



APPLICATION OF TEXT MINING ON SPATIAL VISUAL SENTENCES

MASTER THESIS PRESENTATION

Xiwen Cheng

July 22, 2015

Supervisor: Dr. Michael S. Lew
Leiden Institute of Advanced Computer Science

TABLE OF CONTENTS

1. Introduction

2. Approach

3. Results

4. Conclusions

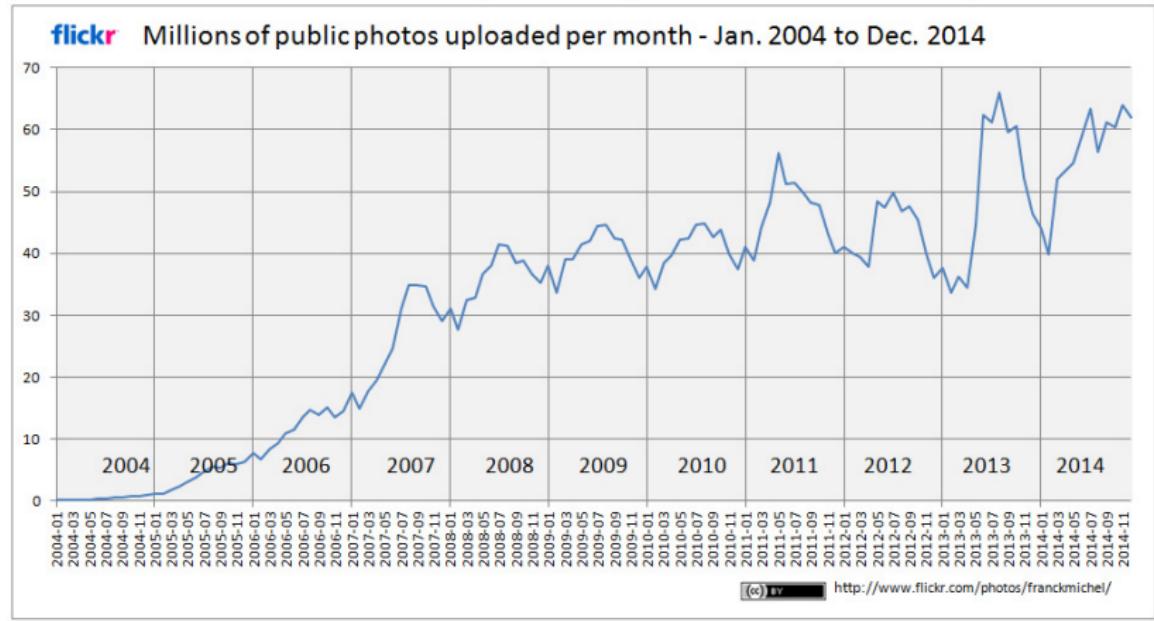
INTRODUCTION



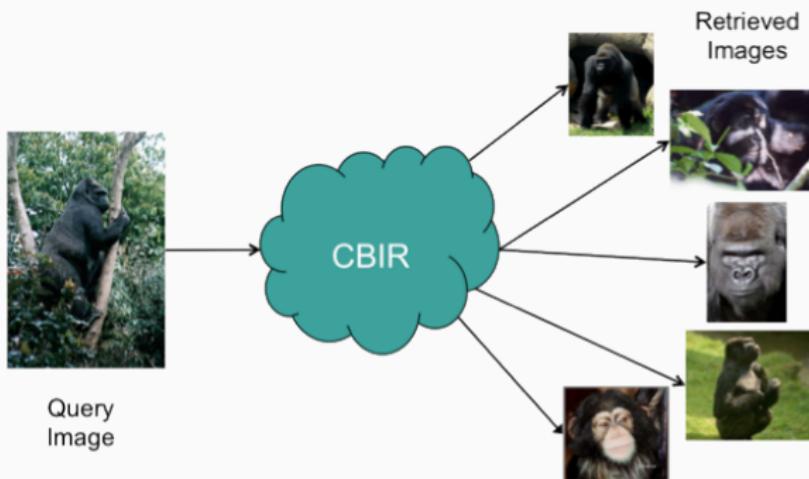
AGE OF DIGITAL PHOTOGRAPHY

- **Affordable portable devices:** smartphones, tablets & DSLR cameras
- **Trends:** Flickr, Instagram, Facebook & Snapchat

OVER 10 BILLION IMAGES AT FLICKR

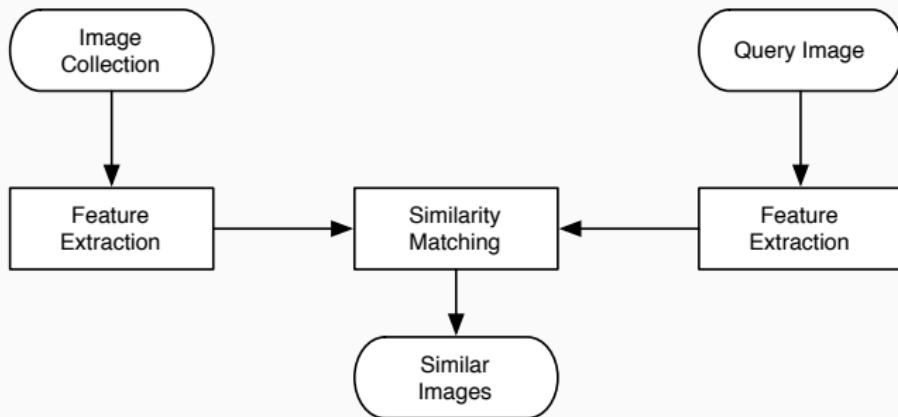


RETRIEVE RELEVANT IMAGES



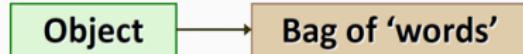
Given one or more sample images as input return a list of images from a collection sorted by relevance.

CONTENT BASED IMAGE RETRIEVAL



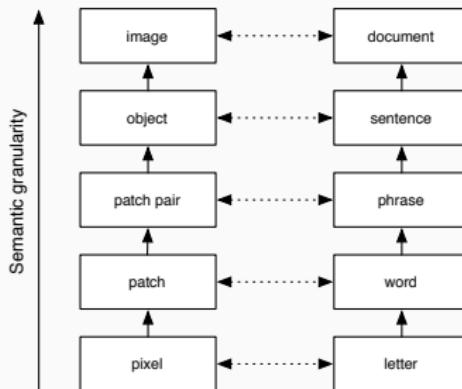
Typical CBIR System

BAG OF VISUAL WORDS



Bag-of-Visual-Words model: inspired by text retrieval method where a document is represented by a set of distinct keywords.

SEMANTIC ANALOGY BETWEEN IMAGE AND TEXT DOCUMENT



Zheng, Qing-Fang, Wei-Qiang Wang, and Wen Gao. "Effective and efficient object-based image retrieval using visual phrases." ACM, 2006.

PIXEL - LETTER



Figure 1: b

PATCH - WORD

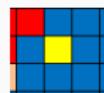


Figure 2: abbbcbffff

PATCH PAIR - PHRASE

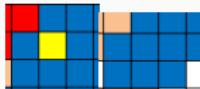


Figure 3: abbbcbffff dbbbbbbbbbb

OBJECT - SENTENCE

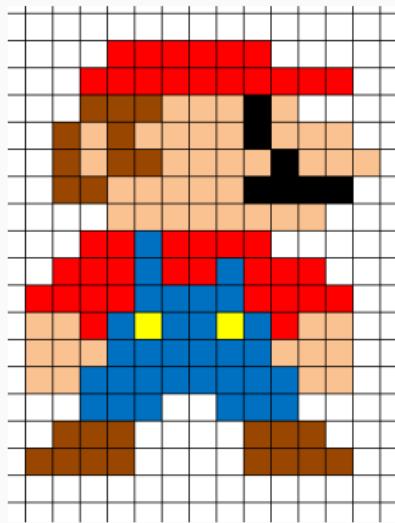


Figure 4: "Mario"

DEFINITION: SENTENCE

“A sequence of words capable of standing alone to make an assertion, ask a question, or give a command, usually consisting of a subject and a predicate containing a finite verb.”

<http://www.wordreference.com/definition/sentence>

SPATIAL VISUAL SENTENCE

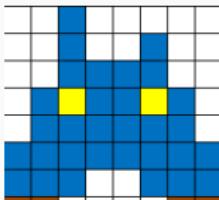
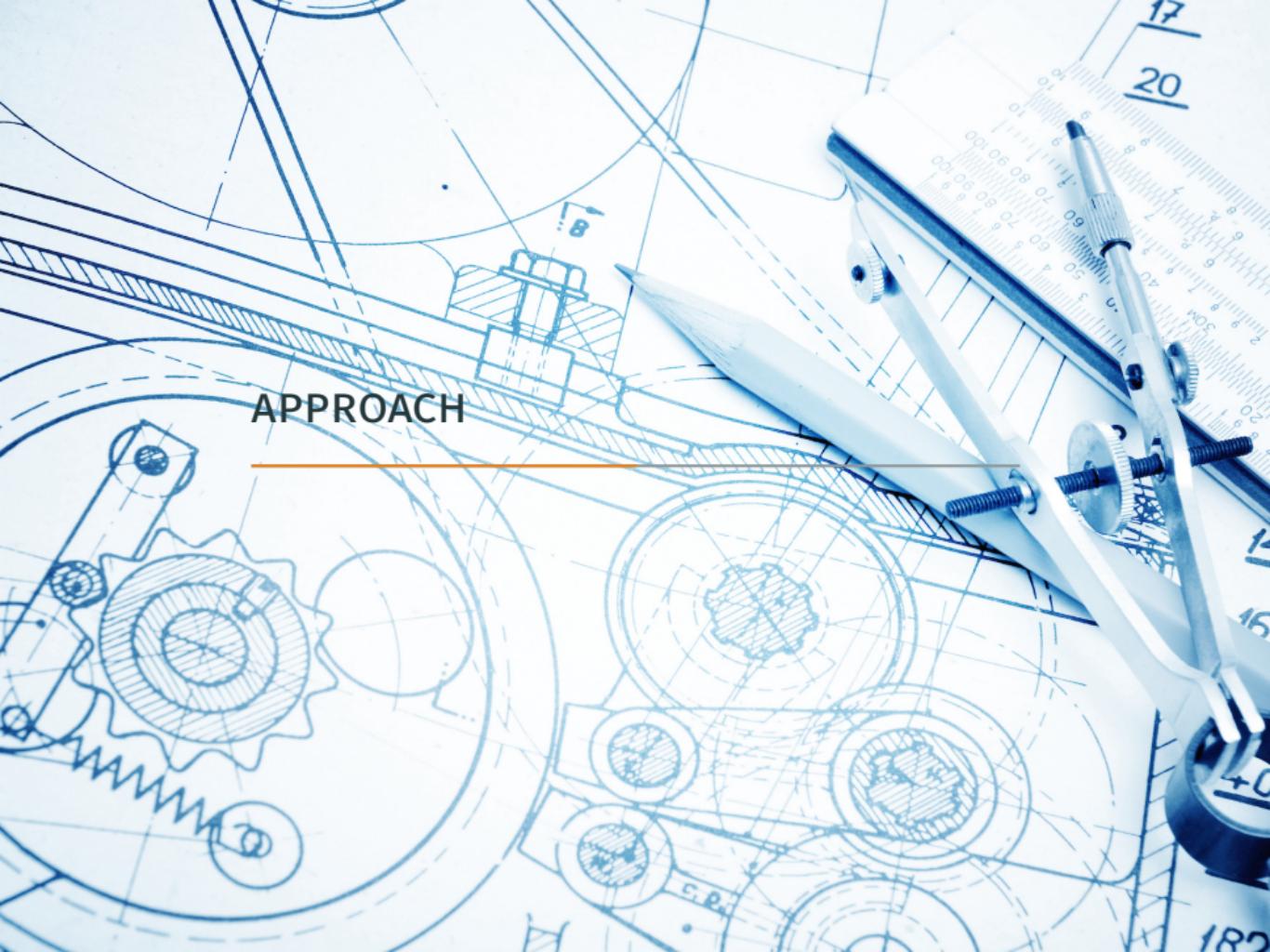


Figure 5: “Blue pant with 2 yellow buttons”

Spatial visual sentence is an ordered sequence of features (**words**) that belong together if they reside (**spatial**) in a common image segment (**visual semantics**).



A detailed technical blueprint of a mechanical system, likely a gear assembly, is shown. The drawing includes various components such as gears, a central housing, and a spring. A pen and a ruler are placed diagonally across the top right corner of the drawing. The word "APPROACH" is printed in bold capital letters at the bottom left of the drawing area.

APPROACH

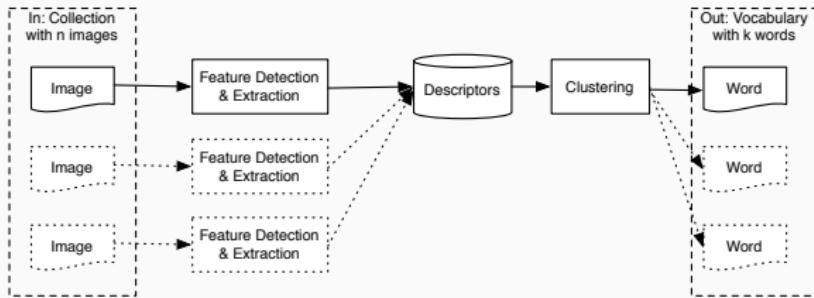
SPATIAL VISUAL SENTENCES ALGORITHM

Input: Collection C containing n images

Output: Album A used for offline search

- 1: Compute Vocabulary
- 2: **for all** Image d in C **do**
- 3: Identify the set of segments S in d
- 4: **for all** Segment s in S **do**
- 5: Compute sentence v by selecting word $w \in d$ if its position
is in s
- 6: **end for**
- 7: Add d with augmented sentences V to A
- 8: **end for**
- 9: **return** A

COMPUTING VOCABULARY



Input collection is a subset of original collection.

Descriptiveness determined by vocabulary size.

VISUAL WORD CONSTRUCTION

For each image $d = (\vec{v}_0, \dots, \vec{v}_i)$ from the original collection C its word representation \vec{d} is computed:

$$\vec{d} = (w_0, \dots, w_n) \text{ where } w_i = \text{best_matching_word}(A, \vec{v}_i) \quad (1)$$

Mark and discard noisy words based on text retrieval concepts:

- Too low occurrences: Too specific
- Too high occurrences: Stop words

SEGMENTATION



(a) Original



(b) Canny Edges

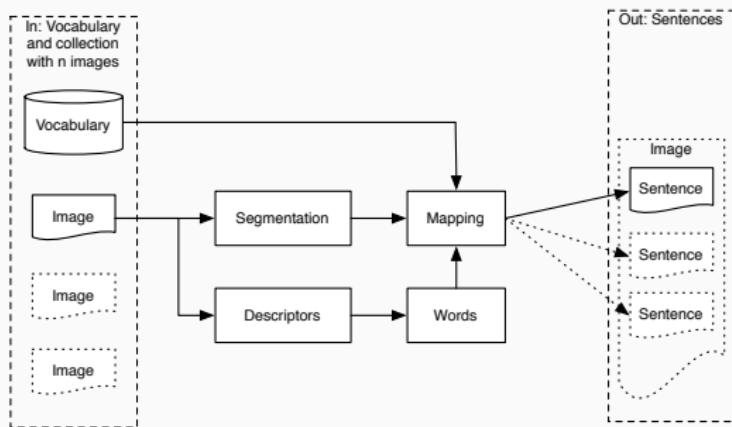


(c) Otsu Threshold



(d) SLIC Superpixels

SPATIAL VISUAL SENTENCE CONSTRUCTION



RESULTS



DATASET: UKBENCH



ukbench00164.jpg



ukbench00165.jpg



ukbench00166.jpg

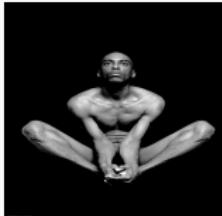


ukbench00167.jpg

DATASET: MIRFLICKR



im6637.jpg:
daffodil,
flower, fleur,
yellow, trum-
pet, blue,
light, sunlight,
conservatory,
105mm



im14091.jpg:
365days,
d40, day359,
nikon, raw,
wednesday,
bsb



im4761.jpg:
tohonochul,
flower, im-
pressed-
beauty



im455.jpg:
rome, roma,
objects, my-
favourites,
life, italy,
eos350d, dig-
ital, canon,
byfotorita, bw,
white, black, ...

TRAINING AND TEST SET

150 in training-set

50 in test-set

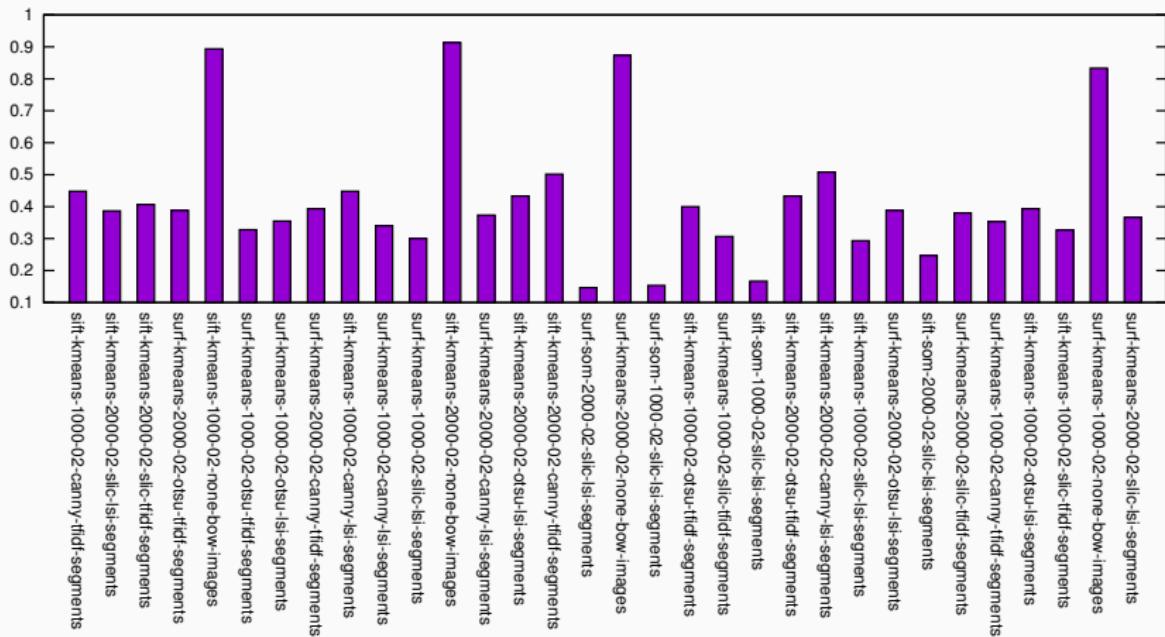
- UKBench: split 3:1 ratio
- MIRFLICKR: just random

ALGORITHM CONFIGURATIONS

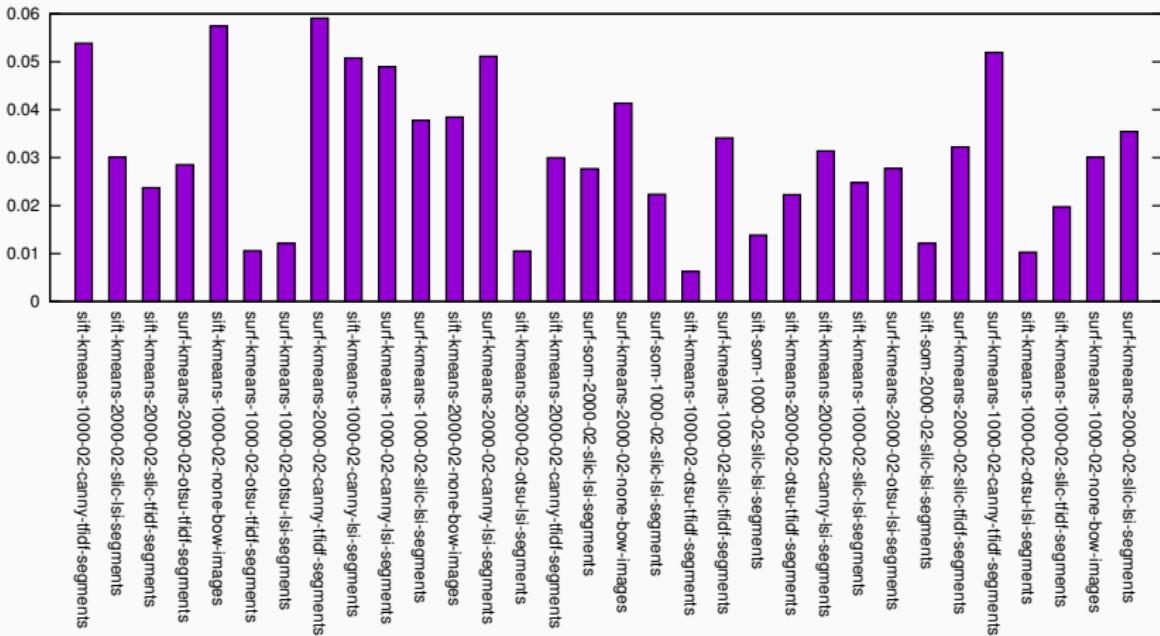
Phase/Parameter			
Feature Descriptors	surf	sift	
Vocabulary clustering	kmeans	som	
Segmentation	otsu	canny	slic
Similarity measure	ratcliff	tfidf	lsi
Corpus mode	images	segments	
Vocabulary size	1000	2000	
Noise ratio (words discarded)	0.20		

Table 1: Parameter value and implementation variations

AVERAGE F-SCORES UKBENCH



AVERAGE F-SCORES MIRFLICKR



CONCLUSIONS



PROMISING

- Standard Bag-of-words performs better than our algorithm
- **SVS performs reasonably well on UKBench**
- But it had trouble retrieving relevant images from MIRFLICKR (So did BoW)

FUTURE WORK

- Family photo album as dataset with recurring *objects*
- Speed up assessment and tuning

QUESTIONS?