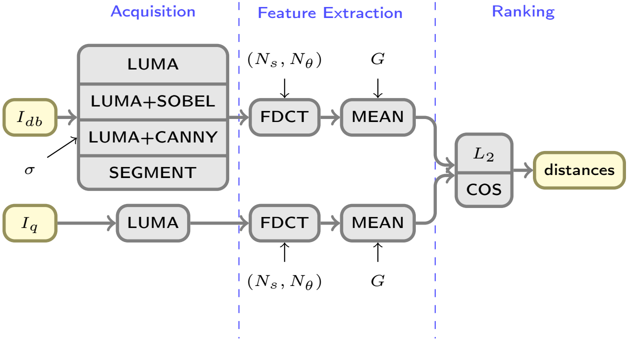
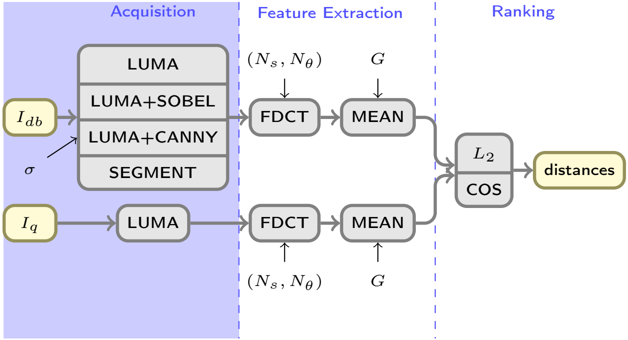


Figure: Local Descriptors

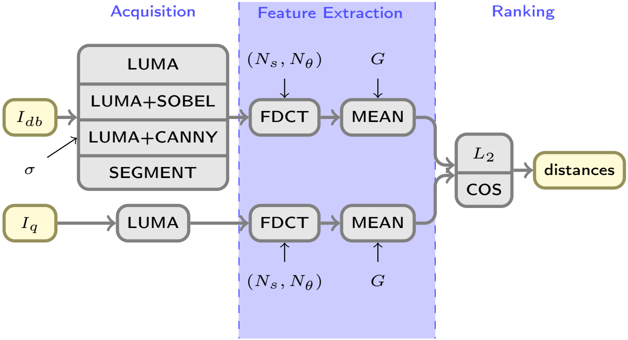
Proposed Retrieval Pipelines (Global)



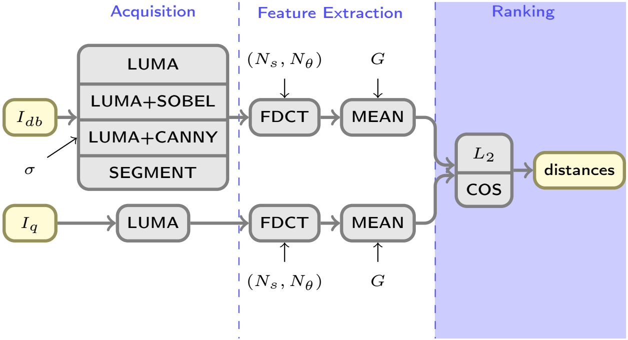
Proposed Retrieval Pipelines (Global)



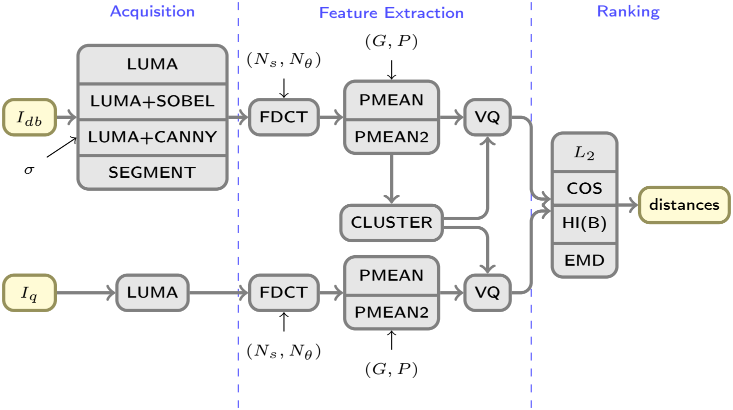
Proposed Retrieval Pipelines (Global)



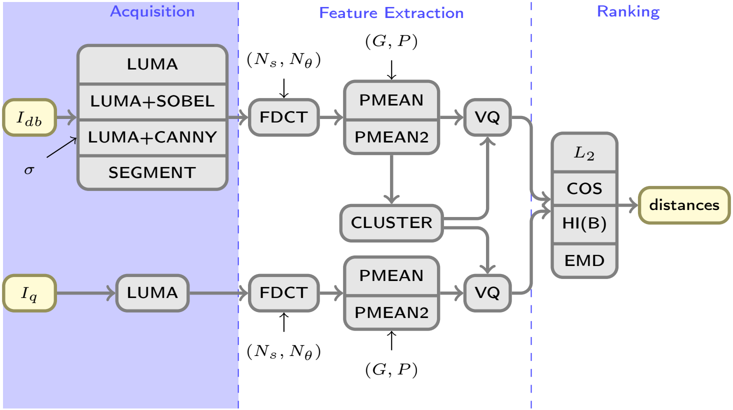
Proposed Retrieval Pipelines (Global)



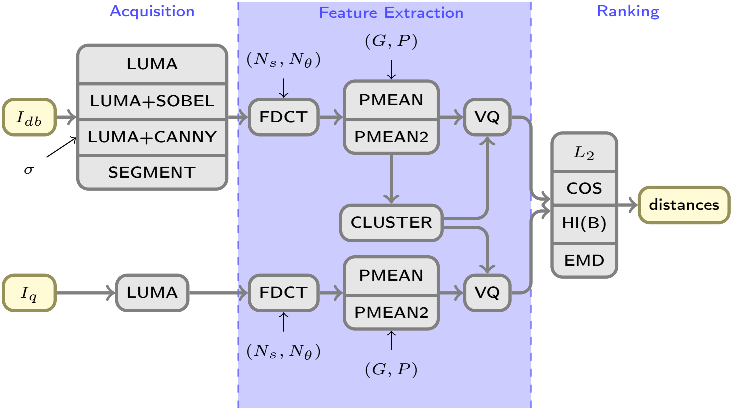
Proposed Retrieval Pipelines (Local)



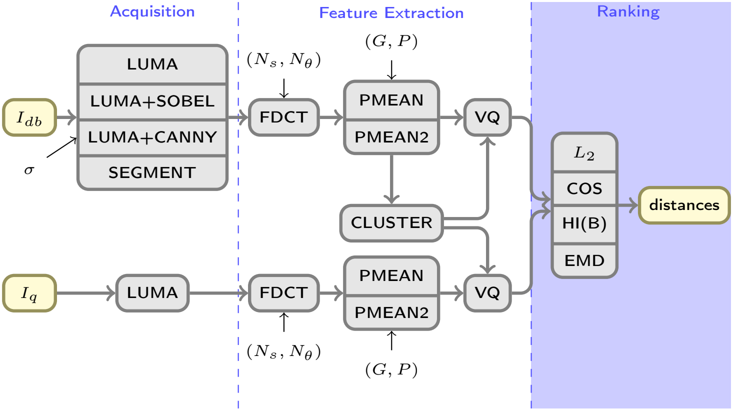
Proposed Retrieval Pipelines (Local)



Proposed Retrieval Pipelines (Local)



Proposed Retrieval Pipelines (Local)



Acquisition

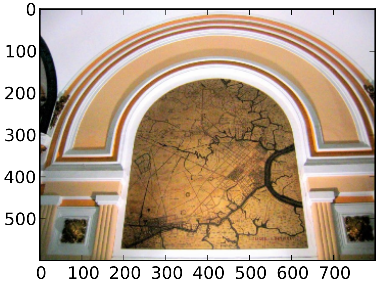


Figure: Original Image

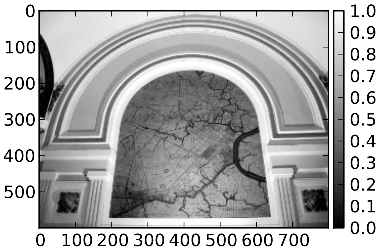


Figure: Luma Conversion

Acquisition

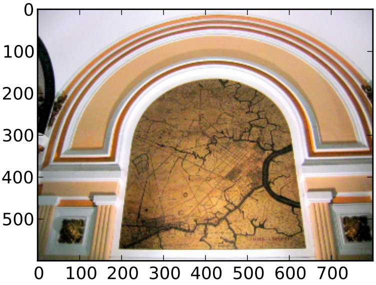


Figure: Original Image

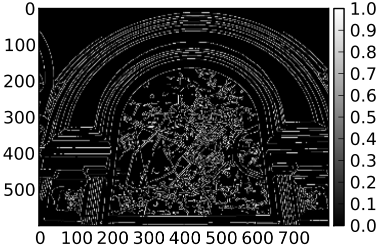


Figure: Canny Operator

Acquisition

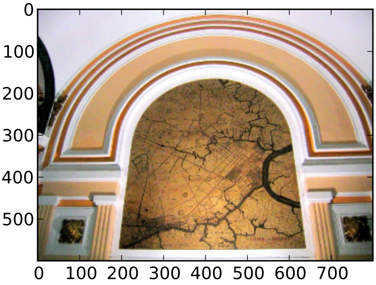


Figure: Original Image

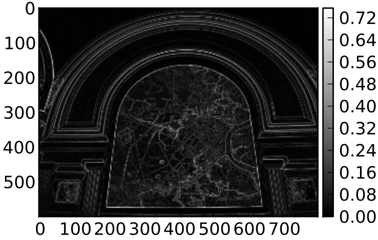


Figure: Sobel Operator

Acquisition

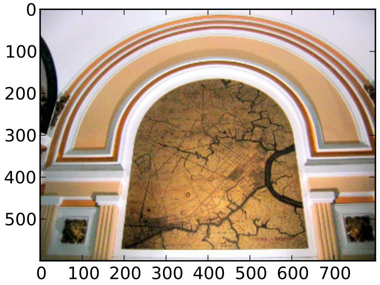


Figure: Original Image

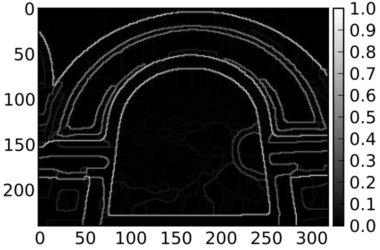


Figure: gPb-owt-ucm Transform

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 and applied in frequency domain as $\hat{\varphi}_{j,l,k}$ using the forward and inverse Fourier Transforms:

$$c(j,l,k) := \langle f, \varphi_{j,l,k} \rangle = \int_{\mathbb{R}^2} f(x) \overline{\varphi_{j,l,k}(x)} dx$$

Constructing the Curvelets

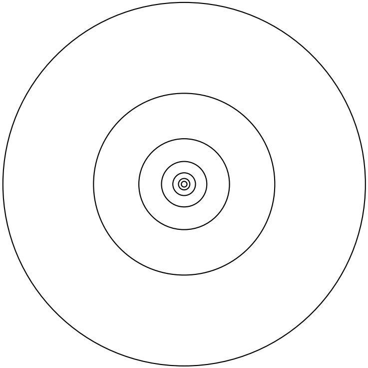


Figure: Frequency Domain

Figure: Spatial Domain

Constructing the Curvelets

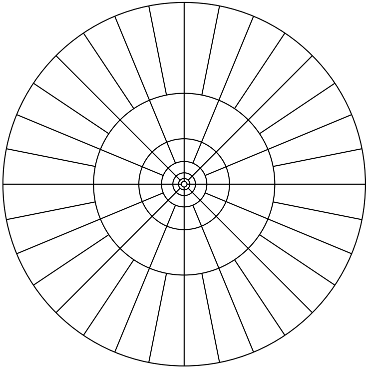


Figure: Frequency Domain

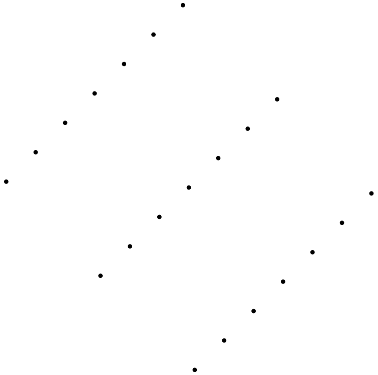


Figure: Spatial Domain

Constructing the Curvelets

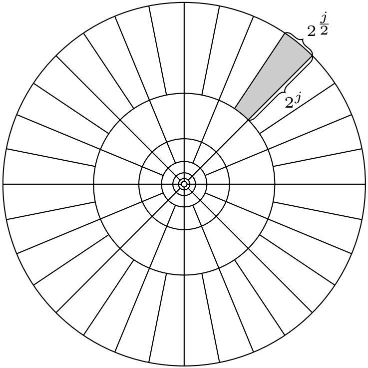


Figure: Frequency Domain

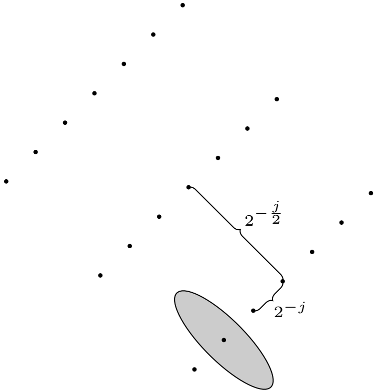


Figure: Spatial Domain

Constructing the Curvelets

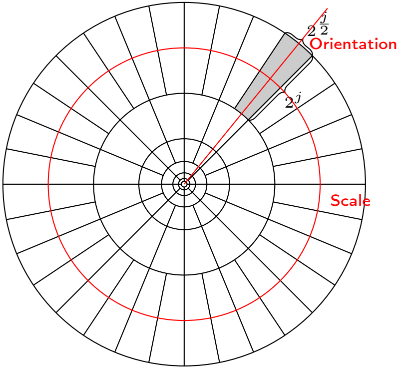


Figure: Frequency Domain

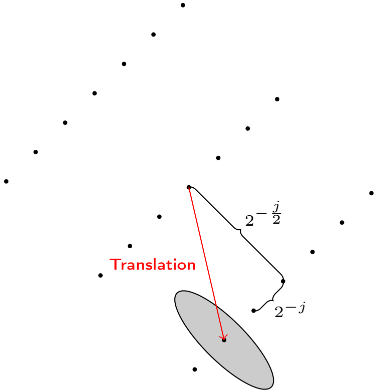


Figure: Spatial Domain

Example Curvelets

Figure: Frequency Domain

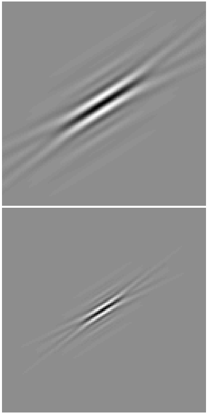


Figure: Spatial Domain

The Fast Discrete Curvelet Transform

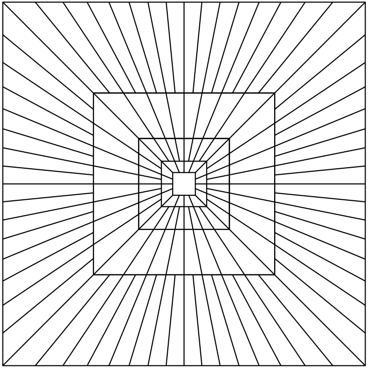


Figure: Frequency Domain

Figure: Parallelogram Support

The Fast Discrete Curvelet Transform

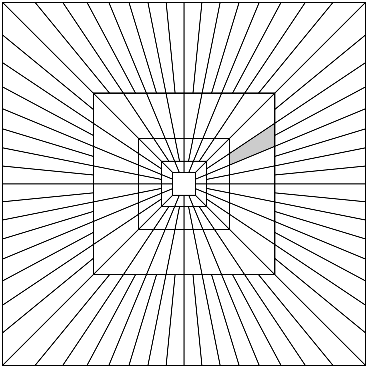


Figure: Frequency Domain

Figure: Parallelogram Support

The Fast Discrete Curvelet Transform

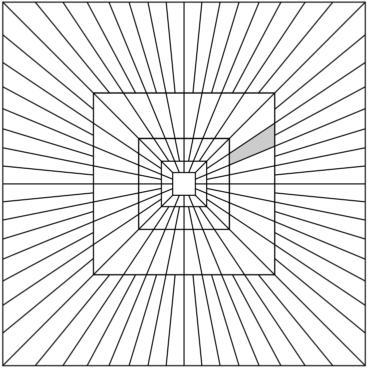


Figure: Frequency Domain

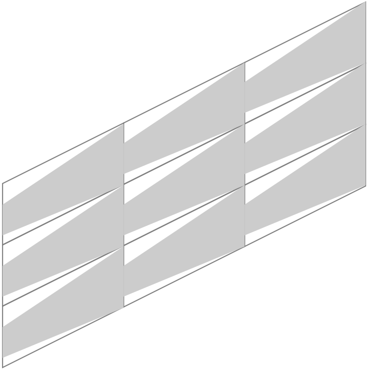


Figure: Parallelogram Support

The Fast Discrete Curvelet Transform

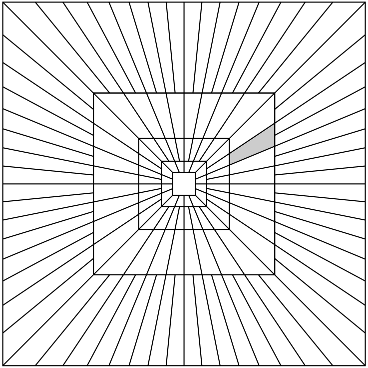


Figure: Frequency Domain

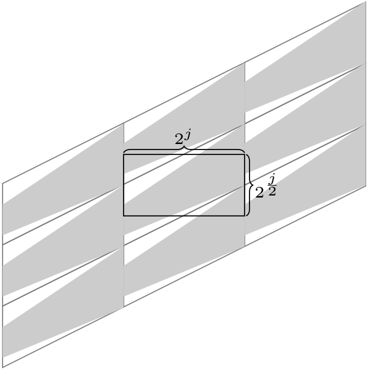


Figure: Parallelogram Support

Global Feature Extraction

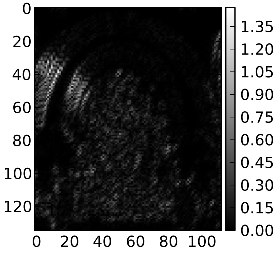


Figure: Curvelet coefficients at a specific scale and angle

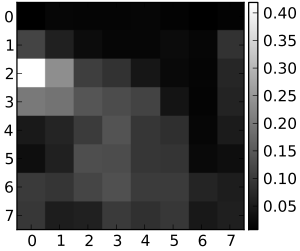


Figure: Mean values on an 8×8 grid

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

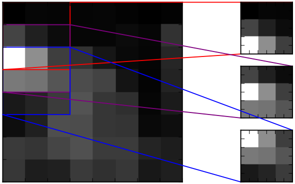


Figure: 8×8 mean coefficient grid sampled using 3×3 window

- 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 cluster S_i, $i = 1, \ldots, k$ the center of which it is closest to
- Image signature is the number of occurences of each "visual word" in the image:

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

Distance Metrics

L_2 $d_{EUCL}(p,q) = \sqrt{\sum_{i=1}^{n}(q_i - p_i)^2}$

Cosine $d_{COS}(p,q) = 1 - \frac{p \cdot q}{\|p\|\|q\|}$

Histogram Intersection (HI) $d_{HI}(P,Q) = 1 - \frac{\sum_{i=1}^{n}\min(p_i,q_i)}{\sum_{i=1}^{n} q_i}$

Earth Mover's Distance (EMD) $d_{EMD}(P,Q) = \frac{\sum_{i=1}^{n}\sum_{j=1}^{m} d_{i,j} f_{i,j}}{\sum_{i=1}^{n}\sum_{j=1}^{m} f_{i,j}}$

TF-IDF Weighting

Term t_i occurs $tc_{i,j}$ times in document $d_j \in D$ with length n_j and is present in m_i documents overall.

Term Frequency $tf_{i,j} = \frac{tc_{i,j}}{n_j}$

Inverse Document Frequency $idf_i = \log \frac{|D|}{m_i}$

Total Term Weight $w_{i,j} = tf_{i,j} \cdot idf_i = \frac{tc_{i,j}}{n_j} \cdot \log \frac{|D|}{m_i}$

Cross-Domain Dataset

Figure: Example images from "Sketch-based image retrieval: benchmark and bag-of-features descriptors", Eitz et al., 2011

Cross-Domain Benchmark

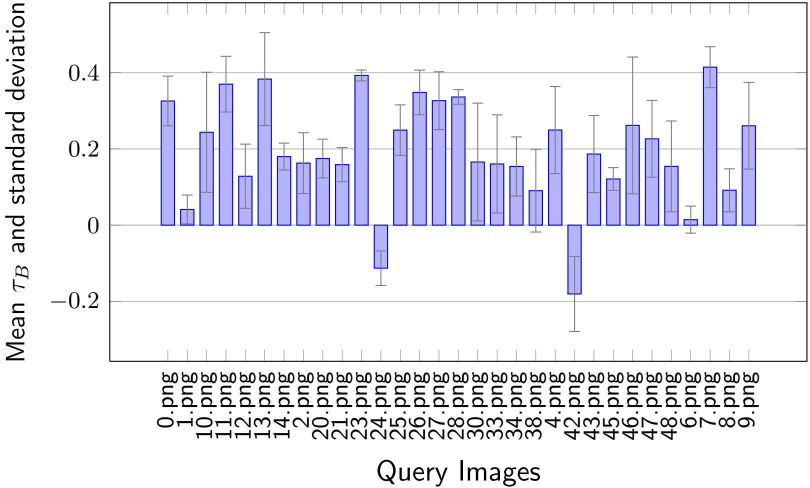
- 31 user study-based ground-truth rankings of 40 images with corresponding query sketches (Eitz et al., 2011)
- Kendall rank correlation coefficient $-1 \leq \tau_B \leq 1$
- τ_B is based on the number of similarly ordered pairs of measurements between two distributions
- $\tau_B = 1$ means same ordering, $\tau_B = -1$ means inverted ordering
- independent of the scaling differences between the two distributions

Cross-Domain Results

Preproc.	Sampling	G	P	σ	Metric	Mean τ_B correlation coefficient
CANNY	MEAN	8		1.5	COS	0.188
CANNY	PMEAN	8	3	1.5	HI	0.22
CANNY	PMEAN2	8	3	1.5	COS	0.19
LUMA	MEAN	12			COS	0.191
SOBEL	MEAN	12			COS	0.2
SOBEL	PMEAN	8	3		HIB	0.187

Table: Best performing pipeline configurations

Cross-Domain Distribution



Intra-Domain Dataset

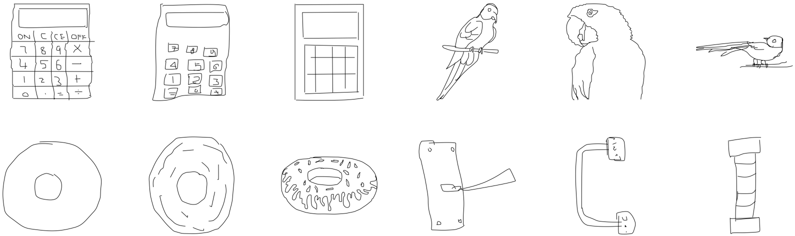


Figure: Example sketches from four categories from "How do humans sketch objects?", Eitz et al., 2012

Intra-Domain Benchmark

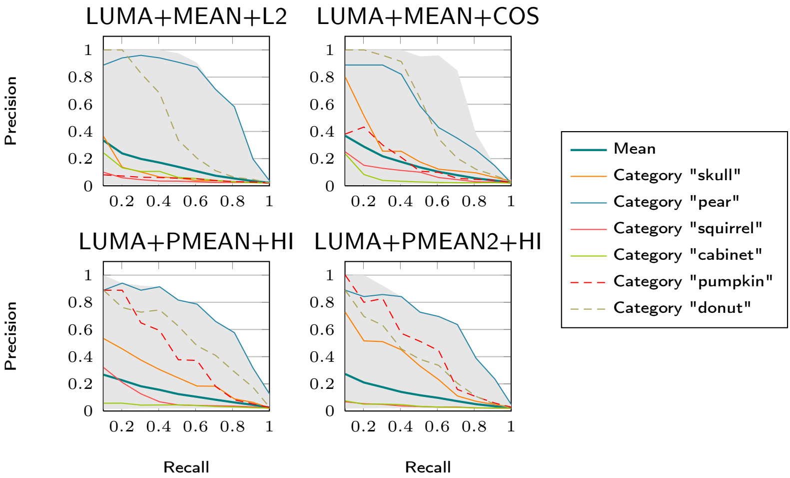
- 50 categories with 80 hand-drawn sketches each (Eitz et al., 2012)
- Precision-recall statistics

$$recall = \frac{\text{number of correct positive results}}{\text{total number of positives}}$$

$$precision = \frac{\text{number of correct positive results}}{\text{total number of results}}$$

- no edge-detecting preprocessing

Intra-Domain Results



Discussion and Conclusions

- Retrieval performance comparable to other descriptors
- For cross-domain retrieval, local LUMA+CANNY+HI performs best
- For intra-domain retrieval, global descriptors work better
- Large performance differences between queries

$\Rightarrow$ Possibly much better results for narrower problem statements and specialized applications