FELIX STÜRMER

SKETCH-BASED IMAGE RETRIEVAL USING CURVELETS

SKETCH-BASED IMAGE RETRIEVAL USING CURVELETS

FELIX STÜRMER

An Evaluation of Curvlet-Based Cross-Domain Descriptors for Sketch-Based Image Retrieval

January 2012 – version 0.1

ABSTRACT

Short summary of the contents. . .

ACKNOWLEDGMENTS

acknowledgments go here...

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LISTINGS

ACRONYMS

1 INTRODUCTION

1.1 MOTIVATION

Paragraph about increase in visual data, mobile cameras, medicine, etc...

At the core of the research into content-based image retrieval (CBIR) lies the need to be able to access the growing repositories of visual data in a convenient and efficient manner. In this context "convenient" describes the ability for the user to express the query without a complex reformulation of the intent to make it accessible to the query processor. At the same time the computational efficiency becomes more important as the amount of data to search grows. This issue becomes even more critical as the use of mobile, power-limited devices increases across many areas of application, such as autonomous vehicles or handheld augmented reality devices.

Research into text-based information retrieval has brought into existence many statistical methods to query a potentially large body of text using text as the query input. This preserves the close mapping of the intent of the user to the expression of the query and thereby makes the process accessible to users without knowledge about the internal workings of the retrieval system. Providing the means to access a large amount of visual data using a system with similar properties has turned out not to be an easy problem to solve. Using text-based querying for that purpose depends on the ability to reliably label visual data, which would require solving the general object recognition problem first [3]. To avoid that obstacle and to free the retrieval system from the requirement of translating between textual and visual information, many methods to search an image database using visual similarity have been developed.

While the goals of those systems are very similar, they differ considerably in many aspects of the processing pipeline. The query input ranges from example images over drawings to predicate describing color and shape distribution. Similarly, the structure and content of the databases and the means by which the systems query and rank the results vary significantly. This thesis focuses on evaluating a system that uses hand-drawn sketches as inputs to query databases of either full-color images or contour images. The Fast Discrete Curvelet Transform [1] is used to analyse image segments.

1.2 OUTLINE

Chapter 2 presents the structure of the problem and prior solutions. The following Chapter 3 proposes several variations of a particular solution using the Fast Discrete Curvelet Transform [1]. The experimental setup and its results are documented in Chapter 4 and analysed in Chapter 5. In Chapter 6 several possible conclusions are drawn and pointers towards future research are given.

2 BACKGROUND & RELATED WORK

2.1 THE SEMANTIC GAP

One of the core insights of computer vision in general and content based image retrieval in specific probably is that human perception is inseparably linked to interpretation by the brain. As a human individual there is no way to directly access visual information without them having been filtered and weighted by one's personal experiences and cultural context. Therefore, when people talk about visual similarity of images, it usually includes a large degree of semantic similarity unconciously added to the perception. The difference between that mode of perception and the current algorithmic ways to analyse visual data has been eloquently coined *the semantic gap* by [3]:

> 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.

Having had that realisation can guide the decision of a researcher or designer of such systems.

2.2 THE SENSORY GAP

In addition to the semantic ambiguity described above, another major obstacle of computer vision impacts a CBIR system: *the sensory gap*. This term has also been coined by [3], where it's defined as follows:

> 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.

That terse definition includes a multitude of conditions, that can affect an image, which a CBIR system operates on:

ILLUMINATION The brightness or direction of the illumination can hide or accent edges and texture properties in the scene. Similarly, the color of the illumination influences the recorded color information in the image.

RESOLUTION The imaging resolution sets a lower limit on the size of features that can be correctly recognised by any algorithm. As in all signal processing applications, aliasing of high frequency components of the image can introduce further ambiguities. [2]

OCCLUSION Depending on the viewpoint of the recording and the composition of the scene, distinguishing parts of depicted objects may be occluded by other objects or objects may be only partially inside the recorded image.

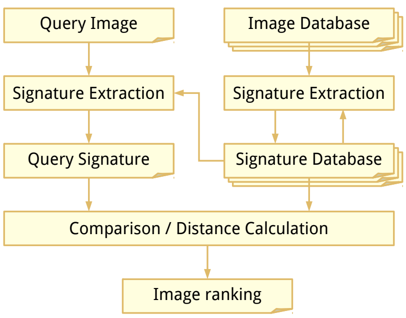
PERSPECTIVE An object's proportions can be distorted by the imaging perspective.

An ideal CBIR system would use feature extraction and comparison methods that can account and correct for such conditions.

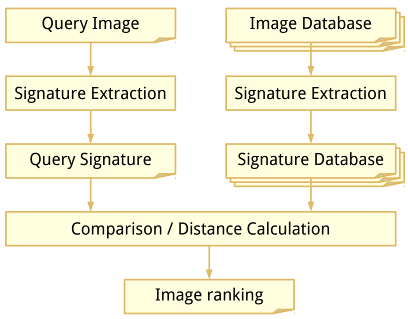
2.3 ANATOMY OF A CBIR SYSTEM

The inner workings of most CBIR systems can best be examined by looking at the processing pipeline each query has to go through. The coarse sequence of computational steps is almost the same in all such systems (Figure 1):

1. Acquire the image.
2. Extract the signature using a feature extraction algorithm.
3. Compare the signature to a database containing the signatures of the images to search within.
4. Rank the images by similarity using the comparison results.



(a) Local features



(b) Global features

Figure 1: Coarse structure of a CBIR system

OLD STUFF BELOW

Most approaches can be characterized by looking at three stages in their processing pipeline:

INPUT FORMAT The structure of the input data determines the amount of information available to the subsequent processing steps. Possible preprocessing steps include color space conversion, scaling and edge extraction.

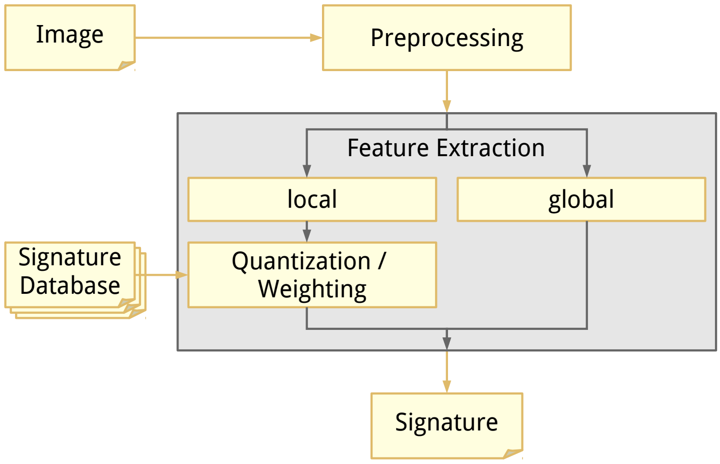


Figure 2: Signature extraction in CBIR systems

EXTRACTED FEATURES Many algorithms produce a large number of coefficients that can be reduced to a set of feature coefficients using by techniques such as vector quantization or principal component analysis (PCA).

DISTANCE METRIC In order to rank the images according to similarity a metric is used to calculate the distance in feature space between two sets of feature coefficients. The selection of a metric is often closely coupled with the feature extraction algorithm.

2.3.1 *input format*

Complete vs incomplete sketches, intra-/cross-domain

2.3.2 *features*

- bag of features from k-means clustered visual words [video google]
- histogram of oriented gradients [chalechale + refs]

2.3.3 *metric*

- after ranking using euclidean distance, rank by spatial similarity [video google]
- Earth Mover's distance? [rubnerljcv00]

3 PROPOSED SOLUTION

Proposed solution goes here...

3.1 INPUT FORMAT

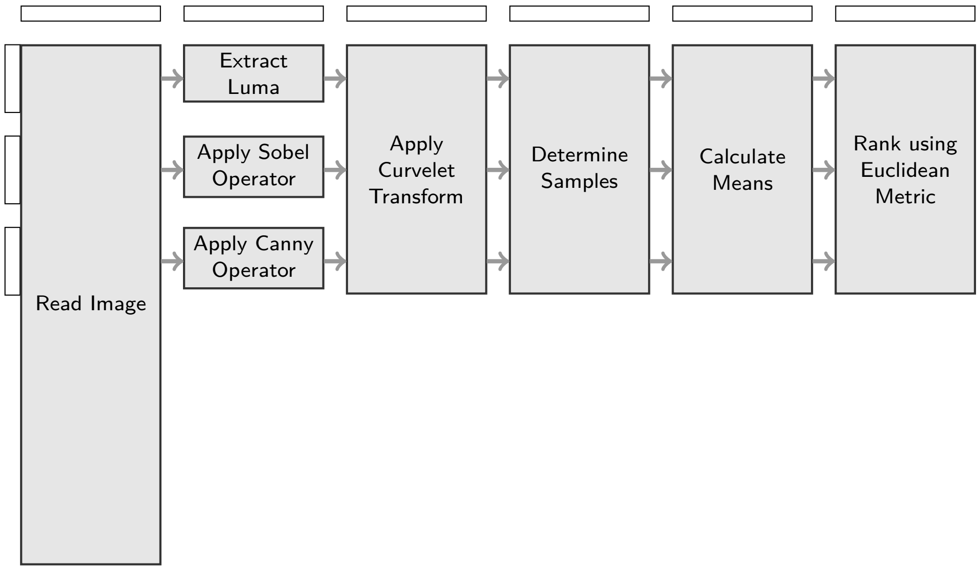
- Luma component (Y') of Y'UV representation
- Gradient magnitude of Sobel operator of luma component
- Canny edge map of luma component
- gPb

3.2 FEATURE EXTRACTION

- Global features: mean and standard deviation
- Local features: visual words via k-means clustering
- great comparison of sampling for k-means clustered vws [nowak06]

3.3 DISTANCE METRIC

- Euclidean Distance
- cosine distance?
- EMD?


Read Image
Extract Luma
Apply Sobel Operator
Apply Canny Operator
Apply Curvelet Transform
Determine Samples
Calculate Means
Rank using Euclidean Metric

4 EXPERIMENTAL RESULTS

Experimental results go here. . .

5 ANALYSIS

Analysis goes here…

6 CONCLUSION

Conclusion goes here. . .

BIBLIOGRAPHY

[1] E. Candes et al. "Fast discrete curvelet transforms." In: *Multiscale modeling and simulation* 5.3 (2006), 861–899.

[2] C. E. Shannon. "Communication in the presence of noise." In: *Proceedings of the IEEE* 86.2 (1998), 447–457. URL: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=659497.

[3] A. W.M Smeulders et al. "Content-based image retrieval at the end of the early years." In: *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 22.12 (2000), 1349–1380.

COLOPHON

This document was typeset using the typographical look-and-feel `classicthesis` developed by André Miede. The style was inspired by Robert Bringhurst's seminal book on typography "*The Elements of Typographic Style*". `classicthesis` is available for both LaTeX and LyX:

http://code.google.com/p/classicthesis/

Happy users of `classicthesis` usually send a real postcard to the author, a collection of postcards received so far is featured here:

http://postcards.miede.de/

DECLARATION

Put your declaration here.

Berlin, January 2012

Felix Stürmer