Low-Cost Visual Feature Representations for Image Retrieval

Ramon Figueiredo Pessoa

Advisor: Jefersson A. Santos

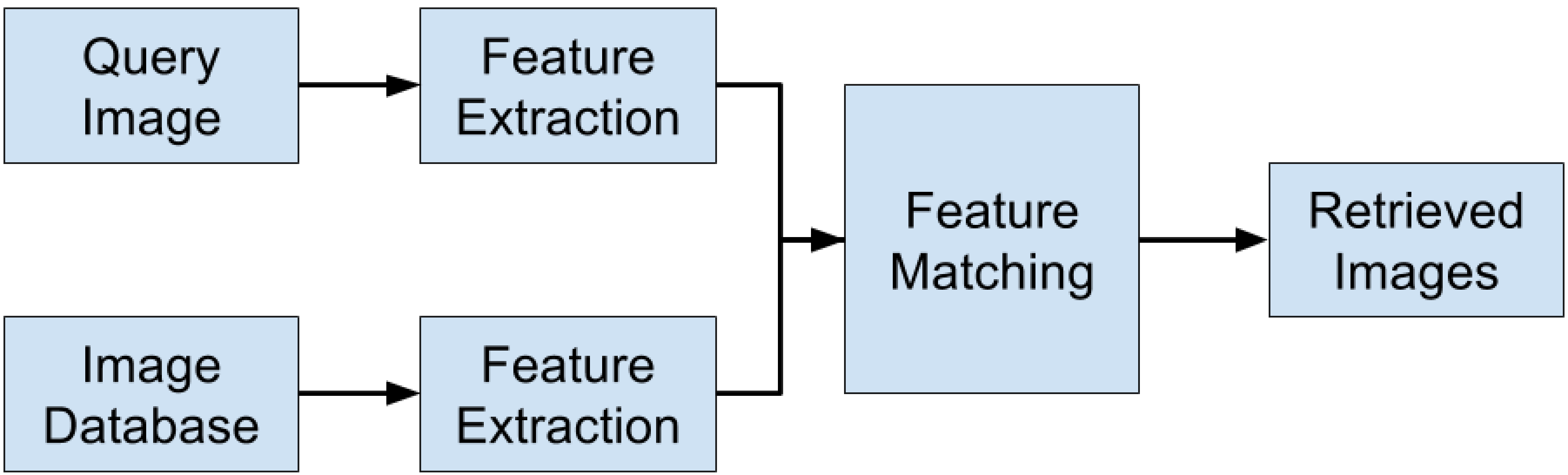
Co-Advisor: William Robson Schwartz

Master Thesis

Outline

- Introduction
- Background
- Benchmark Datasets
- Low-Cost Representation for Mobile Image Search
- Spatial Feature Representation for Mobile Image Search
- Conclusion and Future Words

Content-Based Image Retrieval (CBIR)



Block diagram of a basic CBIR system

CBIR: Semantic Gap

Query: Palmeiras

Relevance is defined by the user

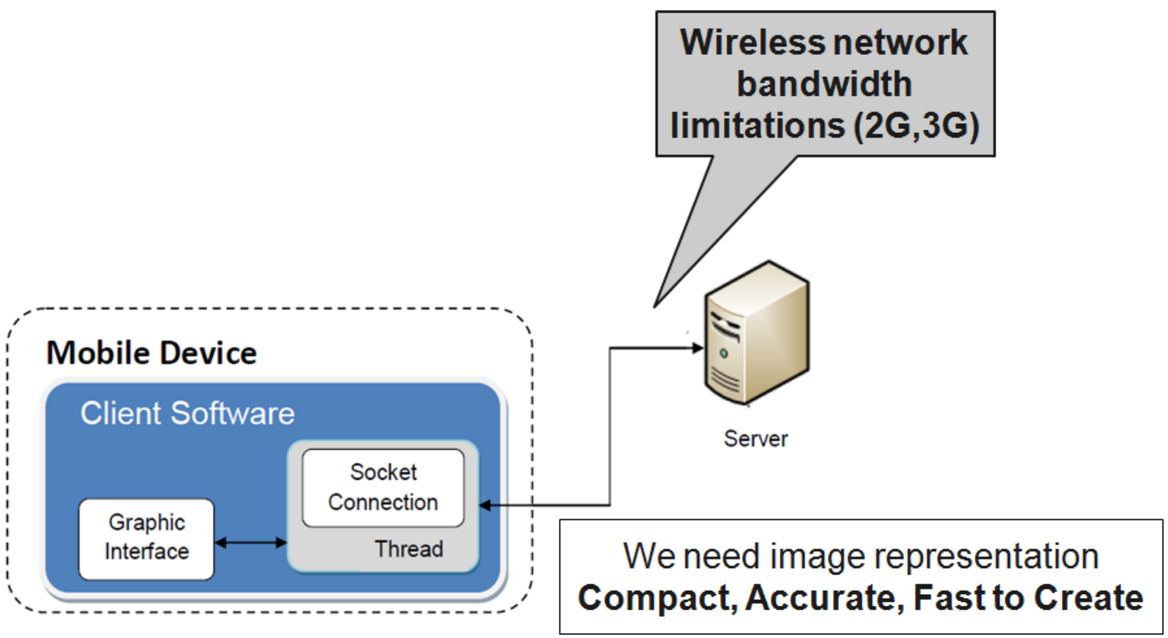
Introduction

- In 2014, the number of smartphone users worldwide achieved around 1.75 billion [Cisco, 2015]
- Several challenges/opportunities in image/video processing on mobile devices emerged: Annotation, categorization, retrieval
- **Problem:** Limited battery life and wireless network bandwidth limitations
- **We need:** compact representations of image which are fast to compute, accurate to retrieve similar images

Smartphone users worldwide 1.75 billion

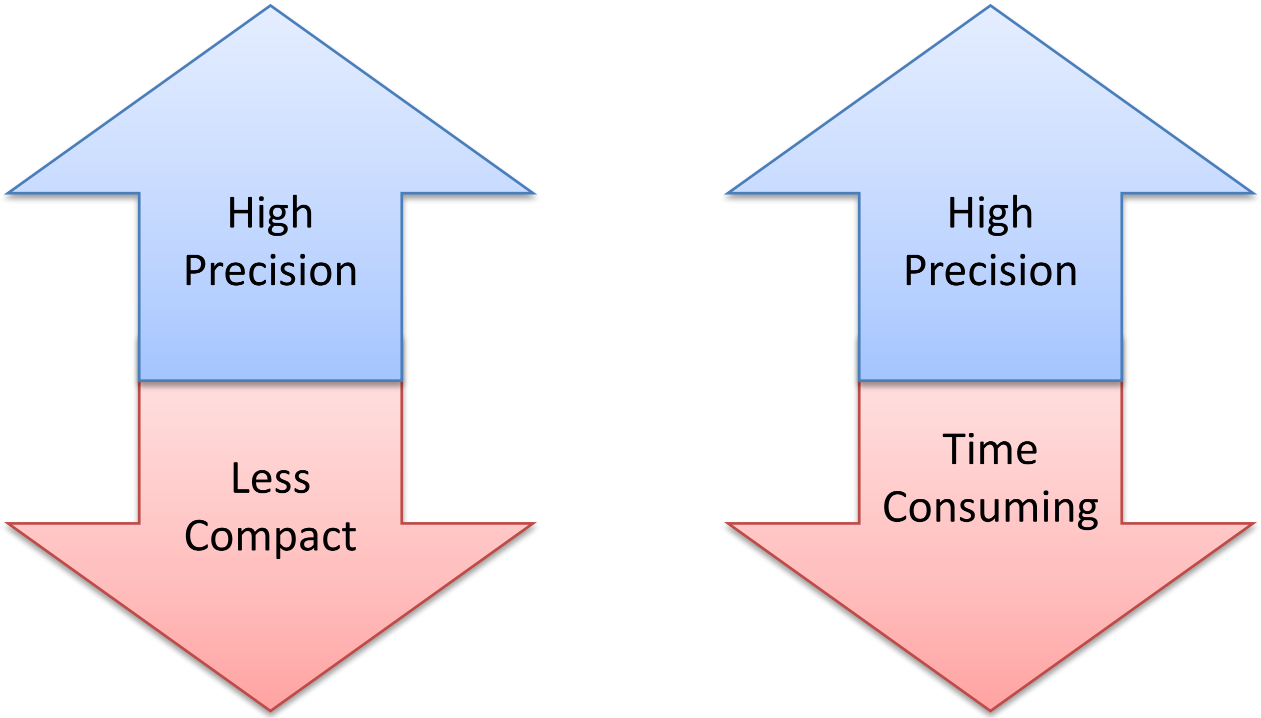
Introduction

- Many works in Mobile Visual Search (MVS)
  - Image Classification and Image Retrieval **(Mobile Image Search)**
- Use client-server architecture
- Client (mobile) sends images or several feature vectors to be processed the server



Introduction

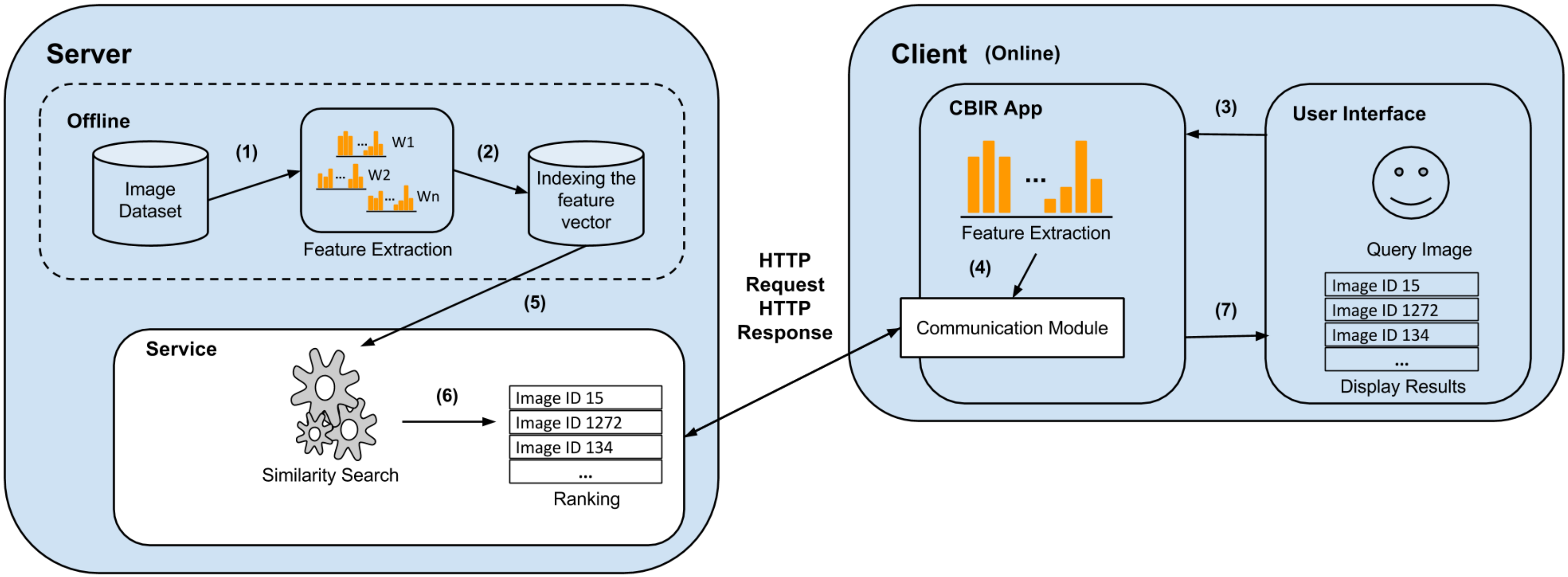
- We need image representation
  - Compact, Accurate, Fast to Create
  - Triple trade-off problem in mobile devices



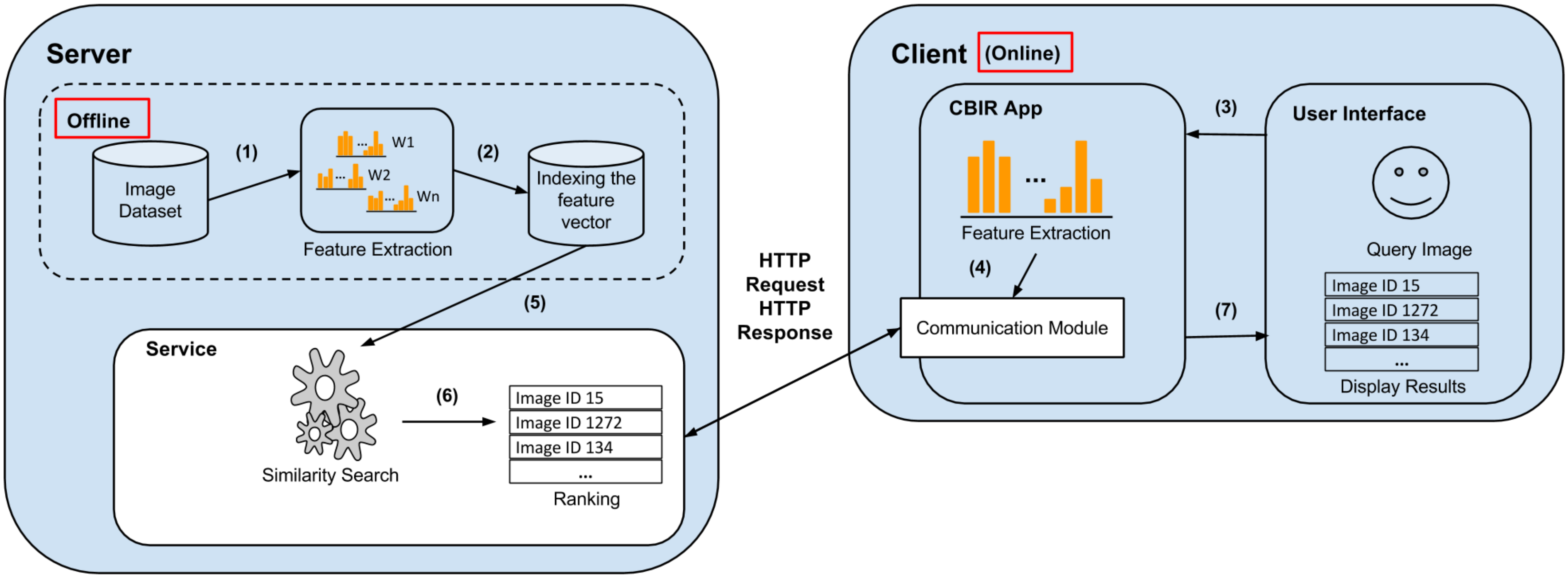
Introduction

- **State-of-the-art: Deep Learning**
- If you run deep learning on mobile devices, using the GPU to adequately characterize the images
  - Time consuming
  - Smartphone battery
  - Sometimes the feature vector will be very large
  - Wireless network bandwidth limitation (2G, 3G)

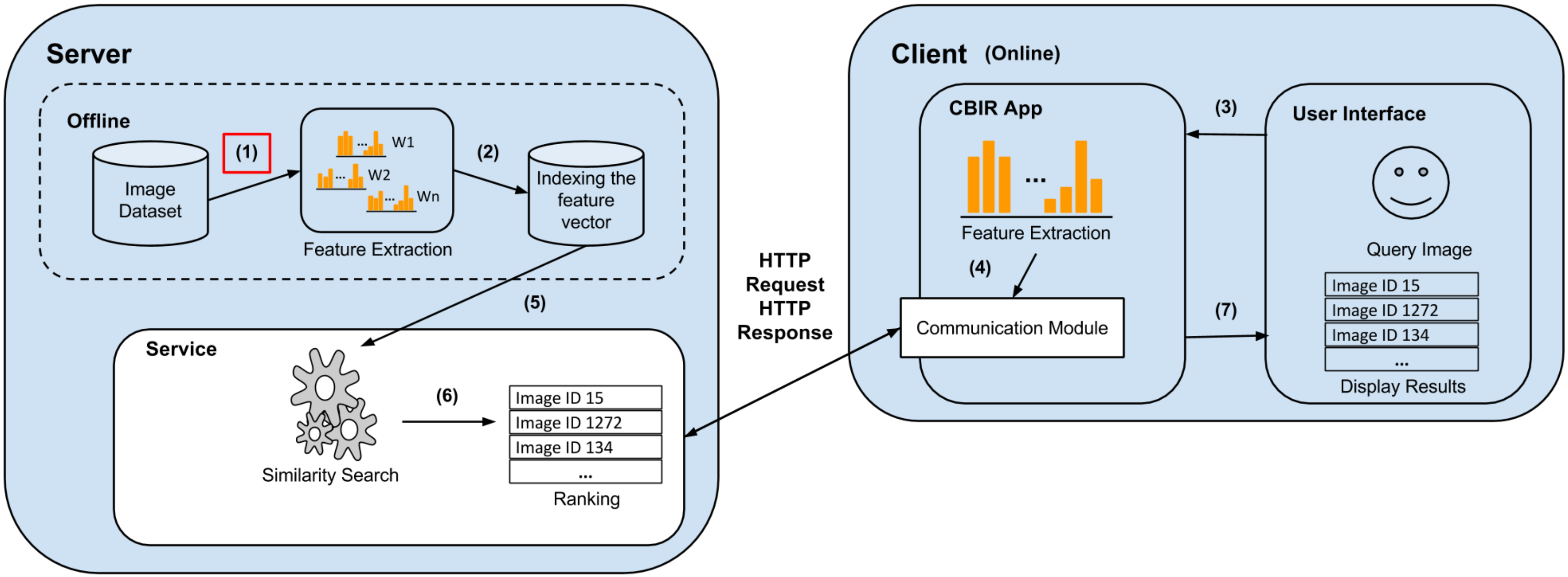
Mobile Image Search



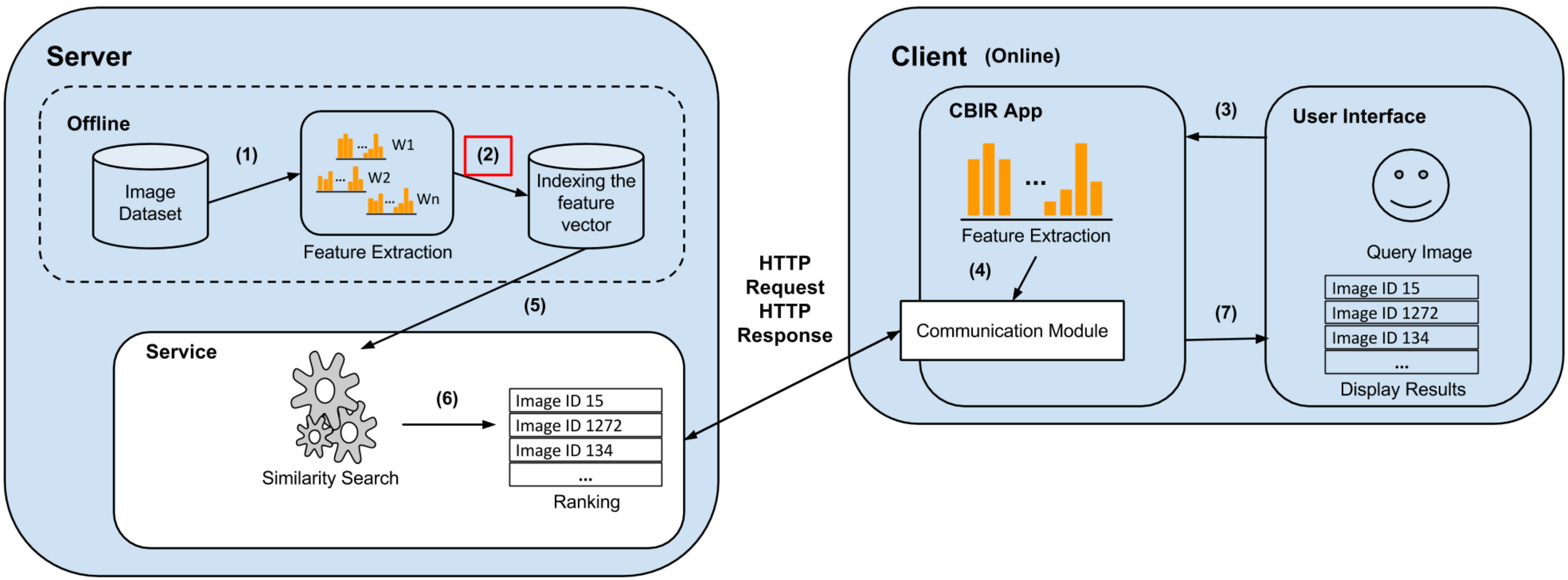
Mobile Image Search



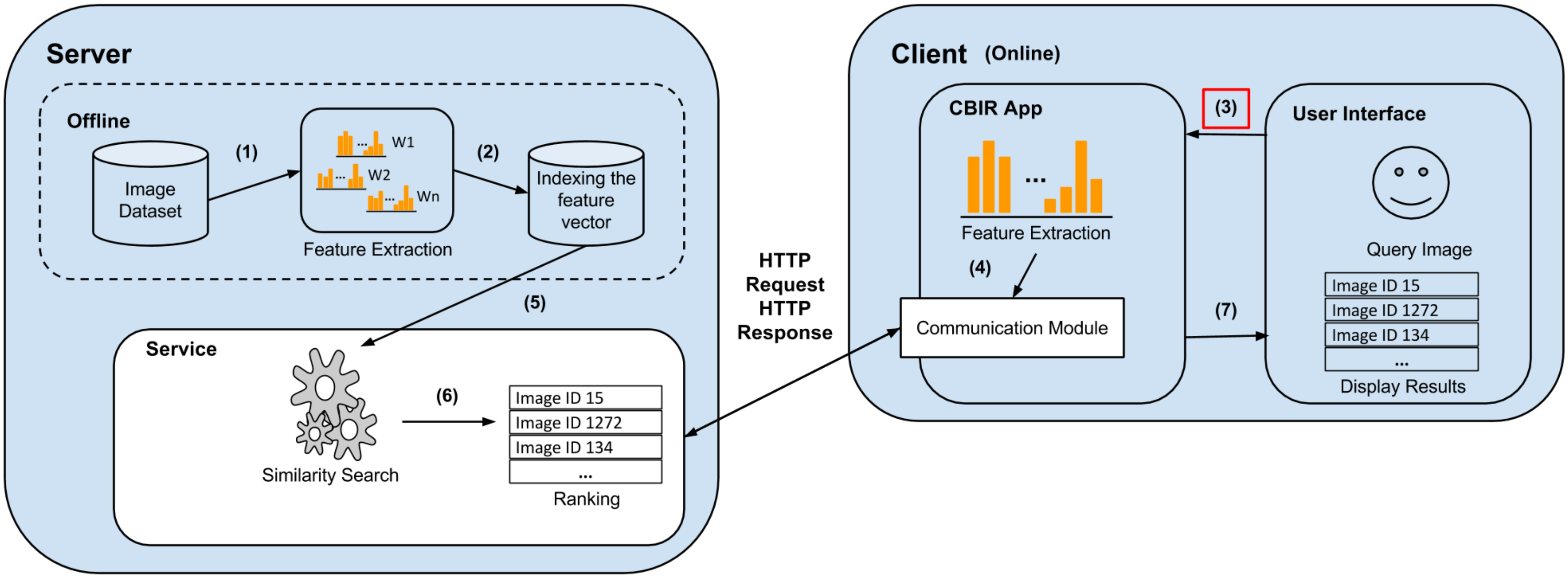
Mobile Image Search



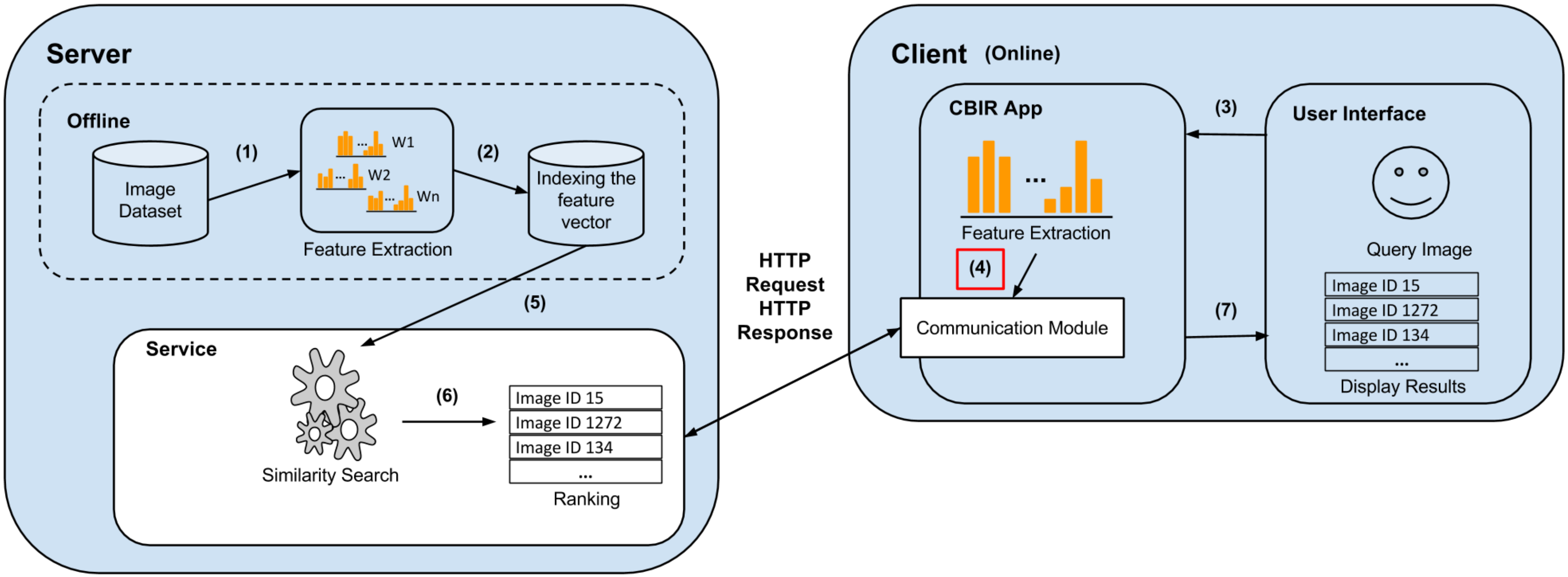
Mobile Image Search



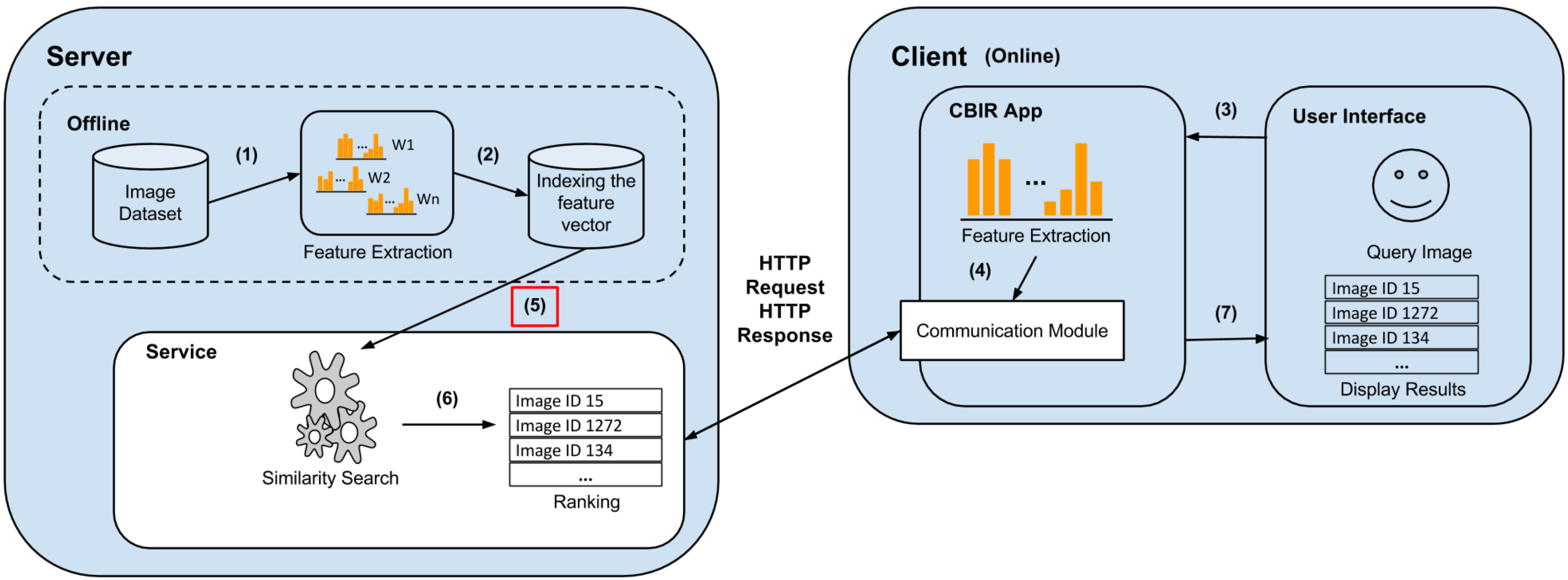
Mobile Image Search



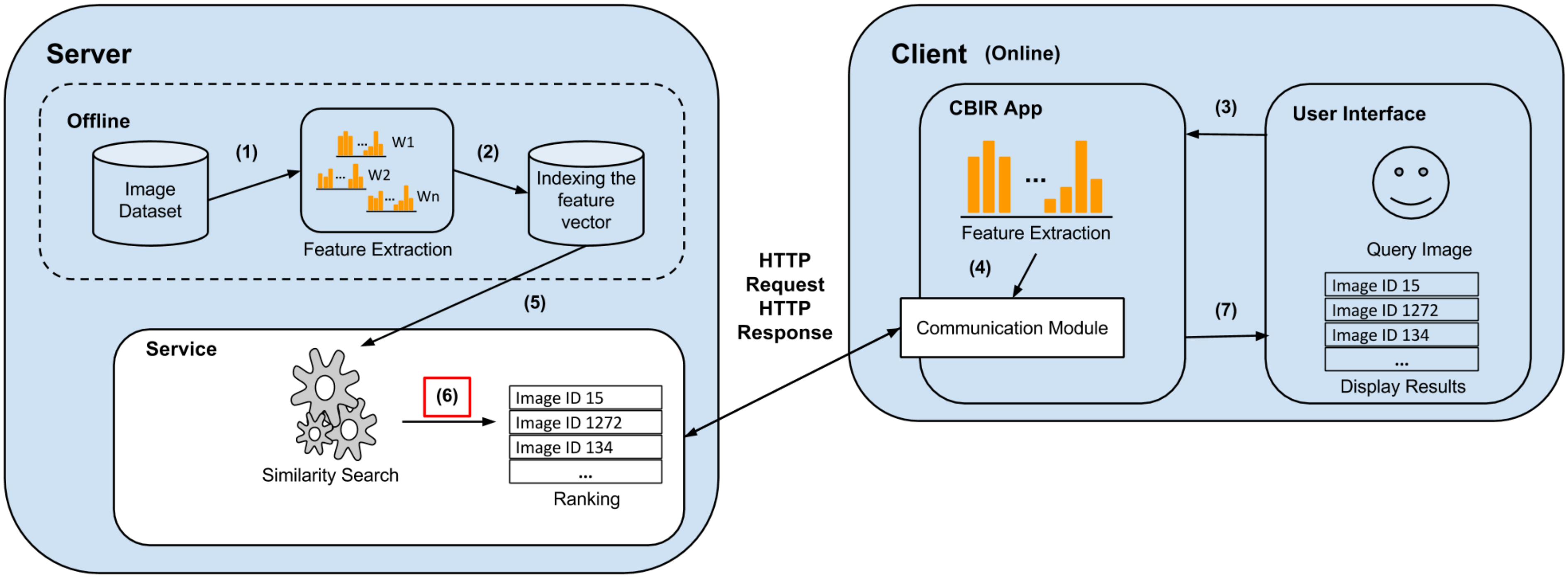
Mobile Image Search



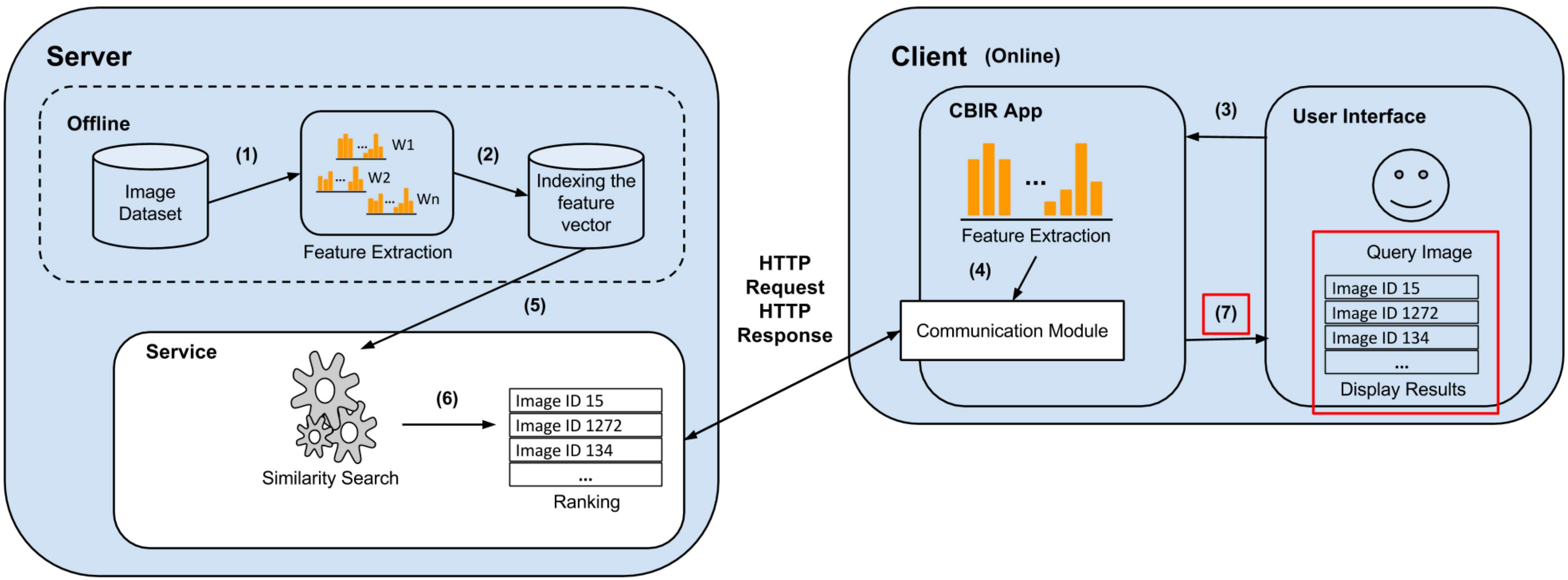
Mobile Image Search



Mobile Image Search



Mobile Image Search



Related Works

- [Kumar and Lu, 2010]
  - Analysis that suggests which cloud computing can potentially save energy for mobile users
- [Girod et al., 2011]
  - Stanford Product Search system
  - Perform feature extraction and compression on the client (mobile), to reduce system latency

Related Work

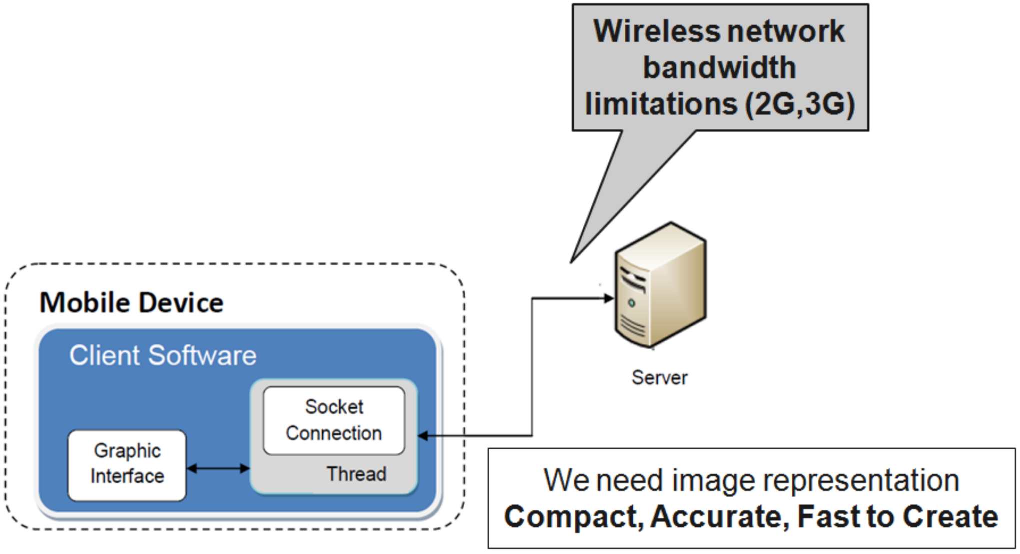
- [Ascenso and Pereira, 2013]
  - Present the lossless compression of binary image feature
  - Lower the energy and bandwidth requirements
- [Zhuang et al., 2014]
  - Propose a novel scheme to quantify the spatial context information and convert it into binary strings

Research Challenges

- Challenges in CBIR
  - Finding a relevant visual content representation
  - Making retrieval scalable to large image datasets
- Challenges in Mobile Image Search
  - Variations in image capturing conditions such as different illumination
  - Limitations of battery and memory usage
  - High network cost incurred by data transmission

Objectives

- Robust and efficient feature extraction algorithms that fit mobile device constraints
- Fast and effective feature vector compression
- Analyze the best representations to represent images in a content-based image retrieval architecture

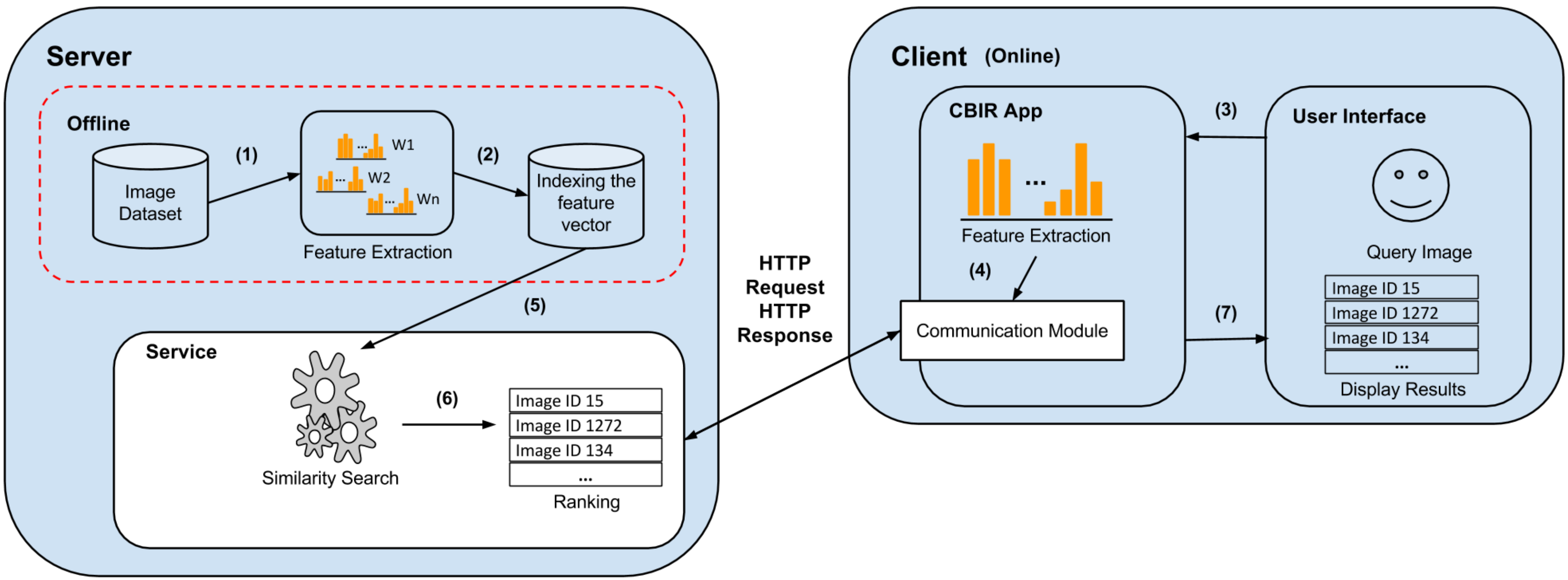


Contributions

1. A comparative study of Low-Cost Representation for Mobile Image Search
   - Binary descriptors using mid-level representation
   - Impact of dense sampling and sparse sampling to compute descriptors using bags of words strategies
   - Global descriptors (color, texture, and shape)
   - Image features compression techniques

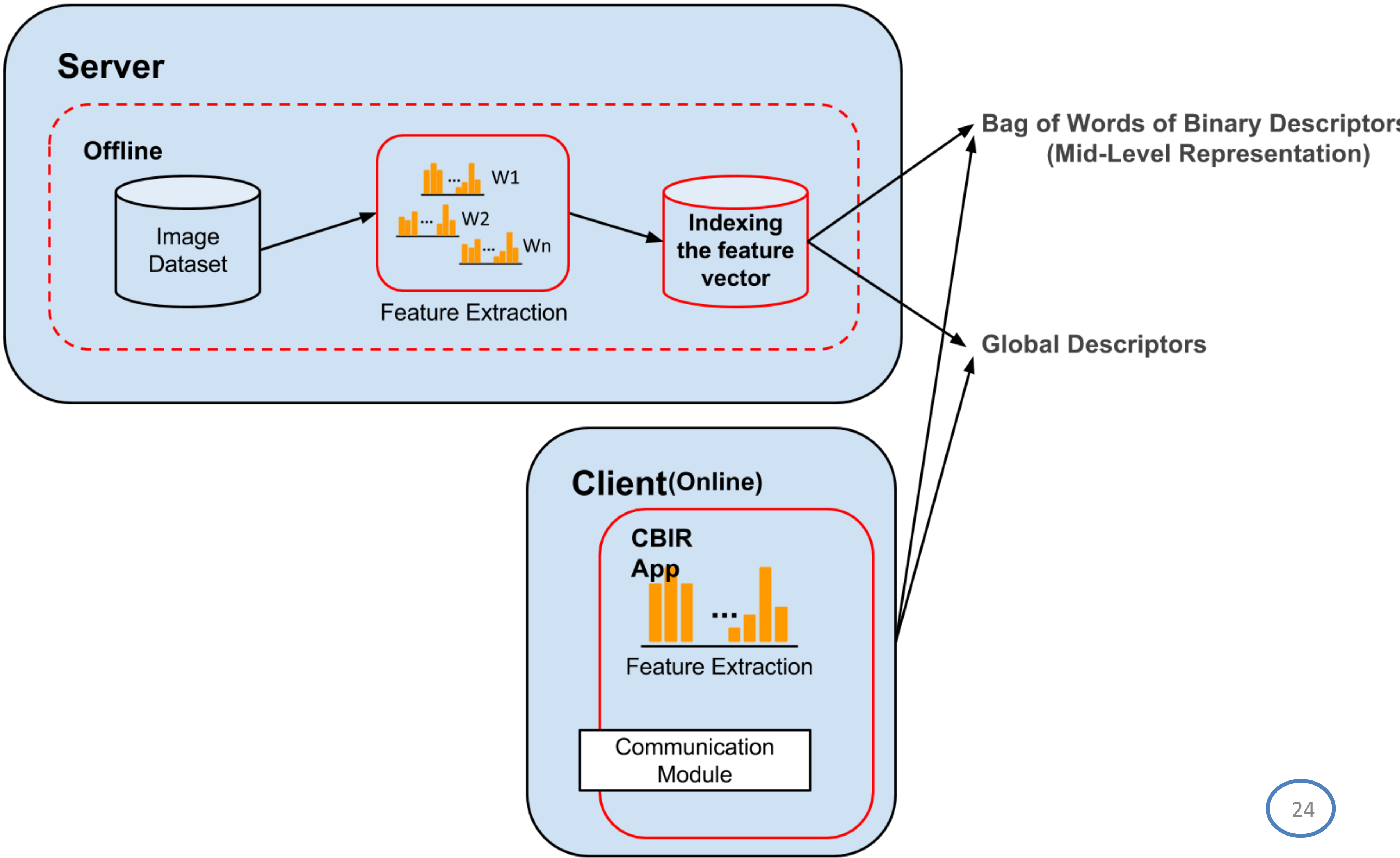
2. We propose two new bag of visual words representations that include spatial information
   - Improve the quality of image representation on mobile devices
   - BOBGraph (Spatial Bag of BIC graph)
   - BOBSlic (Spatial Bag of Slic BIC)

Background

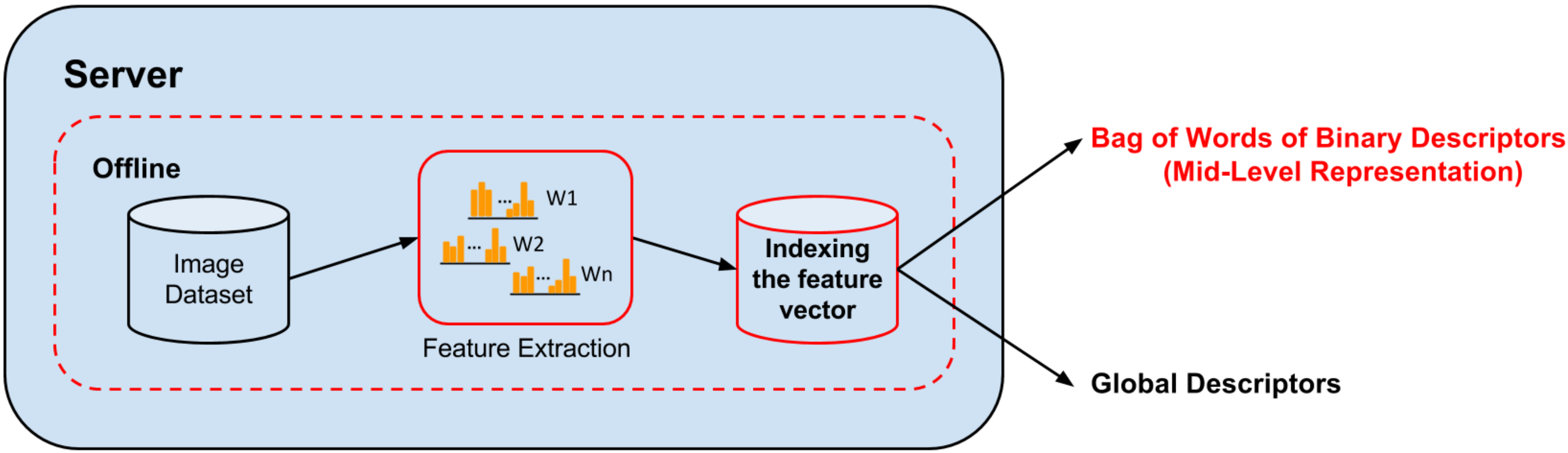


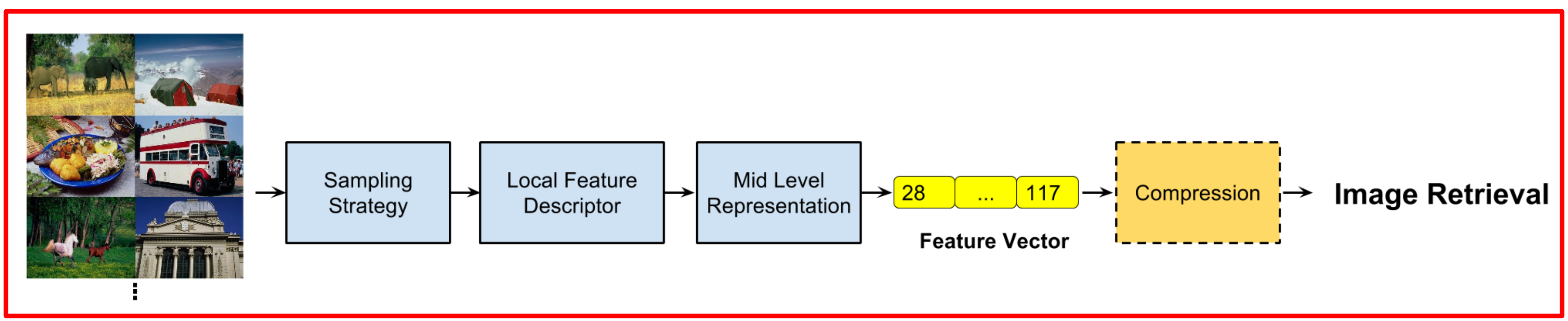
Server
Offline
Image Dataset
(1)
W1
W2
Wn
Feature Extraction
(2)
Indexing the feature vector
(5)
Service
Similarity Search
(6)
Image ID 15
Image ID 1272
Image ID 134
...
Ranking
HTTP Request
HTTP Response
Client (Online)
CBIR App
...
Feature Extraction
(4)
Communication Module
(3)
(7)
User Interface
Query Image
Image ID 15
Image ID 1272
Image ID 134
...
Display Results

Background: Feature Extraction

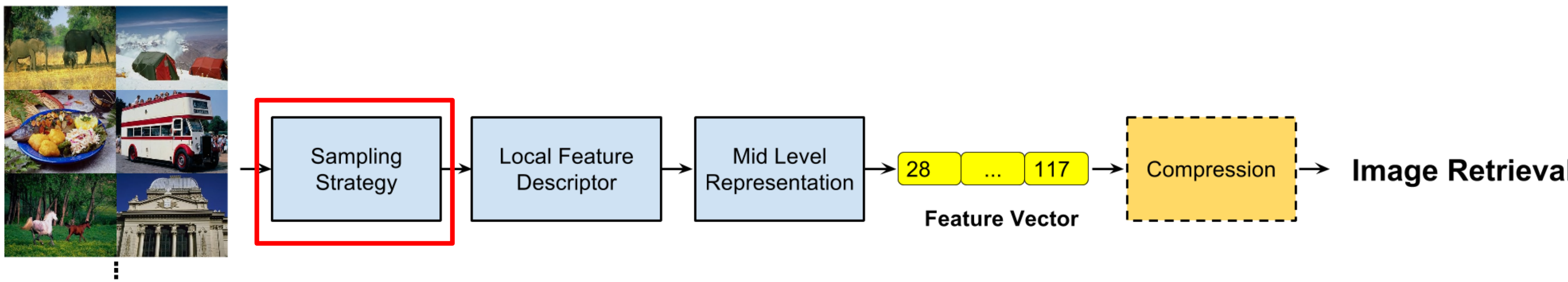


Background: Mid-Level Representations

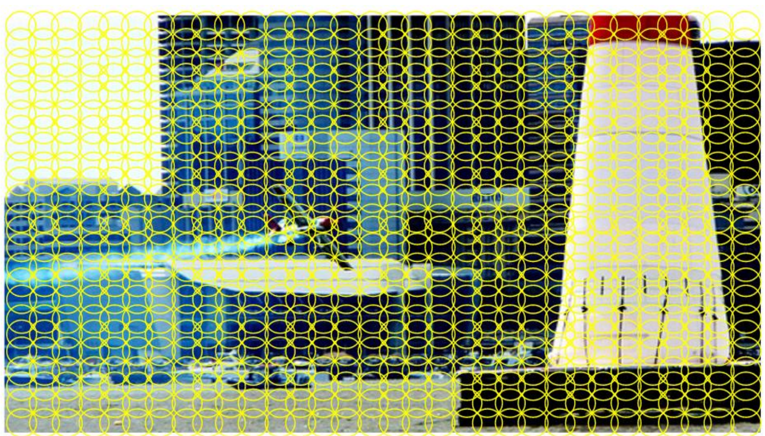




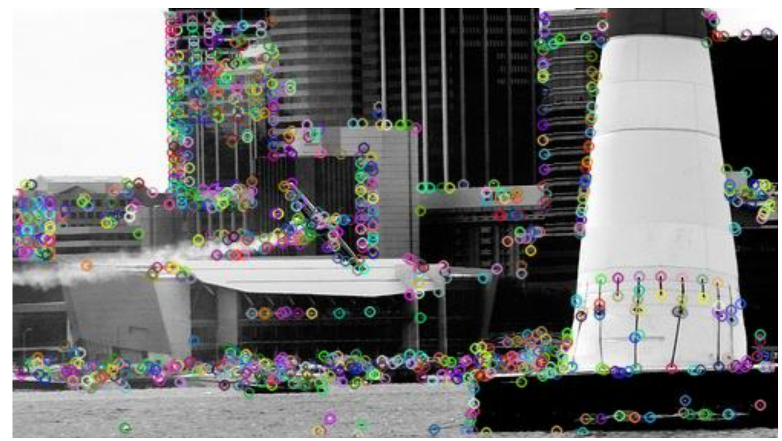
Background: Mid-Level Representations



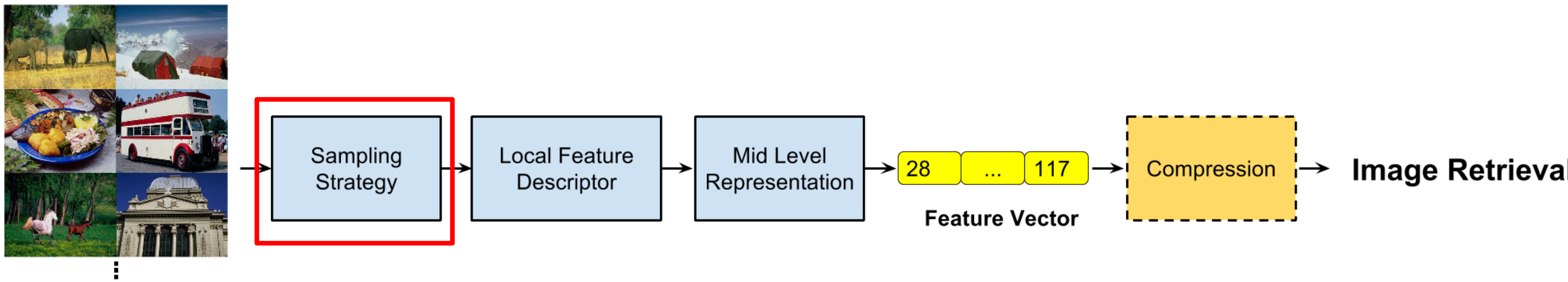
Dense Sampling



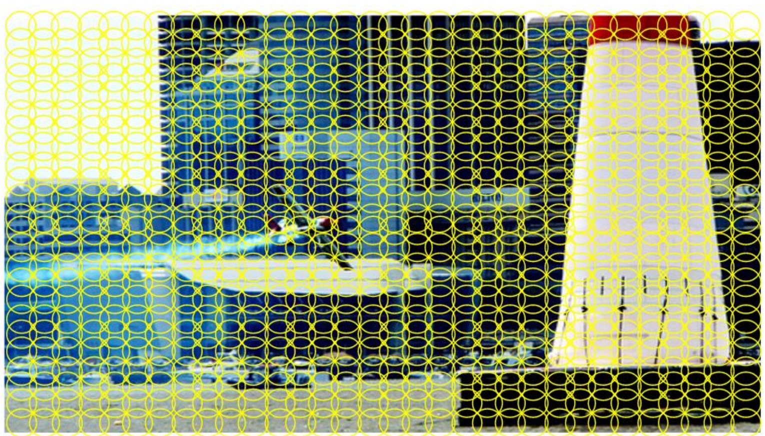
Sparse Sampling



Background: Mid-Level Representations

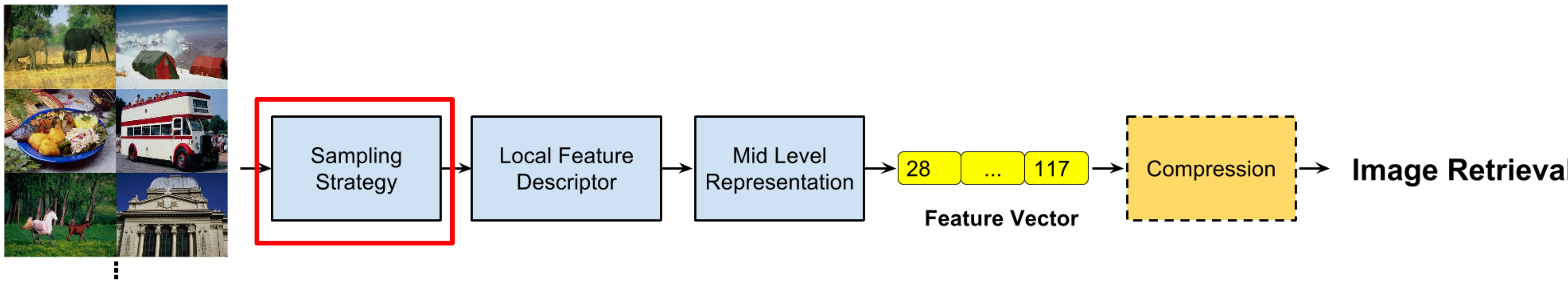


Dense Sampling

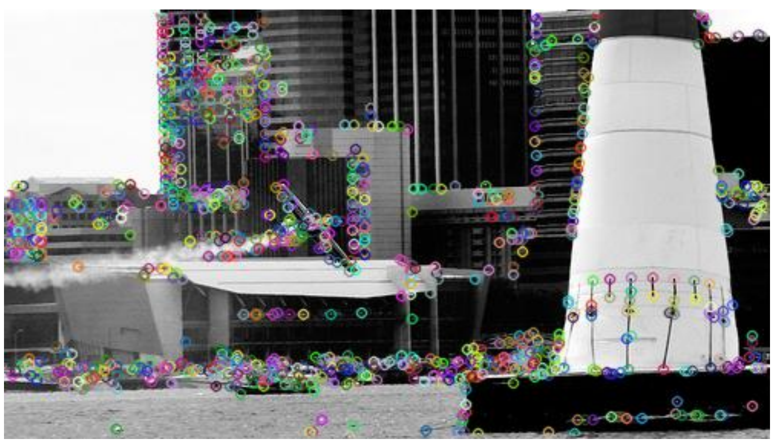


Dense Sampling 6x6 pixels

Background: Mid-Level Representations

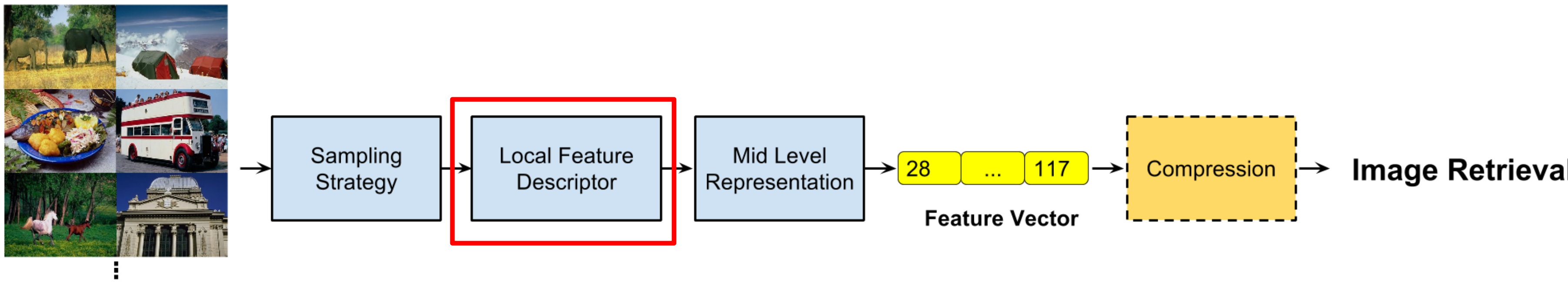


Sparse Sampling



1. **FAST:** Features from Accelerated Segment Test
2. **GFTT:** Good Features to Track
3. **GFTT Harris:** Good Features to Track with Harris
4. **MSER:** Maximally Stable Extremal Regions
5. **ORB Detector:** Oriented FAST and Rotated BRIEF
6. **SURF Detector:** Speeded-Up Robust Features

Background: Mid-Level Representations



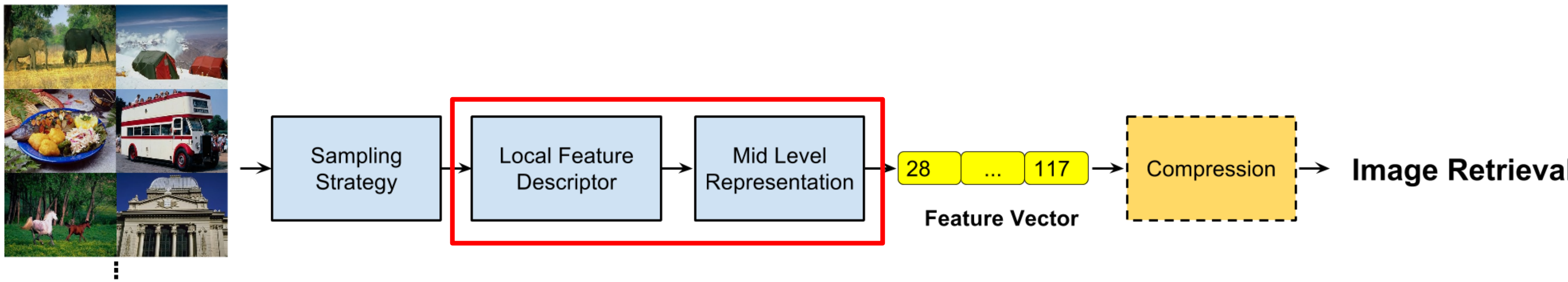
Binary Descriptors

1. **BRIEF:** Binary Robust Independent Elementary Features
2. **ORB:** Oriented FAST and Rotated BRIEF
3. **BRISK:** Binary Robust Invariant Scalable Keypoints
4. **FREAK:** Fast Retina Keypoint
5. **BinBoost:** Boosting Binary keypoint descriptors

Non-Binary Descriptors

1. **SIFT:** Scale Invariant Feature Transform
2. **SURF:** Speeded Up Robust Feature

Background: Mid-Level Representations



Binary Descriptors

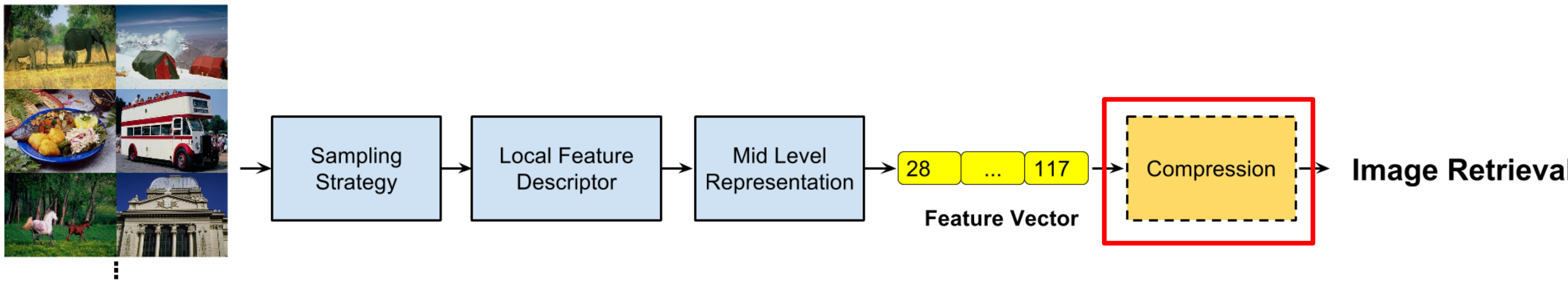
1. **BRIEF:** Binary Robust Independent Elementary Features
2. **ORB:** Oriented FAST and Rotated BRIEF
3. **BRISK:** Binary Robust Invariant Scalable Keypoints
4. **FREAK:** Fast Retina Keypoint
5. **BinBoost:** Boosting Binary keypoint descriptors

Low-complexity binary descriptors

(1) Time required for extracting

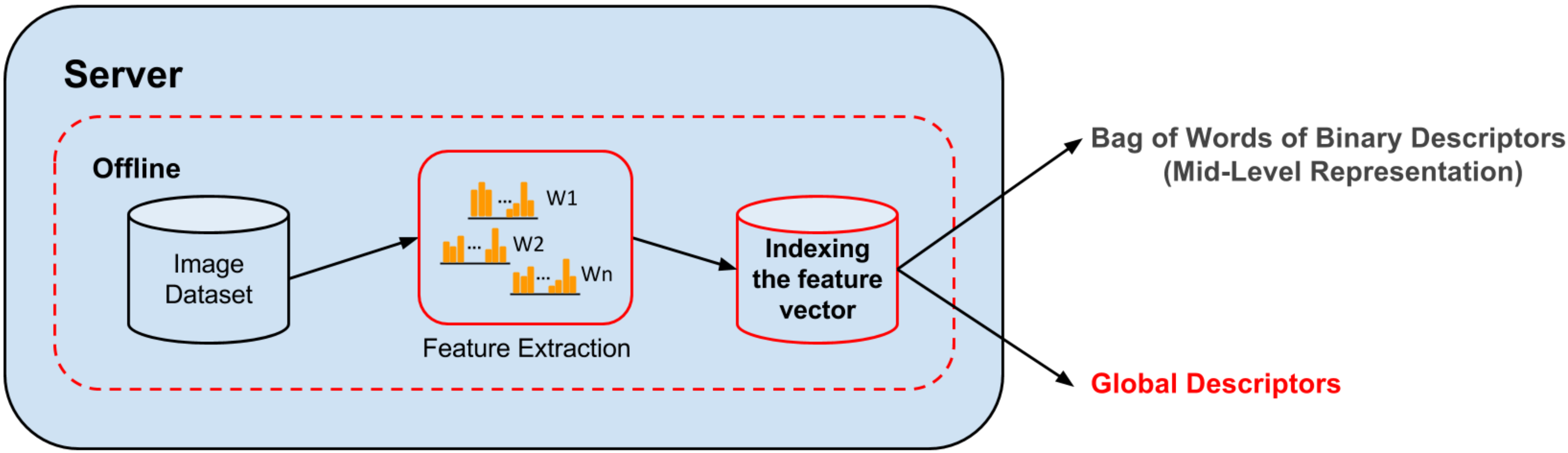
(2) Small size of extracted feature vector

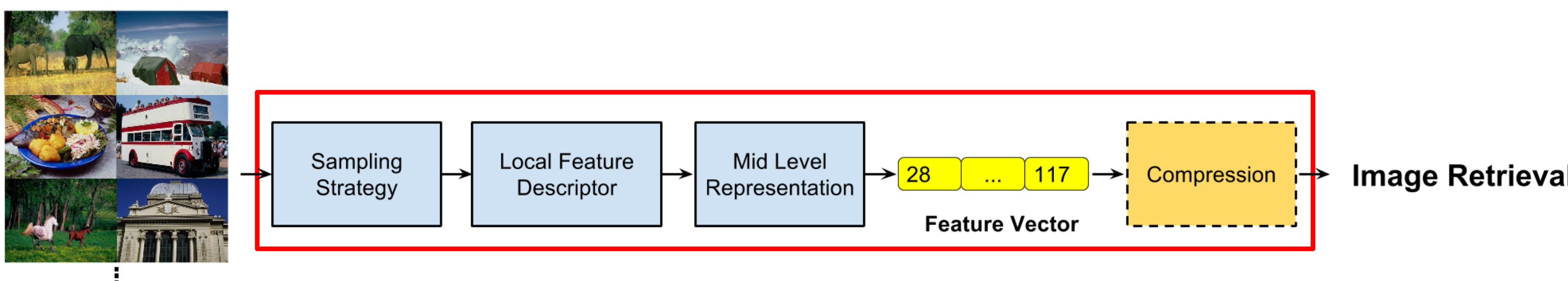
Background: Mid-Level Representations



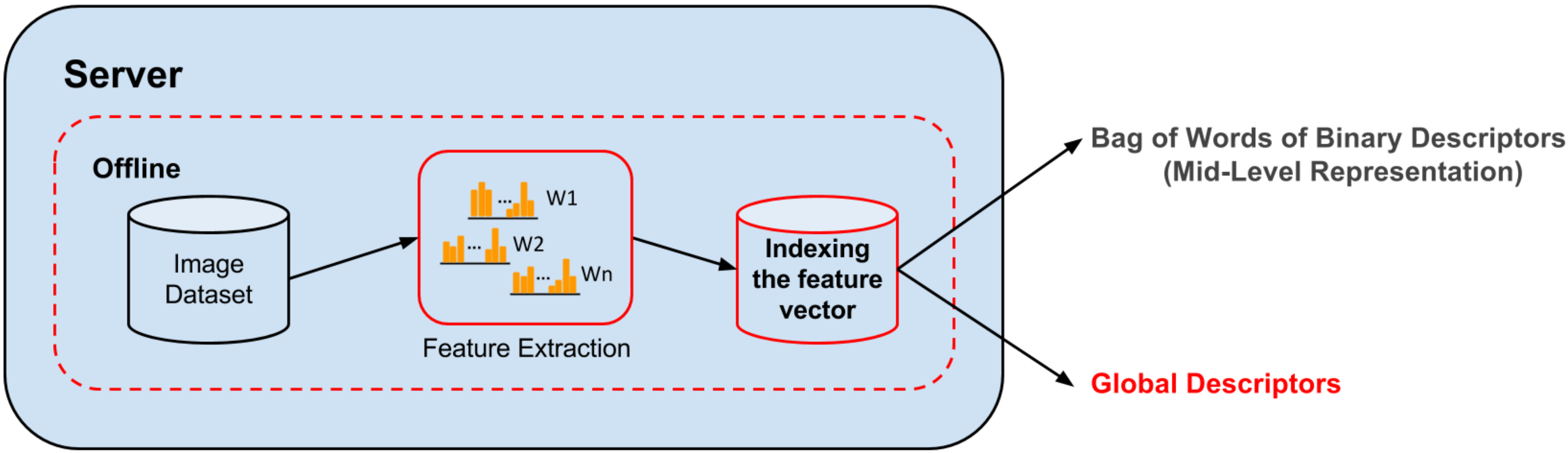
- Lossless compression
  - Huffman encoding
  - Error Enconding
  - Run-length encoding
- Lossy compression
  - Soft-MAX Truncated
    - **[10.2, 3.45, 11.89]** to **[10, 3, 11]**
    - Other solution: Normalize and kept the exponent of the number
  - Soft-MAX using Ranges

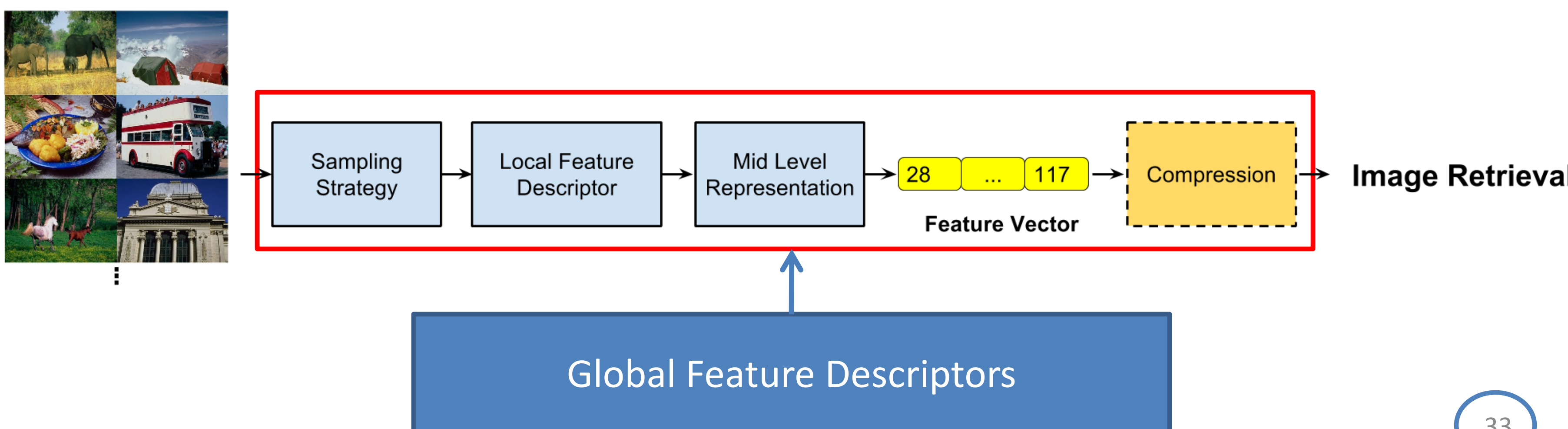
Background: Global Feature Descriptors



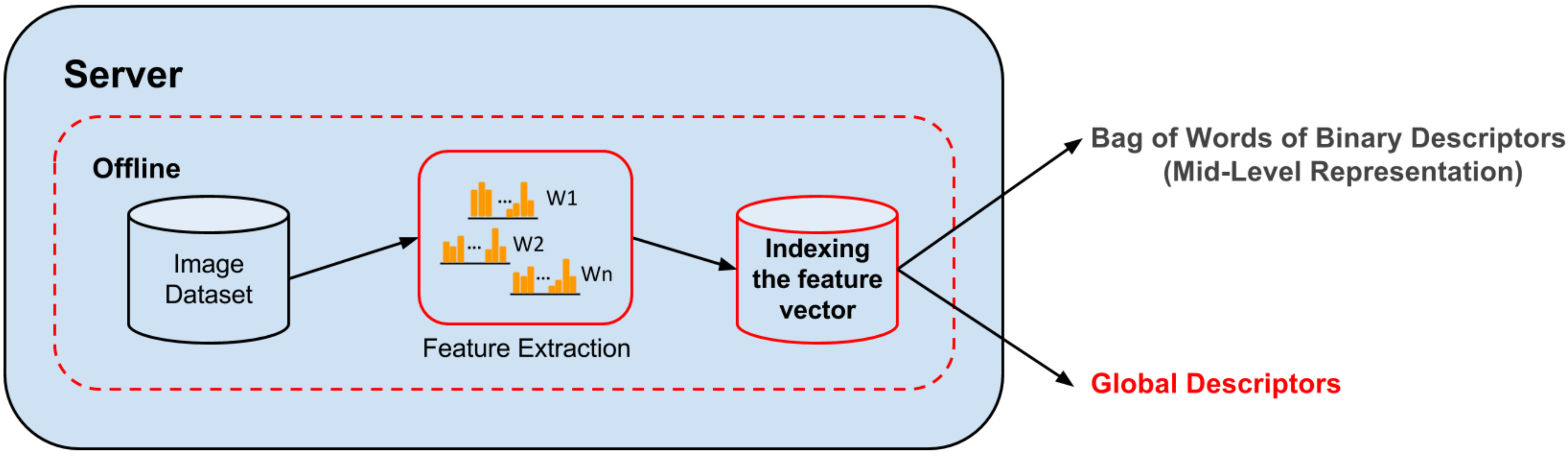


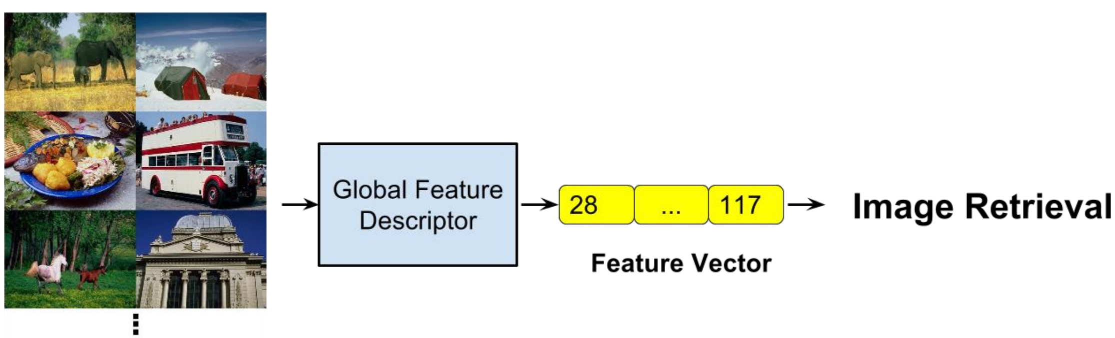
Background: Global Feature Descriptors



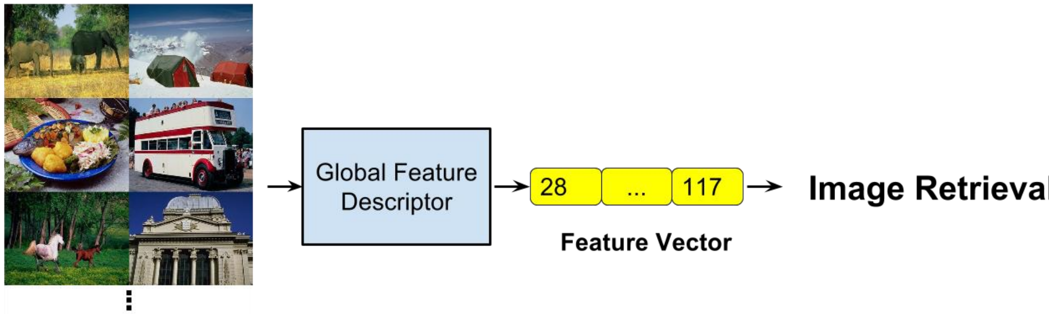


Background: Global Feature Descriptors



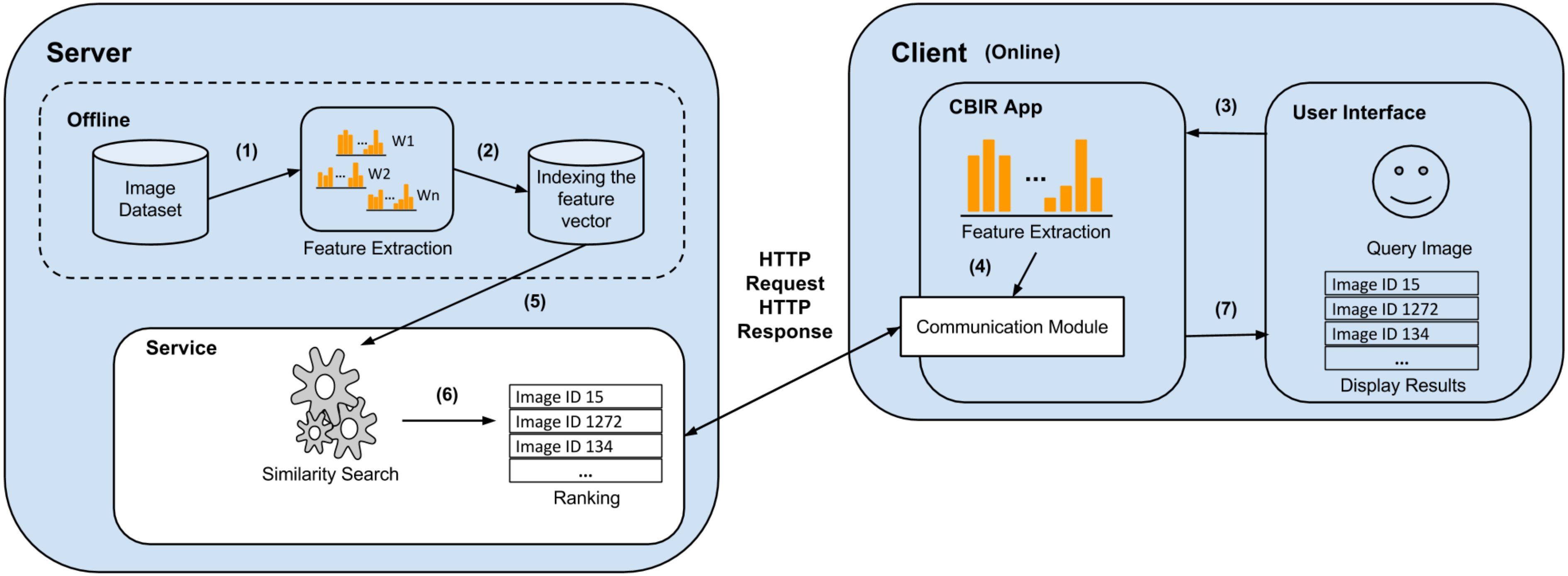


Background: Global Feature Descriptors

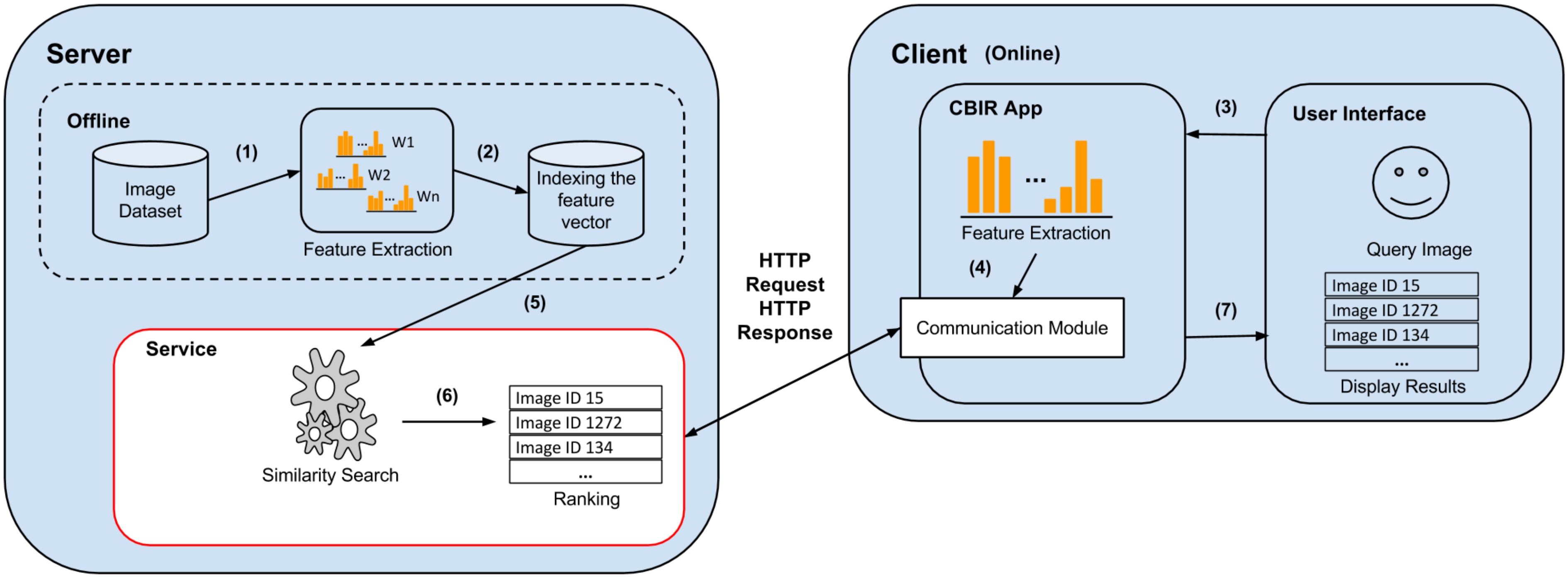


Color (10)	Vector Size	Texture (5)	Vector Size	Shape (2)	Vector Size
ACC	256	LAS	256	EOAC	288
BIC	128	LBP	10	SPYTEC	16
CGCH	64	QCCH	40		
ColorBitmap	306	SASI	64		
CSD	64	UNSER	32		
CWHSV	64				
CWLUV	127				
GCH	64				
JAC	256				
LCH	64				

Background: Similarity Search



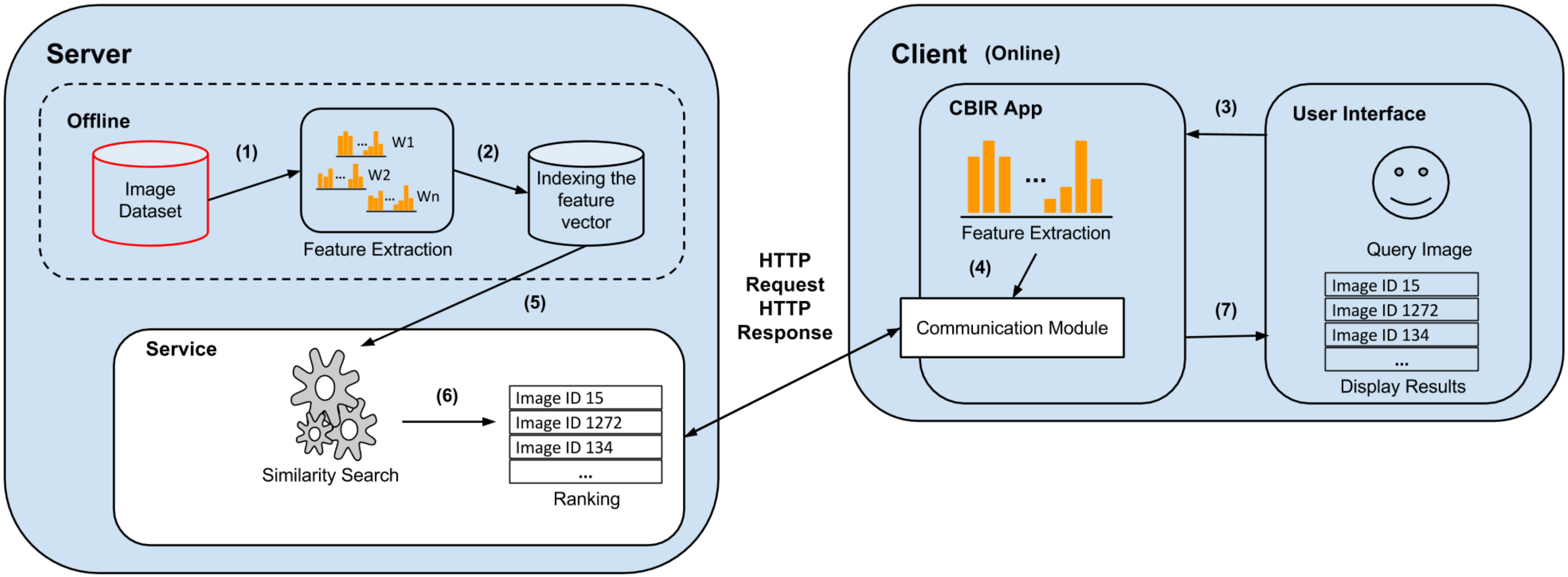
Background: Similarity Search



We use Euclidean distance and Manhattan distance

We statistically showed that **Manhattan** was more accurate than **Euclidean**

Benchmark Datasets



We use 10 datasets

Benchmark Datasets

Dataset	N. Images	N. Classes	Measure	Multi-Label
1: 15Scences	4,485	15	MAP, P@10	No
2: OxBuild11	567	11	MAP, P@10	No
3: ParisLandmarks	6,392	12	MAP, P@10	No
4: ZuBuD	1,005	201	MAP, P@5	No
5: SMVS692	3,460	692	MAP, P@5	No
6: catech101	9,144	102 (with Backgroud class)	MAP, P@10	No
7: caltech256	30,607	257 (with Cluster class)	MAP, P@10	No
8: WANG	1,000	10	MAP, P@10	No
9: VOC2007	9,963	20	MAP, P@10	Yes
10: UWdataset	1,109	20	MAP, P@10	Yes

10 Datasets to simulated more real sceneries of Image Retrieval

Benchmark Datasets

Dataset	N. Images	N. Classes	Measure	Multi-Label
1: 15Scences	4,485	15	MAP, P@10	No
2: OxBuild11	567	11	MAP, P@10	No
3: ParisLandmarks	6,392	12	MAP, P@10	No
4: ZuBuD	1,005	201	MAP, P@5	No
5: SMVS692	3,460	692	MAP, P@5	No
6: catech101	9,144	102 (with Backgroud class)	MAP, P@10	No
7: caltech256	30,607	257 (with Cluster class)	MAP, P@10	No
8: WANG	1,000	10	MAP, P@10	No
9: VOC2007	9,963	20	MAP, P@10	Yes
10: UWdataset	1,109	20	MAP, P@10	Yes

4 Datasets of Scenes

Benchmark Datasets

Dataset	N. Images	N. Classes	Measure	Multi-Label
1: 15Scences	4,485	15	MAP, P@10	No
2: OxBuild11	567	11	MAP, P@10	No
3: ParisLandmarks	6,392	12	MAP, P@10	No
4: ZuBuD	1,005	201	MAP, P@5	No
5: SMVS692	3,460	692	MAP, P@5	No
6: catech101	9,144	102 (with Backgroud class)	MAP, P@10	No
7: caltech256	30,607	257 (with Cluster class)	MAP, P@10	No
8: WANG	1,000	10	MAP, P@10	No
9: VOC2007	9,963	20	MAP, P@10	Yes
10: UWdataset	1,109	20	MAP, P@10	Yes

1 Dataset used for Mobile Visual Search

Benchmark Datasets

Dataset	N. Images	N. Classes	Measure	Multi-Label
1: 15Scences	4,485	15	MAP, P@10	No
2: OxBuild11	567	11	MAP, P@10	No
3: ParisLandmarks	6,392	12	MAP, P@10	No
4: ZuBuD	1,005	201	MAP, P@5	No
5: SMVS692	3,460	692	MAP, P@5	No
6: catech101	9,144	102 (with Backgroud class)	MAP, P@10	No
7: caltech256	30,607	257 (with Cluster class)	MAP, P@10	No
8: WANG	1,000	10	MAP, P@10	No
9: VOC2007	9,963	20	MAP, P@10	Yes
10: UWdataset	1,109	20	MAP, P@10	Yes

3 Datasets using Single Label

Benchmark Datasets

Dataset	N. Images	N. Classes	Measure	Multi-Label
1: 15Scences	4,485	15	MAP, P@10	No
2: OxBuild11	567	11	MAP, P@10	No
3: ParisLandmarks	6,392	12	MAP, P@10	No
4: ZuBuD	1,005	201	MAP, P@5	No
5: SMVS692	3,460	692	MAP, P@5	No
6: catech101	9,144	102 (with Backgroud class)	MAP, P@10	No
7: caltech256	30,607	257 (with Cluster class)	MAP, P@10	No
8: WANG	1,000	10	MAP, P@10	No
9: VOC2007	9,963	20	MAP, P@10	Yes
10: UWdataset	1,109	20	MAP, P@10	Yes

2 Datasets using Multi-Label

Benchmark Datasets

Dataset	N. Images	N. Classes	Measure	Multi-Label
1: 15Scences	4,485	15	MAP, P@10	No
2: OxBuild11	567	11	MAP, P@10	No
3: ParisLandmarks	6,392	12	MAP, P@10	No
4: ZuBuD	1,005	201	MAP, P@5	No
5: SMVS692	3,460	692	MAP, P@5	No
6: catech101	9,144	102 (with Backgroud class)	MAP, P@10	No
7: caltech256	30,607	257 (with Cluster class)	MAP, P@10	No
8: WANG	1,000	10	MAP, P@10	No
9: VOC2007	9,963	20	MAP, P@10	Yes
10: UWdataset	1,109	20	MAP, P@10	Yes

Evaluated using MAP, P@10

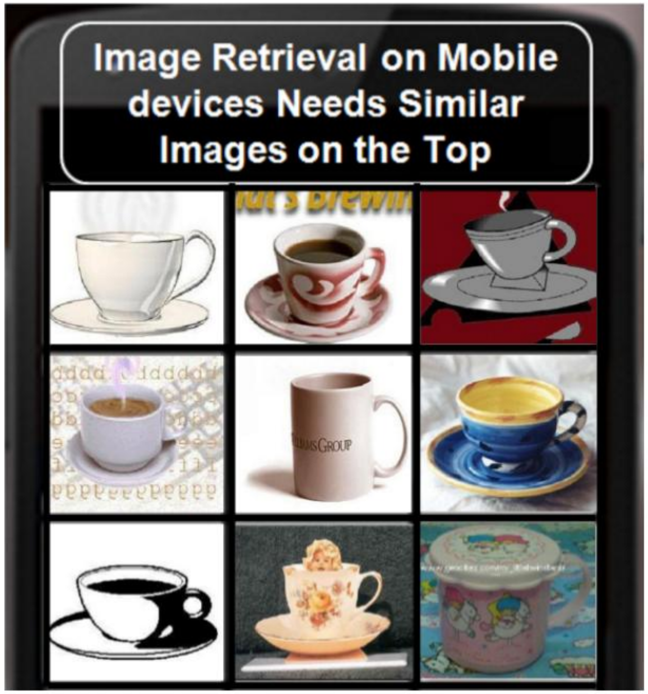
Benchmark Datasets

Dataset	N. Images	N. Classes	Measure	Multi-Label
1: 15Scences	4,485	15	MAP, P@10	No
2: OxBuild11	567	11	MAP, P@10	No
3: ParisLandmarks	6,392	12	MAP, P@10	No
4: ZuBuD	1,005	201	MAP, P@5	No
5: SMVS692	3,460	692	MAP, P@5	No
6: catech101	9,144	102 (with Backgroud class)	MAP, P@10	No
7: caltech256	30,607	257 (with Cluster class)	MAP, P@10	No
8: WANG	1,000	10	MAP, P@10	No
9: VOC2007	9,963	20	MAP, P@10	Yes
10: UWdataset	1,109	20	MAP, P@10	Yes

Evaluated using MAP, P@5

Evaluation metrics

- To evaluate Effectiveness
  - P@10 (precision at top 10), P@5, P@15
  - MAP (Mean Average Precision)
- To evaluate Efficiency
  - Time of computing (seconds)
    - feature extraction and representation
- To evaluate Compactness
  - Compression Ratio (CR)



Contributions

1. A comparative study of Low-Cost Representation for Mobile Image Search

2. We propose two new bag of visual words representations that include spatial information

Low-Cost Representation for Mobile Image Search

- We concentrate our efforts in four main fronts
  1. Binary low-level descriptor selection
  2. Mid-level representation
  3. Low-level global representation analysis
  4. Feasibility analysis of data compression techniques

- We deal with the feature extraction **triple trade-off problem**
  - Efficiency Evaluation (Fast?)
  - Effectiveness Evaluation (Accurate?)
  - Compactness Evaluation (Compact ?)

Low-Cost Representation for Mobile Image Search

- Experiments
  1. Mid-level Representation Analysis (Experiment 1)
     - Efficiency Evaluation (Fast?)
     - Effectiveness Evaluation (Accurate?)
     - Compactness Evaluation (Compact ?)
  2. Low-level Global Representation Analysis (Experiment 2)
     - Efficiency Evaluation (Fast?)
     - Effectiveness Evaluation (Accurate?)
     - Compactness Evaluation (Compact ?)

All Results Analyzed Statistically with 95% confidence

Low-Cost Representation for Mobile Image Search

- Discussion
  - More suitable global descriptors
  - More suitable mid-level representation
  - Paired statistical tests (Global Vs Mid-level)
    - **We point out the more suitable descriptor**
      - Efficiency, Effectiveness, Compactness

Low-Cost Representation for Mobile Image Search

- Experiment 1
  1. Mid-level Representation Analysis
     - Efficiency Evaluation (Fast?)
       - Sampling strategies comparison (15Scences and WANG datasets)
       - Time spent for feature extraction (Caltech101 and VOC2007 datasets)
     - Effectiveness Evaluation (Accurate?)
       - Precision: P@10/P@5 and MAP
       - Time X Precision
     - Compactness Evaluation (Compact ?)
       - Size of Images, Binary Features, Mid-level representation
       - Compresion Ratio X Precision
       - Lossless and Lossy Compression

Low-Cost Representation for Mobile Image Search

- Experiment 1
  1. Mid-level Representation Analysis
     - Efficiency Evaluation (Fast?)
       - Sampling strategies comparison (15Scences and WANG datasets)
       - Time spent for feature extraction (Caltech101 and VOC2007 datasets)
     - Effectiveness Evaluation (Accurate?)
       - Precision: P@10/P@5 and MAP
       - Time X Precision
     - Compactness Evaluation (Compact ?)
       - Size of Images, Binary Features, Mid-level representation
       - Compresion Ratio X Precision
       - Lossless and Lossy Compression

Mid-level Representation Efficiency Evaluation (Sampling)

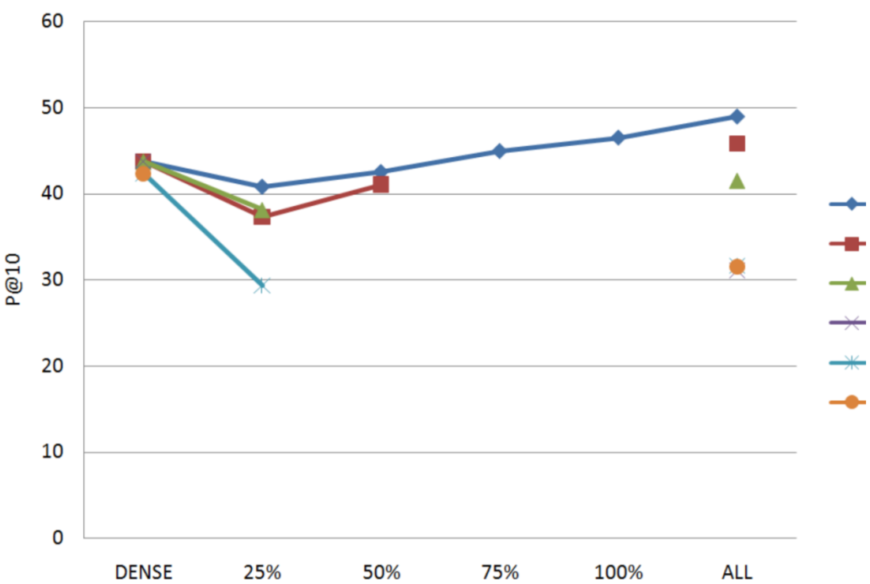
Best combination of binary descriptor with a keypoint detector

FTBKSM	**FAST** + **BRISK** + BoW(SoftMax)	FTBKSM	**FAST** + **BRISK** + BoW(SoftMax)
GTBKSM	**GFTT** + **BRISK** + BoW(SoftMax)	GTBKSM	**GFTT** + **BRISK** + BoW(SoftMax)
HSBKSM	**GFTTHarris** + **BRISK** + BoW(SoftMax)	HSBKSM	**GFTTHarris** + **BRISK** + BoW(SoftMax)
MRBKSM	**MSER** + **BRISK** + BoW(SoftMax)	MROBSM	**MSER** + **ORB**+ BoW(SoftMax)
OBBFSM	**ORB** + **BRIEF** + BoW(SoftMax)	OBBFSM	**ORB** + **BRIEF** + BoW(SoftMax)
SFBFSM	**SURF** + **BRIEF** + BoW(SoftMax)	SFBFSM	**SURF** + **BRIEF** + BoW(SoftMax)

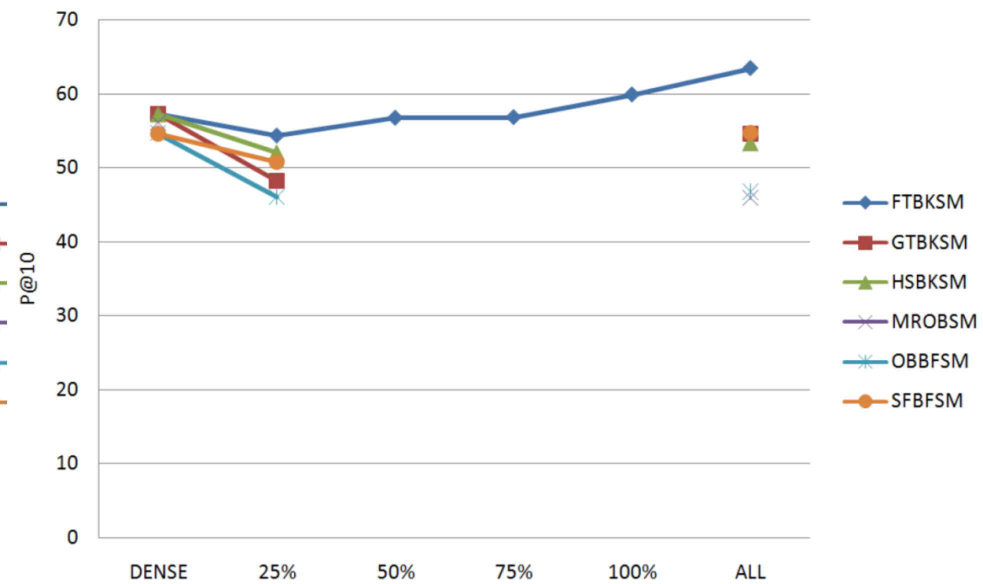
WANG Dataset

15 Scenes Dataset

Mid-level Representation
Efficiency Evaluation (Sampling)



WANG Dataset

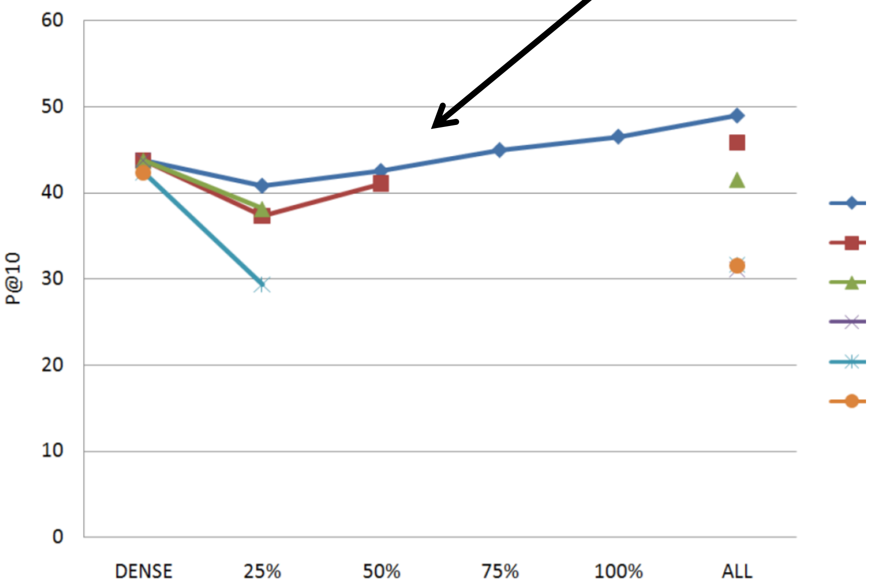


15 Scenes Dataset

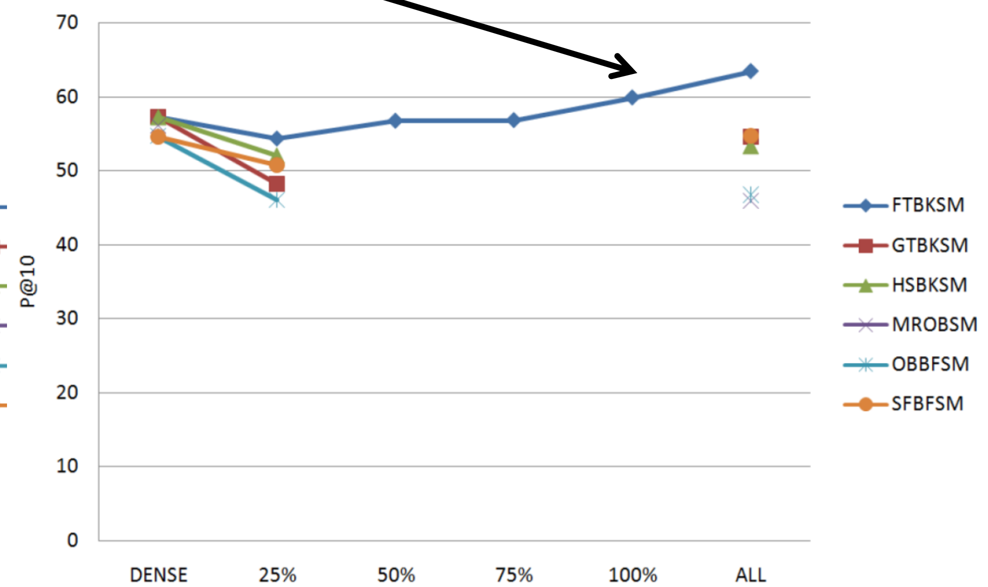
25%, 50%, 75%, 100% = Sparse Sampling using 25% (or 50% or 75% or 100%) of points in Dense Sampling

Mid-level Representation Efficiency Evaluation (Sampling)

FAST + BRISK + BoW(SoftMax)



WANG Dataset

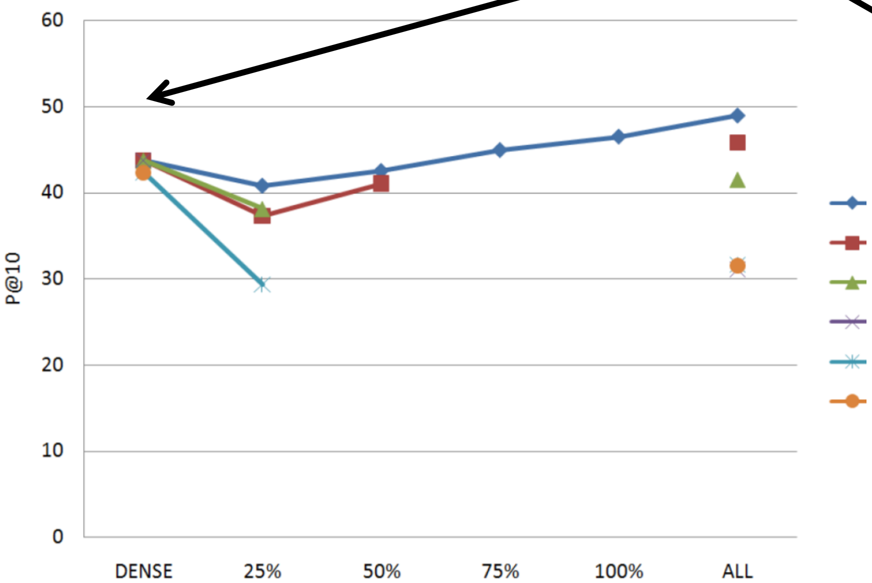


15 Scenes Dataset

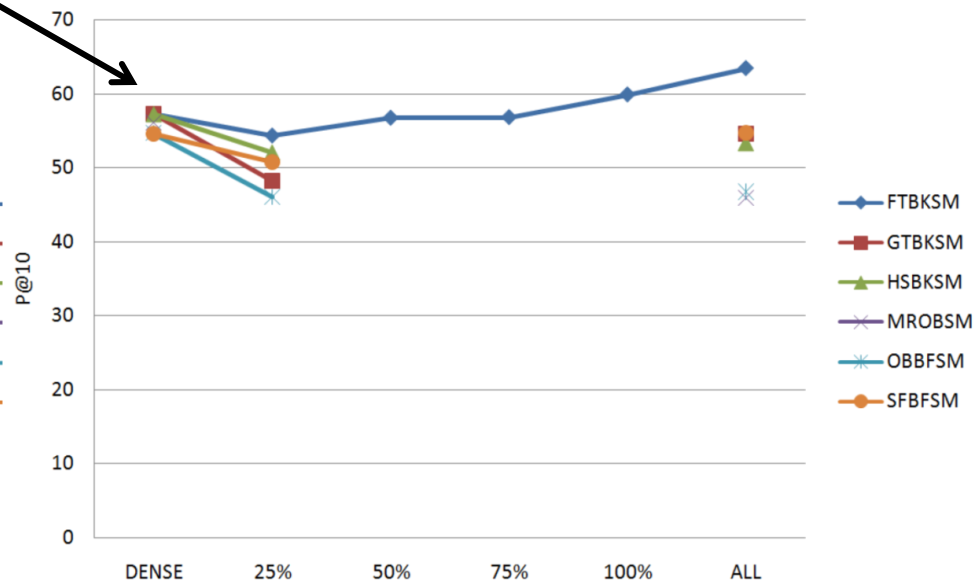
25%, 50%, 75%, 100% = Sparse Sampling using 25% (or 50% or 75% or 100%) of points in Dense Sampling

Mid-level Representation Efficiency Evaluation (Sampling)

DENSE



WANG Dataset

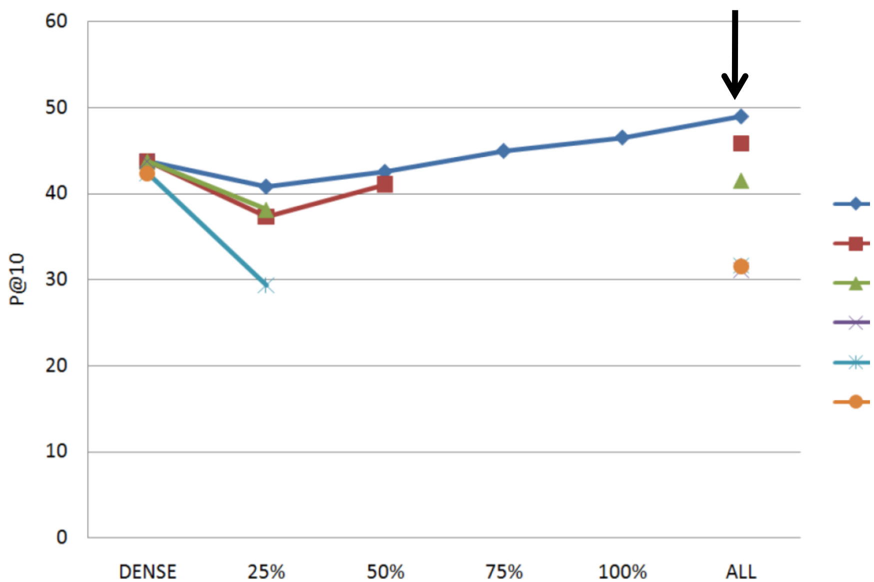


15 Scenes Dataset

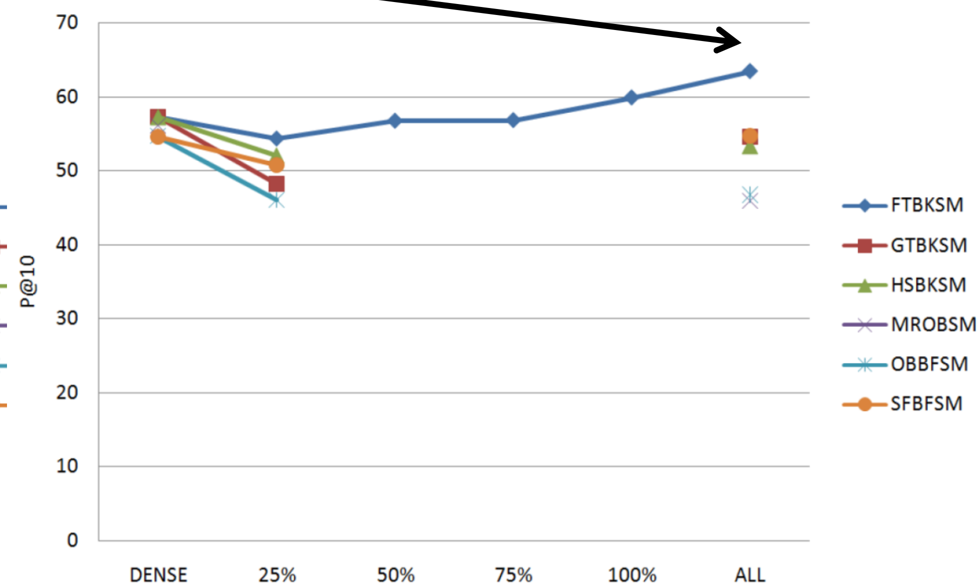
25%, 50%, 75%, 100% = Sparse Sampling using 25% (or 50% or 75% or 100%) of points in Dense Sampling

Mid-level Representation
Efficiency Evaluation (Sampling)

All Keypoints detected



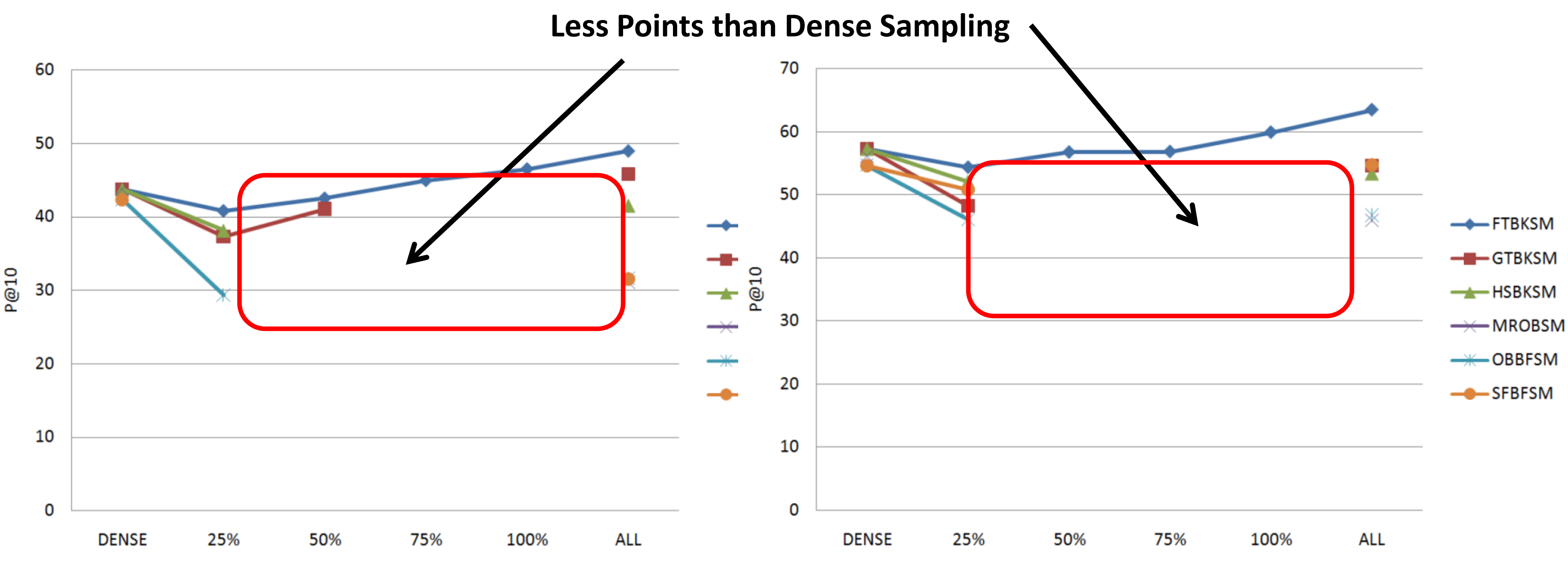
WANG Dataset



15 Scenes Dataset

25%, 50%, 75%, 100% = Sparse Sampling using 25% (or 50% or 75% or 100%) of points in Dense Sampling

Mid-level Representation Efficiency Evaluation (Sampling)

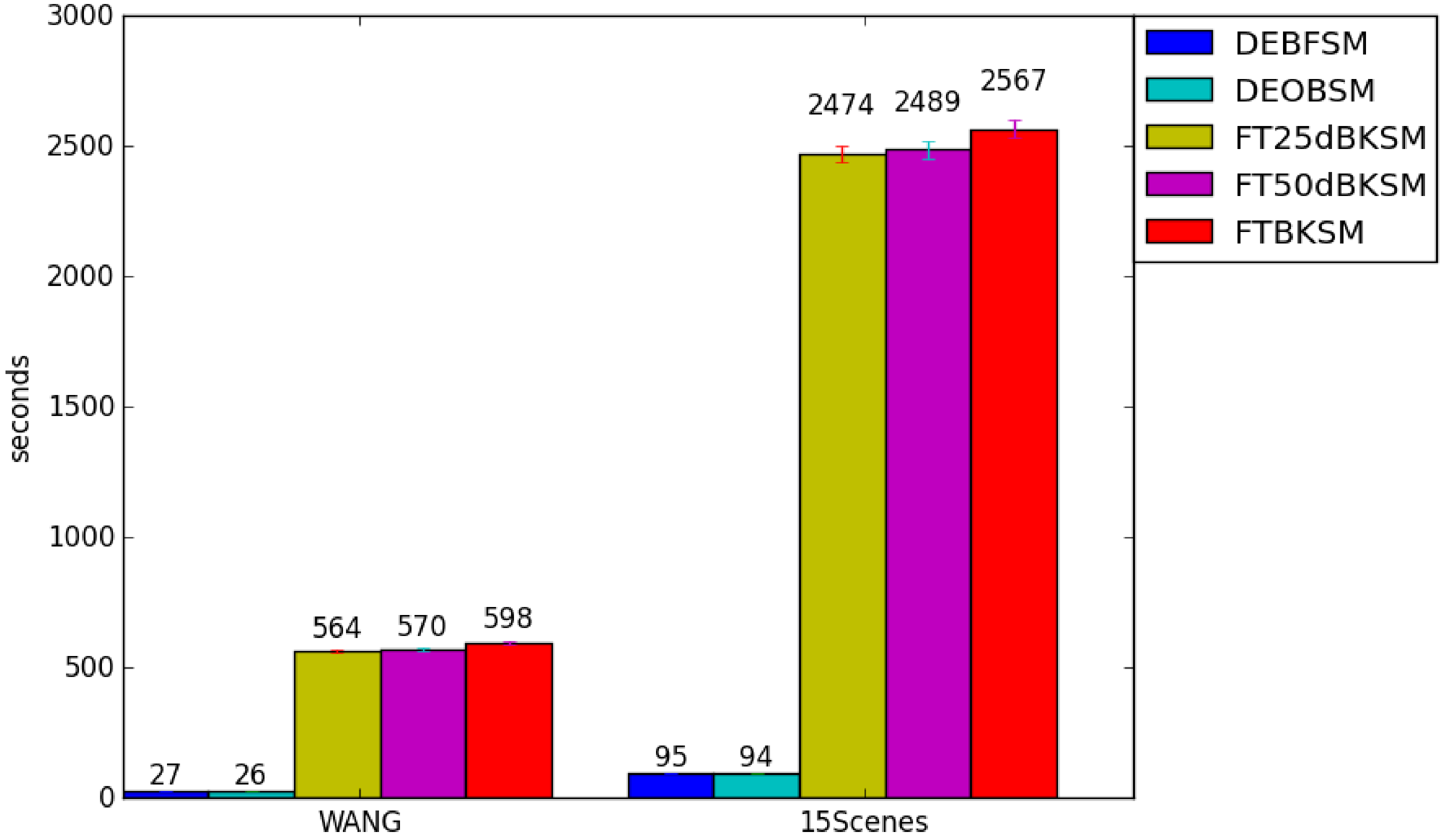


WANG Dataset

15 Scenes Dataset

25%, 50%, 75%, 100% = Sparse Sampling using 25% (or 50% or 75% or 100%) of points in Dense Sampling

Mid-level Representation Efficiency Evaluation (Sampling)



DENSE + BRIEF
DENSE + ORB
FAST + 25% Dense BRISK
FAST + 50% Dense BRISK
FAST + BRISK

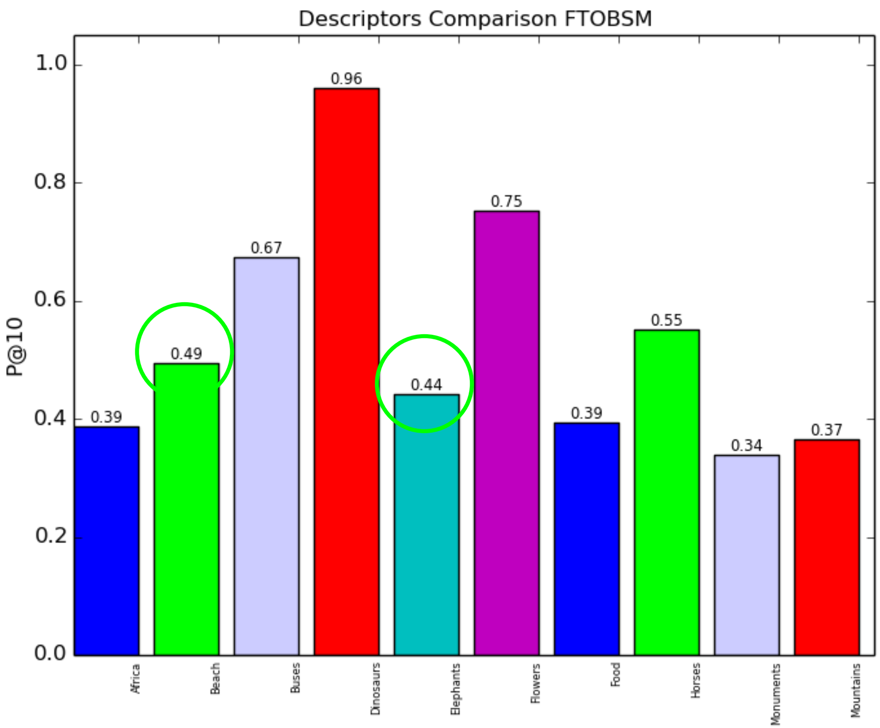
Mid-level Representation Efficiency Evaluation (Sampling)

- More suitable descriptors
  - DEOBSM = Dense + ORB + Bag of Words (Soft MAX)
  - FTOBSM = FAST + ORB + Bag of Words (Soft MAX)
  - FTBKSM = FAST + BRISK + Bag of Words (Soft MAX)
- Dense Vs Sparse
  - Dense is more accurate
  - Paired Statistical Test (95% of confidence)
    - Dense + ORB and FAST + ORB
    - Dense + ORB is the best

Dense X Sparse: WANG dataset

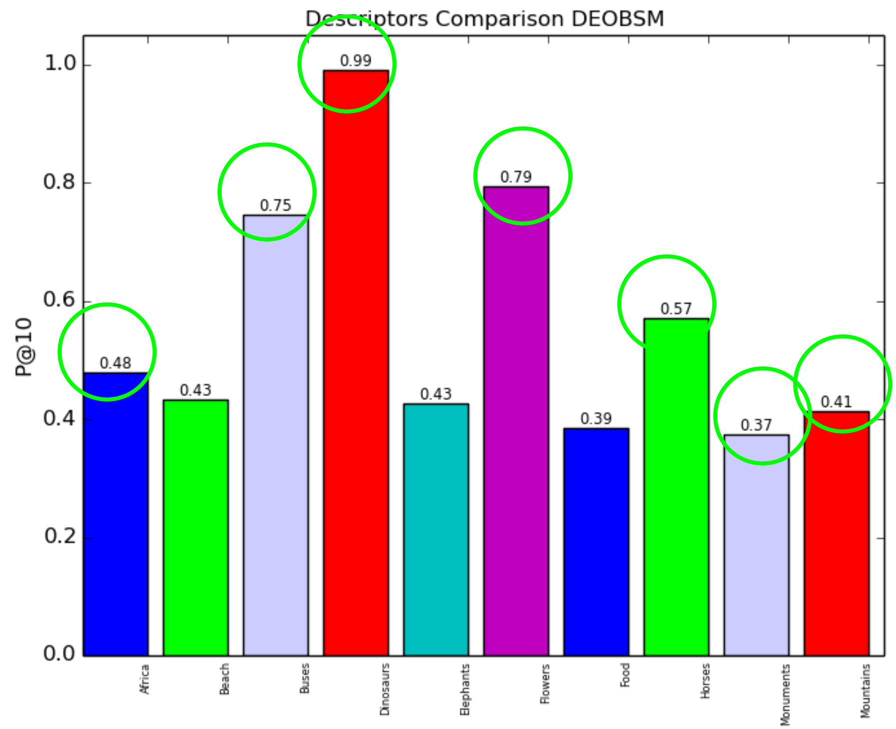
FAST+ ORB and FAST + ORB

Best Classes:
Beach, Elephants



Dense + ORB and FAST + ORB

Best Classes:
Africa, Buses, Dinosaus, Flowers, Horses, Monuments, Montains



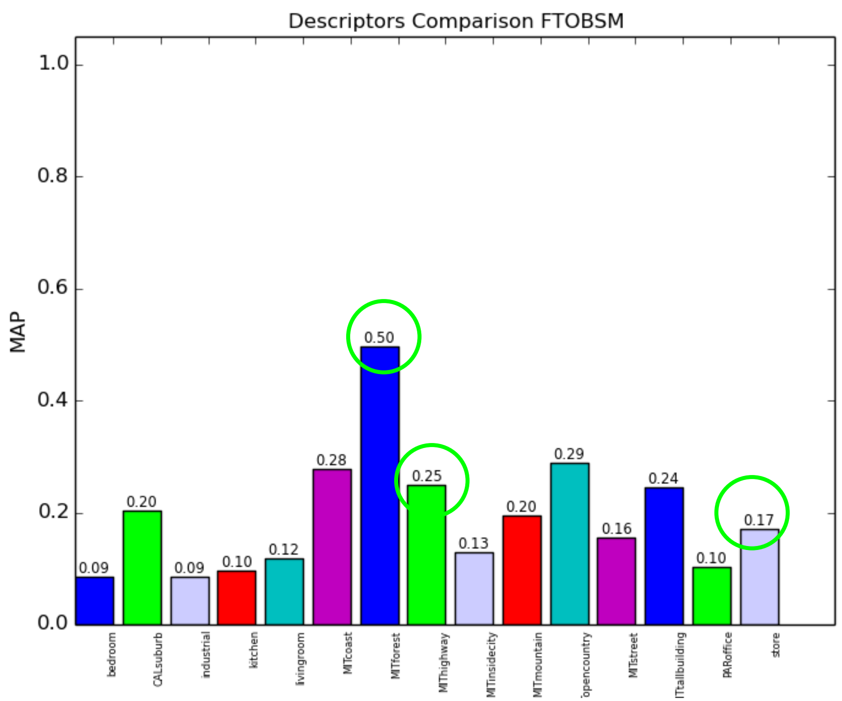
Dense X Sparse: 15Scenes dataset

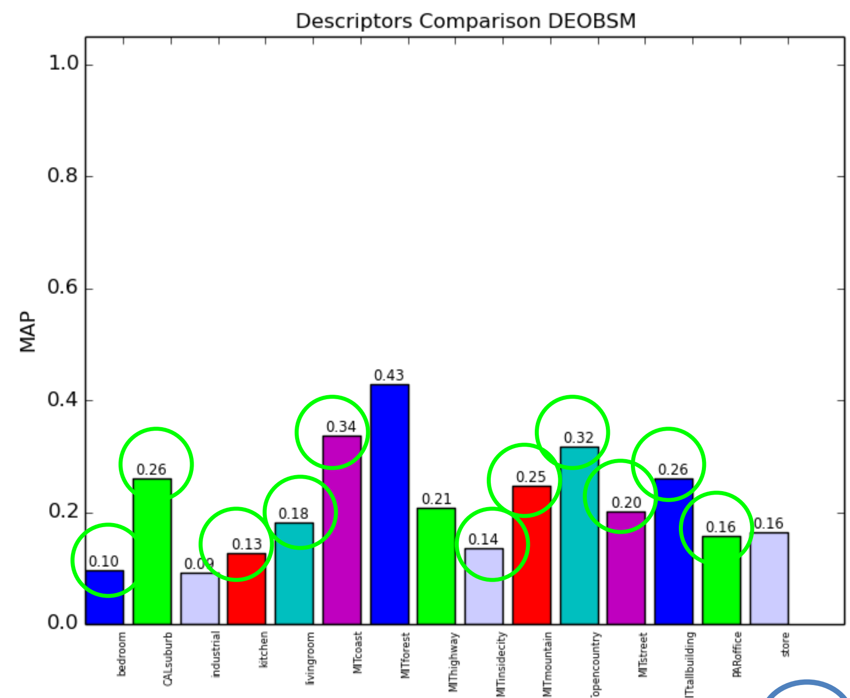
FAST+ ORB and FAST + ORB

Dense + ORB and FAST + ORB

Best Classes:
MITForest, MIThighway, store

Best Classes:
bedroom, CALsuburb, kitchen, livingroom, MITcoast, MITinsidecity, MITmountain, MITopencountry, MITstreet, MITtallbuilding, PARoffice

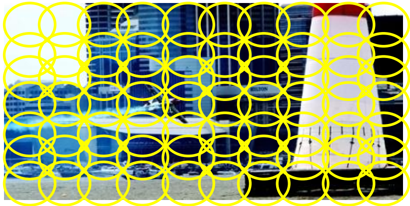




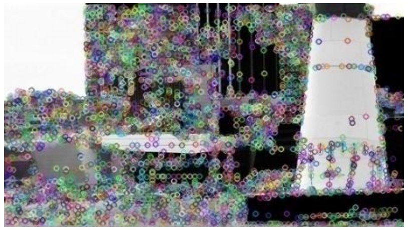
Restriction of Dense Sampling

- To use dense sampling on mobile
  - The images should not be so big
  - If image is big, dense sampling is slow
- For this reason
  - Is important resize images extremely big
  - Perform a more sparse dense sampling
- Experiments were performed before and after the resize, almost no impact on the accuracy

Dense Sampling



FAST Sampling



Low-Cost Representation for Mobile Image Search

- Experiment 1
  1. Mid-level Representation Analysis
     - Efficiency Evaluation (Fast?)
       – Sampling strategies comparison (15Scences and WANG datasets)
       – Time spent for feature extraction (Caltech101 and VOC2007 datasets)
     - Effectiveness Evaluation (Accurate?)
       – Precision: P@10/P@5 and MAP
       – Time X Precision
     - Compactness Evaluation (Compact ?)
       – Size of Images, Binary Features, Mid-level representation
       – Compresion Ratio X Precision
       – Lossless and Lossy Compression

Mid-level Representation Efficiency Evaluation (Time)

Database	N. Images	Classes	Multi Label?
Caltech101	9145	101	No
Pascal VOC 2007	9963	20	Yes

Caltech101

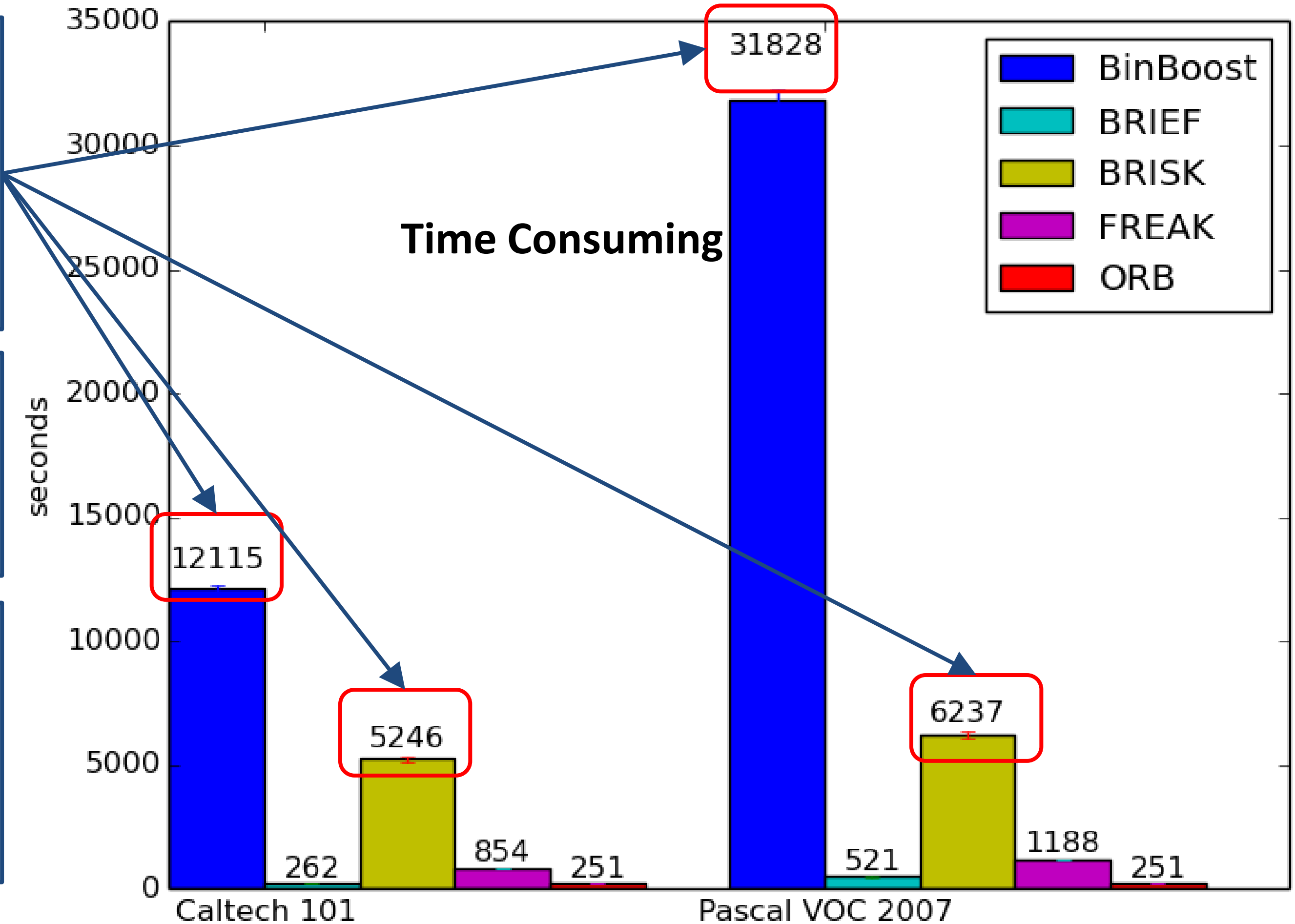
Pascal VOC 2007

Mid-level Representation Efficiency Evaluation (Time)

“BinBoost” and “BRISK” are “Time Consuming”

FREAK is Fast, but has Low Precision

BRIEF and ORB are Fast, with Good Precision

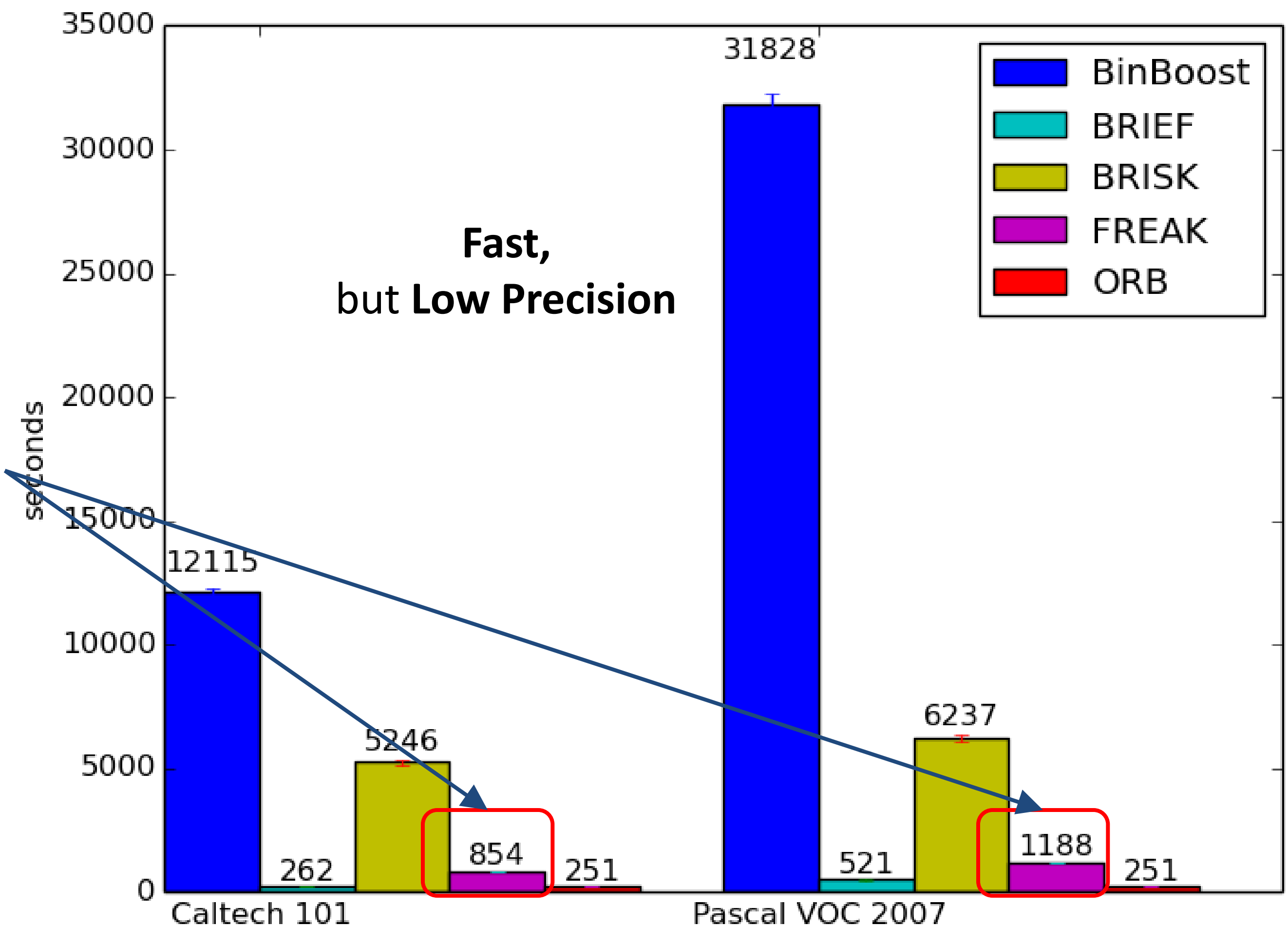


Mid-level Representation Efficiency Evaluation (Time)

"BinBoost" and "BRISK" are "Time Consuming"

FREAK is Fast, but has Low Precision

BRIEF and ORB are Fast, with Good Precision

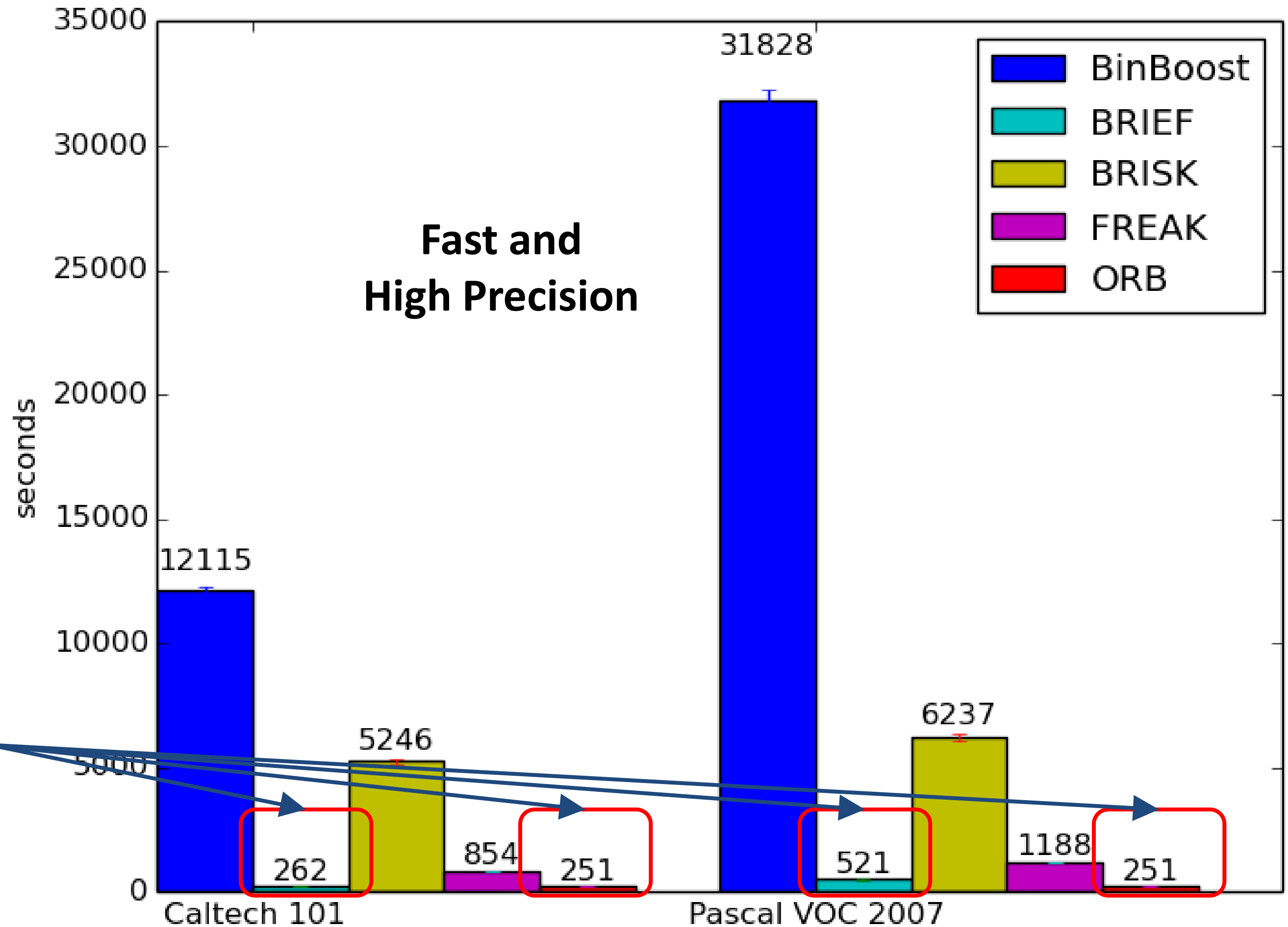


Mid-level Representation Efficiency Evaluation (Time)

"BinBoost" and "BRISK" are "Time Consuming"

FREAK is Fast, but has Low Precision

BRIEF and ORB are Fast, with Good Precision



Low-Cost Representation for Mobile Image Search

- Experiment 1
  1. Mid-level Representation Analysis
     - Efficiency Evaluation (Fast?)
       - Sampling strategies comparison (15Scences and WANG datasets)
       - Time spent for feature extraction (Caltech101 and VOC2007 datasets)
     - Effectiveness Evaluation (Accurate?)
       - Precision: P@10/P@5 and MAP
       - Time X Precision
     - Compactness Evaluation (Compact ?)
       - Size of Images, Binary Features, Mid-level representation
       - Compresion Ratio X Precision
       - Lossless and Lossy Compression

Mid-level Representation Effectiveness Evaluation (Precision)

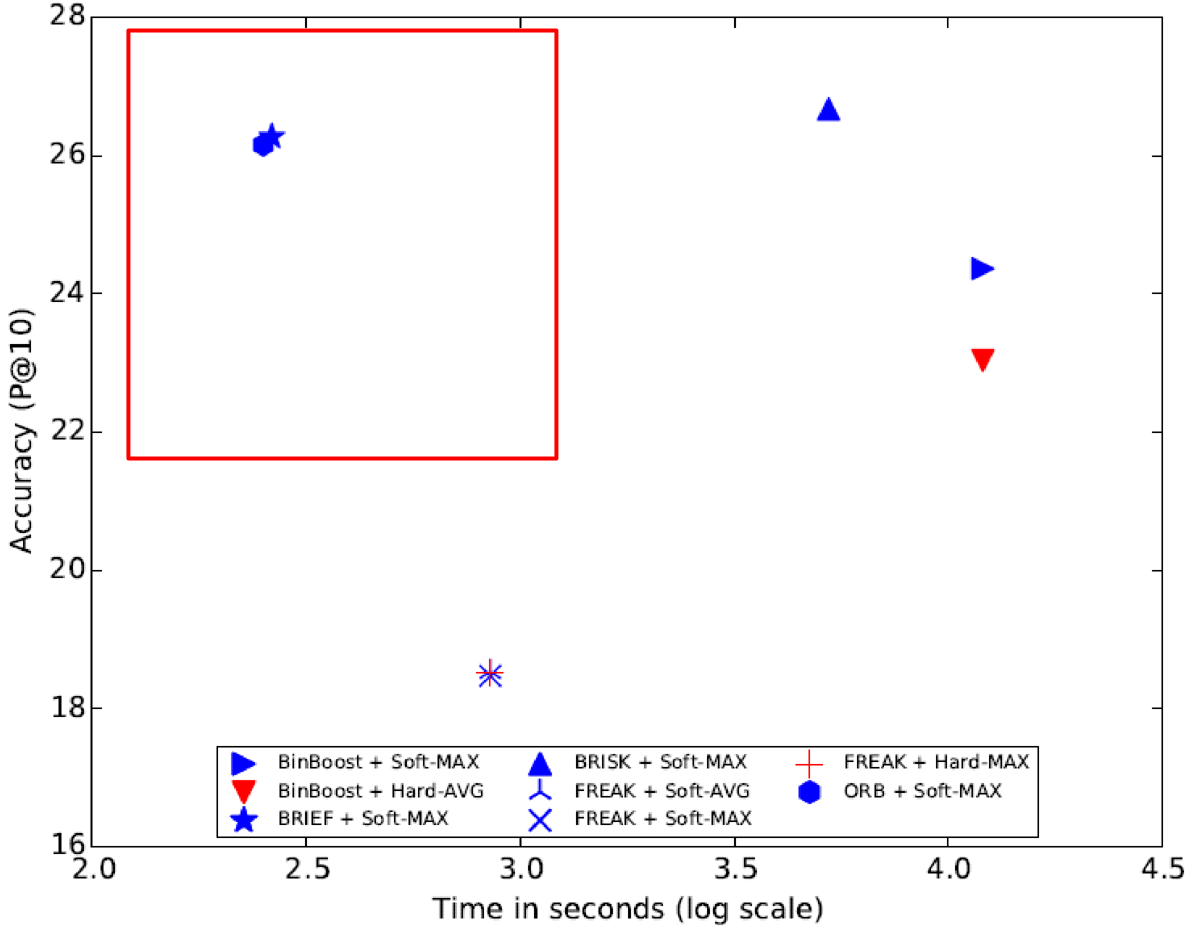
10 datasets: Results indicate **"Soft Assignment + MAX Pooling"** as the best option

Dataset	Descriptor	Soft-AVG	Soft-MAX	Hard-MAX	Hard-AVG
Caltech101 (Euclidean) (P@10)	BinBoost	15.8 +/- 0.3	**24.4 +/- 0.3**	17.6 +/- 0.3	**23.0 +/- 0.4**
	BRIEF	16.0 +/- 0.3	**26.3 +/- 0.4**	18.2 +/- 0.3	18.6 +/- 0.3
	BRISK	19.7 +/- 0.3	**26.7 +/- 0.4**	18.1 +/- 0.3	20.0 +/- 0.3
	FREAK	**18.5 +/- 0.3**	**18.5 +/- 0.3**	**18.5 +/- 0.3**	17.2 +/- 0.3
	ORB	14.1 +/- 0.2	**26.2 +/- 0.4**	17.1 +/- 0.3	17.7 +/- 0.3
VOC2007 (Euclidean) (P@10)	BinBoost	40.2 +/- 0.2	**44.9 +/- 0.2**	42.4 +/- 0.2	43.6 +/- 0.2
	BRIEF	39.8 +/- 0.2	**44.6 +/- 0.2**	39.6 +/- 0.2	**44.2 +/- 0.2**
	BRISK	43.3 +/- 0.2	42.0 +/- 0.2	39.0 +/- 0.2	41.1 +/- 0.2
	FREAK	**41.1 +/- 0.2**	37.7 +/- 0.2	37.9 +/- 0.2	**41.1 +/- 0.2**
	ORB	39.7 +/- 0.9	**45.7 +/- 0.2**	38.8 +/- 0.2	42.3 +/- 0.2

Low-Cost Representation for Mobile Image Search

- Experiment 1
  1. Mid-level Representation Analysis
     - Efficiency Evaluation (Fast?)
       - Sampling strategies comparison (15Scences and WANG datasets)
       - Time spent for feature extraction (Caltech101 and VOC2007 datasets)
     - Effectiveness Evaluation (Accurate?)
       - Precision: P@10/P@5 and MAP
       - Time X Precision
     - Compactness Evaluation (Compact ?)
       - Size of Images, Binary Features, Mid-level representation
       - Compresion Ratio X Precision
       - Lossless and Lossy Compression

Mid-level Representation
Effectiveness Evaluation (Precision X Time)



Best: BRIEF + SoftMAX or ORB + SoftMAX

Low-Cost Representation for Mobile Image Search

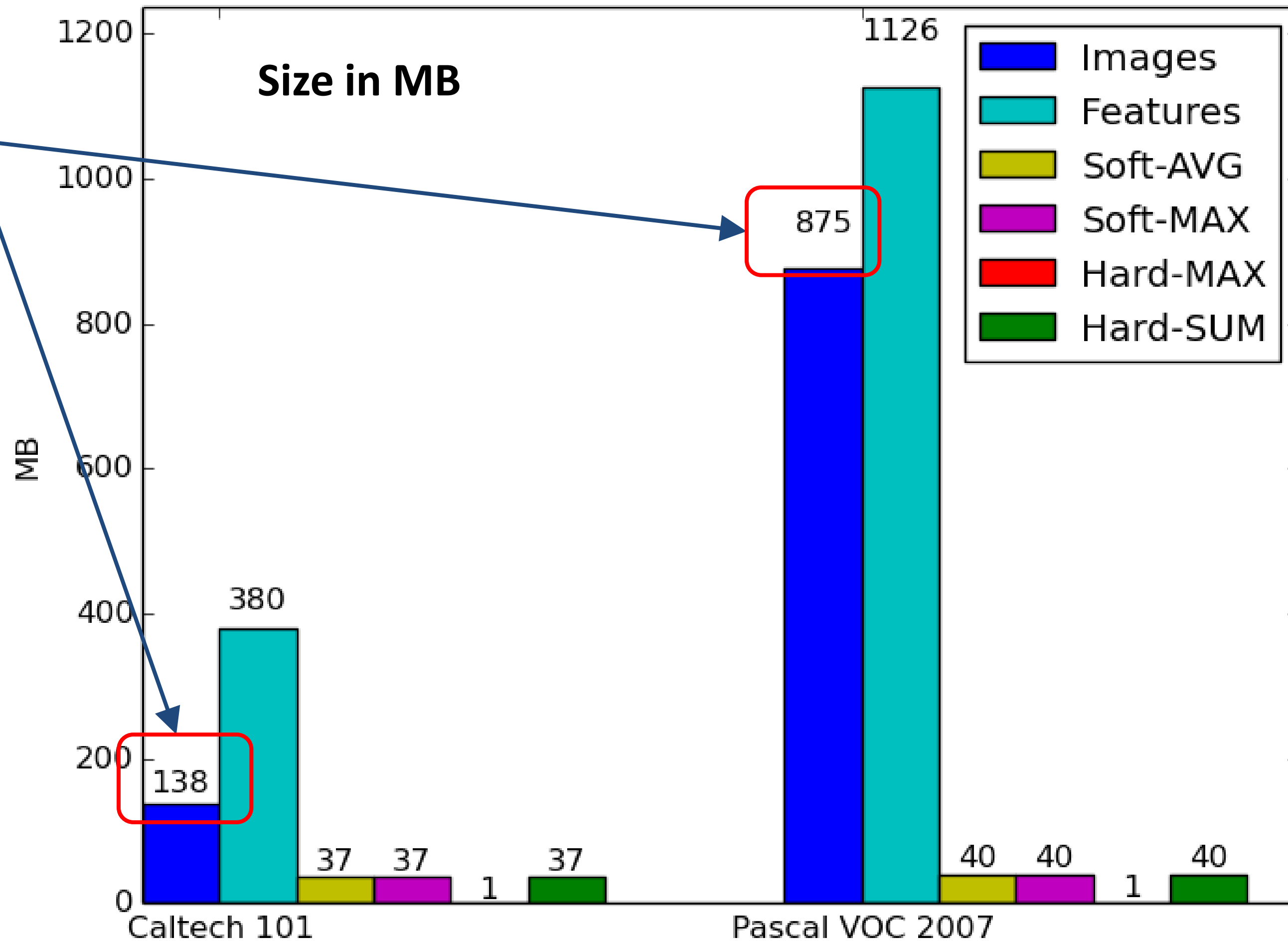
- Experiment 1
  1. Mid-level Representation Analysis
     - Efficiency Evaluation (Fast?)
       - Sampling strategies comparison (15Scences and WANG datasets)
       - Time spent for feature extraction (Caltech101 and VOC2007 datasets)
     - Effectiveness Evaluation (Accurate?)
       - Precision: P@10/P@5 and MAP
       - Time X Precision
     - Compactness Evaluation (Compact ?)
       - Size of Images, Binary Features, Mid-level representation
       - Compresion Ratio X Precision
       - Lossless and Lossy Compression

Mid-level Representation Compactness Evaluation

Size All Images

Size All Binary Features (ORB) of All Images

Size All Mid-Level Repres. (ORB) of All Images

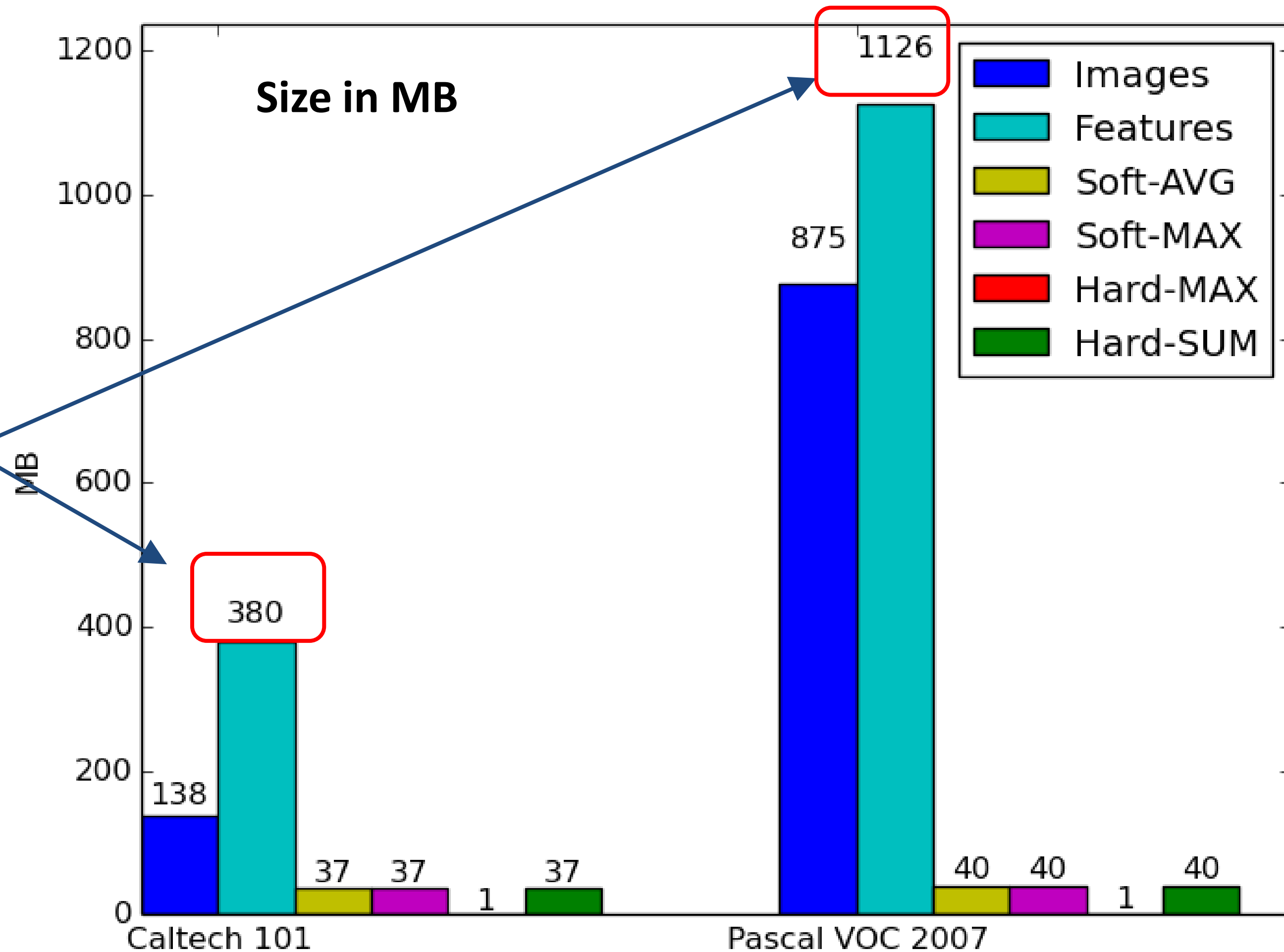


Mid-level Representation Compactness Evaluation

Size All Images

Size All Binary Features (ORB) of All Images

Size All Mid-Level Repres. (ORB) of All Images

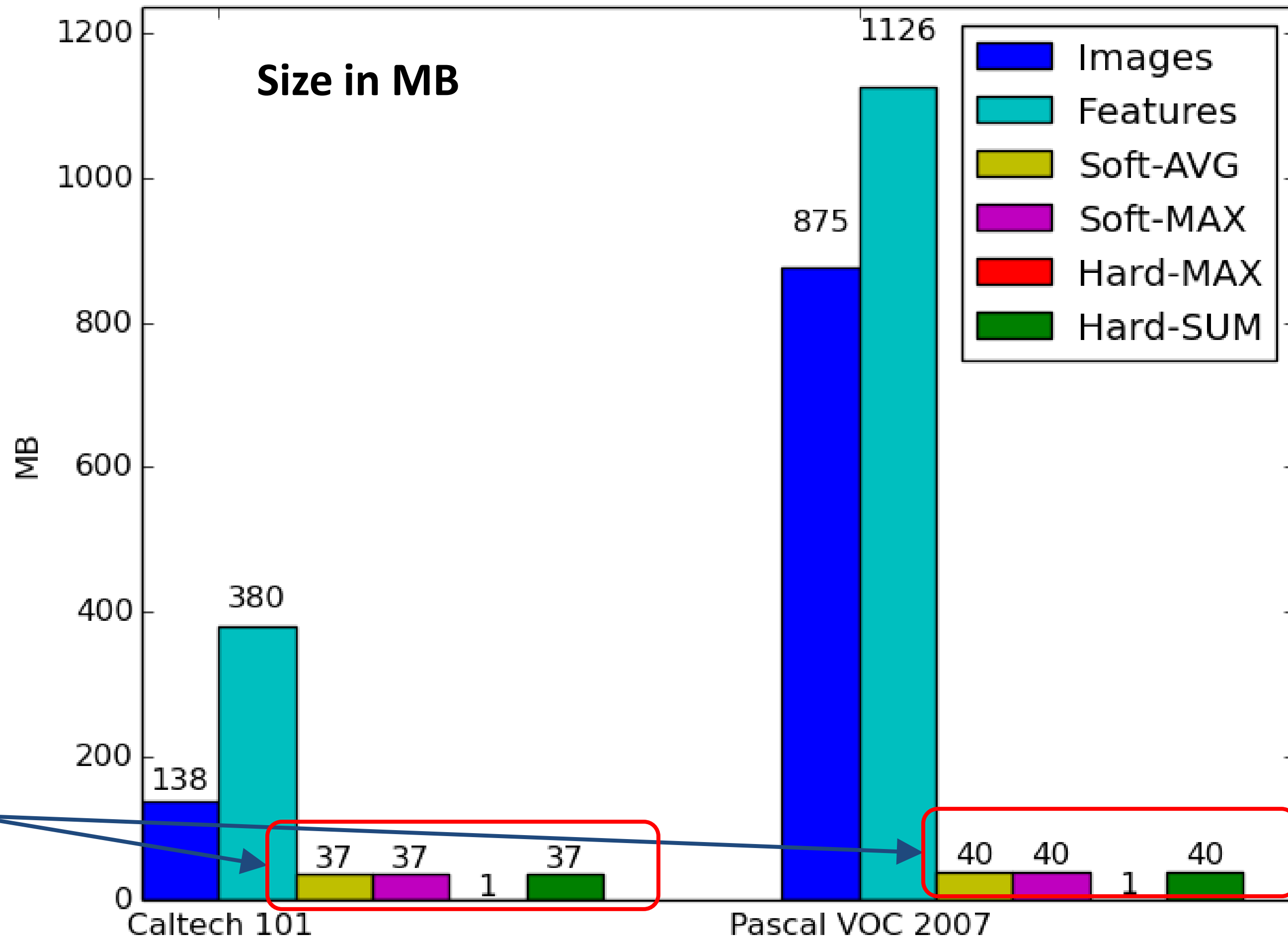


Mid-level Representation Compactness Evaluation

Size All Images

Size All Binary Features (ORB) of All Images

Size All Mid-Level Repres. (ORB) of All Images

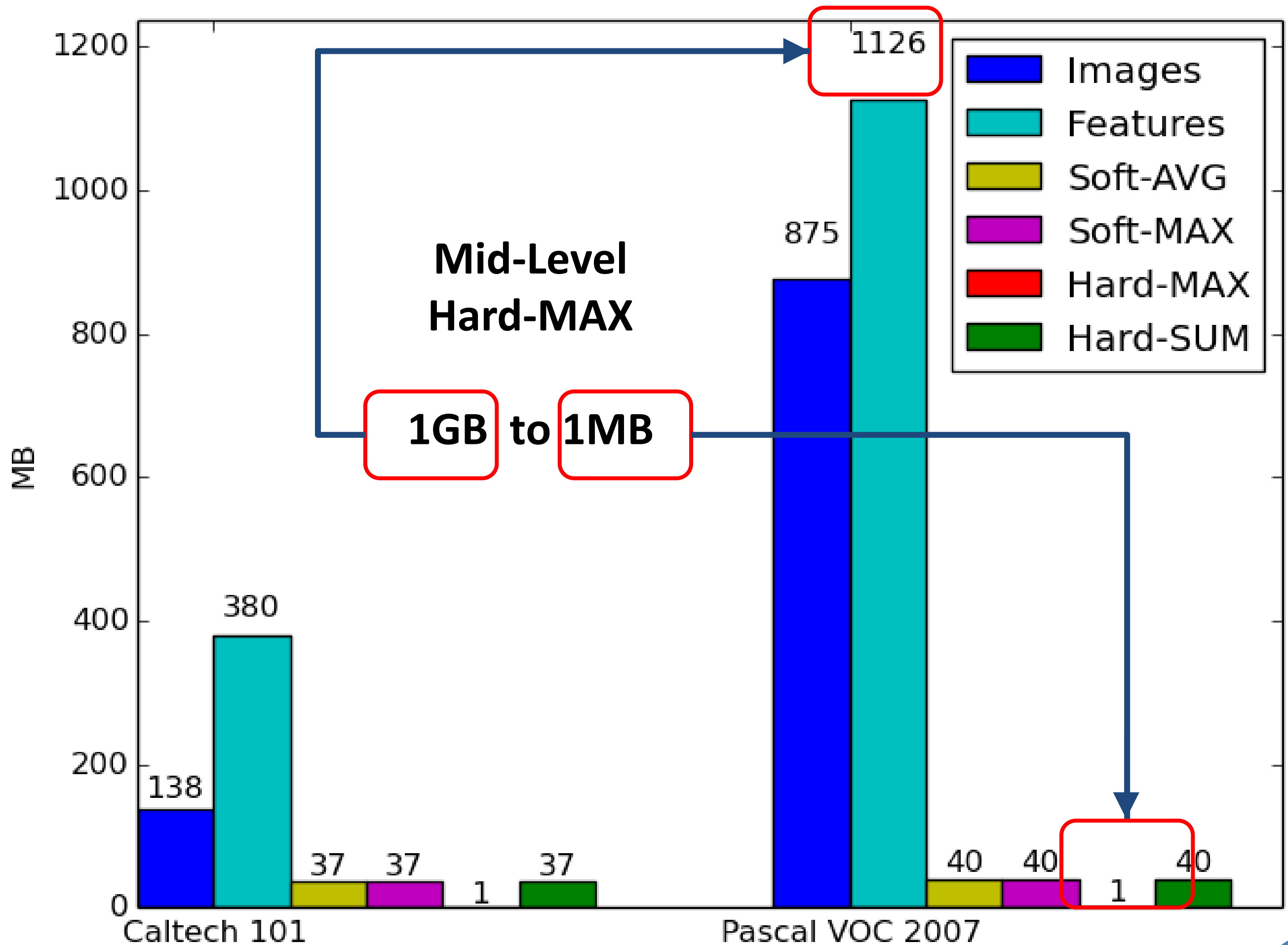


Mid-level Representation Compactness Evaluation

Size All Images

Size All Binary Features (ORB) of All Images

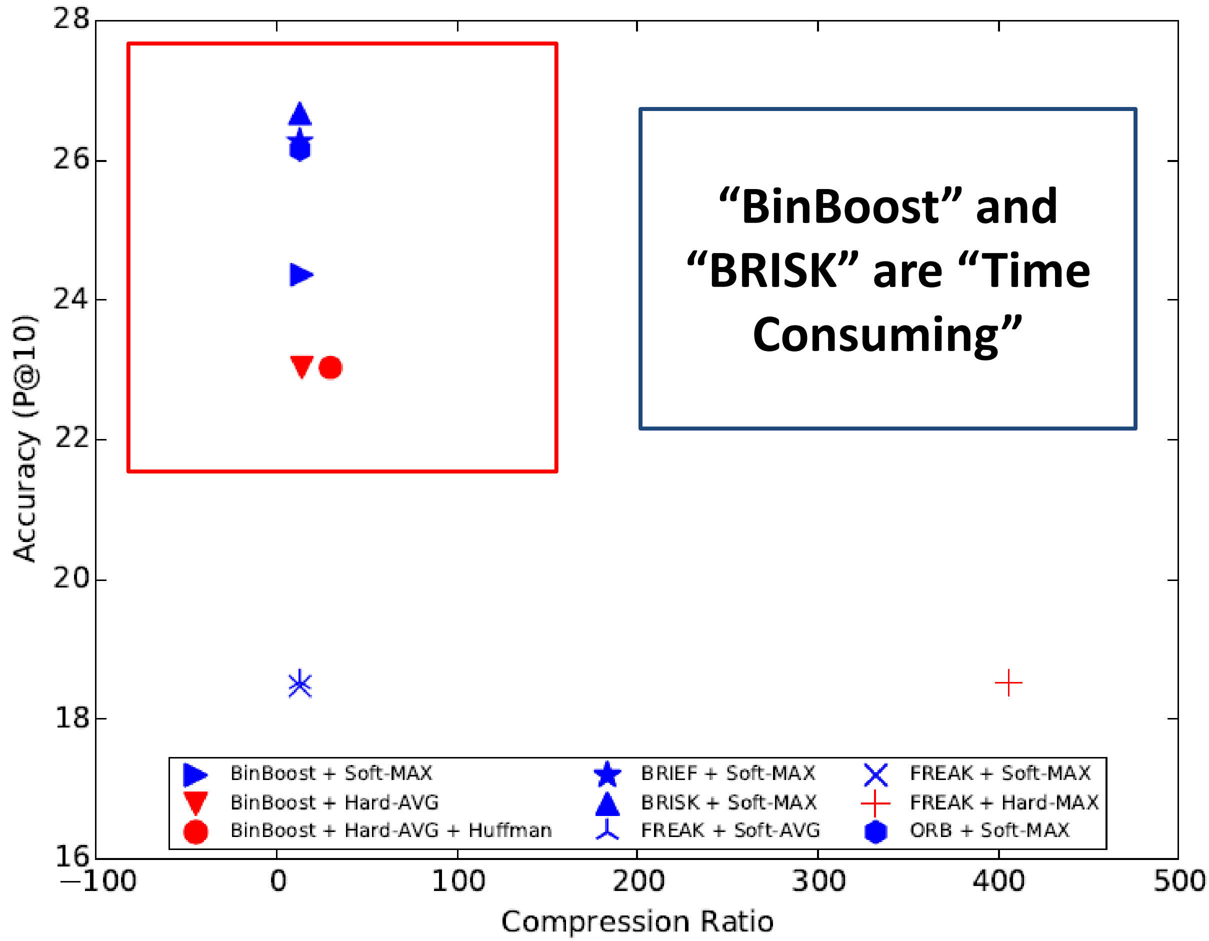
Size All Mid-Level Repres. (ORB) of All Images



Low-Cost Representation for Mobile Image Search

- Experiment 1
  1. Mid-level Representation Analysis
     - Efficiency Evaluation (Fast?)
       - Sampling strategies comparison (15Scences and WANG datasets)
       - Time spent for feature extraction (Caltech101 and VOC2007 datasets)
     - Effectiveness Evaluation (Accurate?)
       - Precision: P@10/P@5 and MAP
       - Time X Precision
     - Compactness Evaluation (Compact ?)
       - Size of Images, Binary Features, Mid-level representation
       - Compresion Ratio X Precision
       - Lossless and Lossy Compression

Mid-level Representation
Compactness Evaluation (Precision X CR)

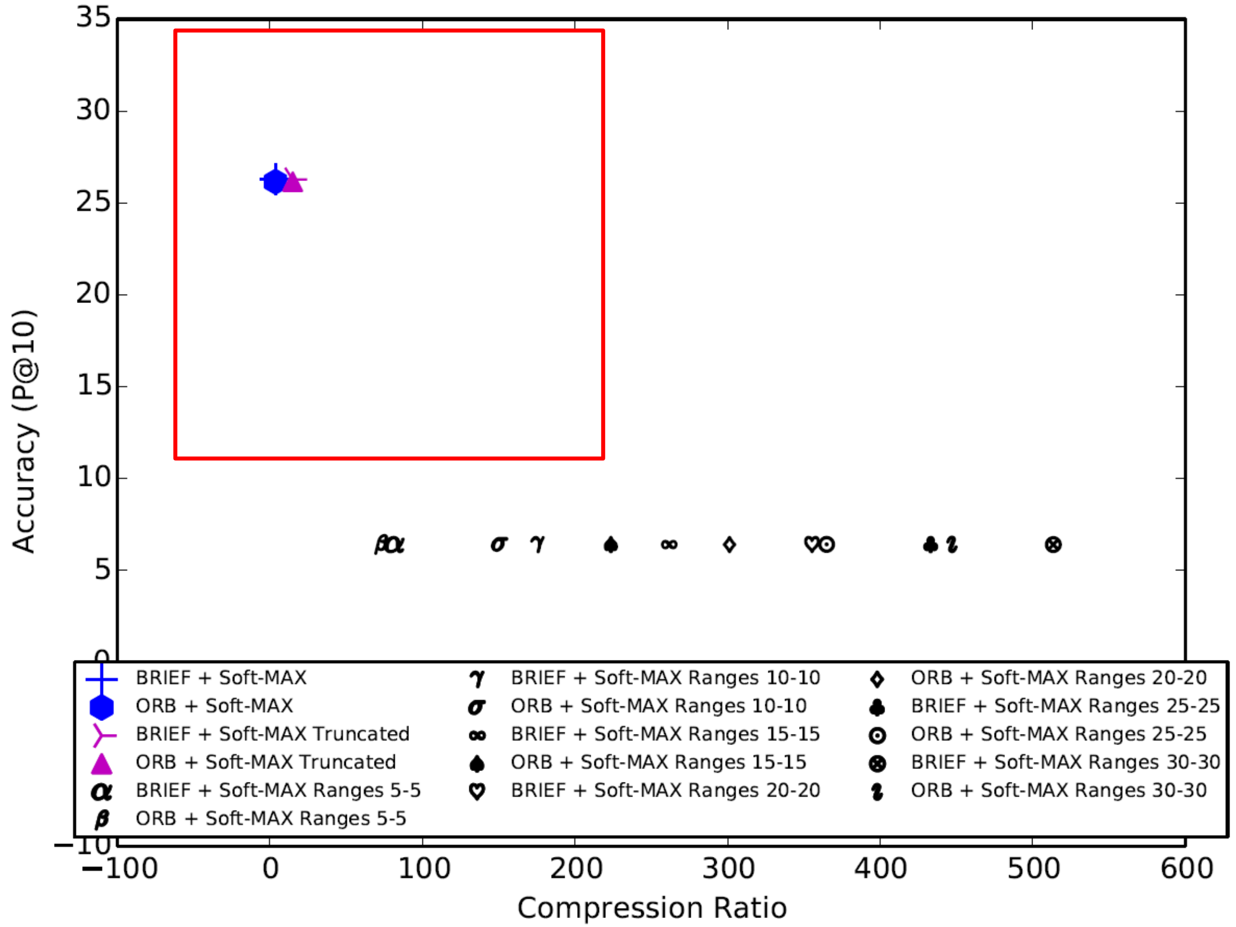


Best: BRIEF + SoftMAX or ORB + SoftMAX

Low-Cost Representation for Mobile Image Search

- Experiment 1
  1. Mid-level Representation Analysis
     - Efficiency Evaluation (Fast?)
       - Sampling strategies comparison (15Scences and WANG datasets)
       - Time spent for feature extraction (Caltech101 and VOC2007 datasets)
     - Effectiveness Evaluation (Accurate?)
       - Precision: P@10/P@5 and MAP
       - Time X Precision
     - Compactness Evaluation (Compact ?)
       - Size of Images, Binary Features, Mid-level representation
       - Compresion Ratio X Precision
       - Lossless and Lossy Compression

Mid-level Representation Compactness Evaluation (Precision X CR)



Best: BRIEF + SoftMAX or ORB + SoftMAX

Low-Cost Representation for Mobile Image Search

- Experiment 1
  1. Mid-level Representation Analysis
     - Efficiency Evaluation (Fast?)
       - Sampling strategies comparison (15Scences and WANG datasets)
       - Time spent for feature extraction (Caltech101 and VOC2007 datasets)
     - Effectiveness Evaluation (Accurate?)
       - Precision: P@10/P@5 and MAP
       - Time X Precision
     - Compactness Evaluation (Compact ?)
       - Size of Images, Binary Features, Mid-level representation
       - Compresion Ratio X Precision
       - Lossless and Lossy Compression

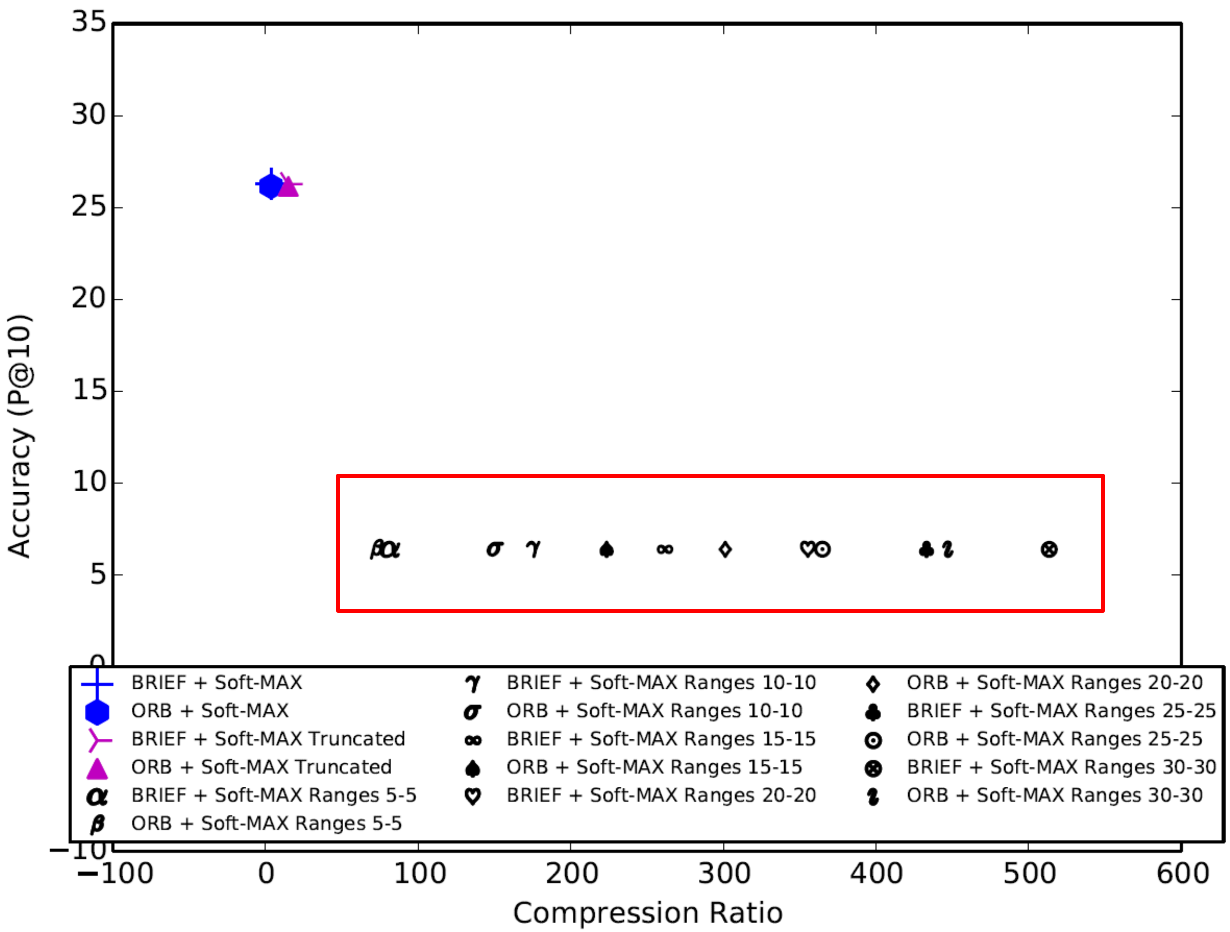
Mid-level Representation
Compactness Evaluation (Precision X CR)

	CR Caltech 101	CR VOC 2007
BRIEF + SoftMAX	3.69	21.7
BRIEF + SoftMAX Truncated	14.9	85
ORB + SoftMAX	3.75	21.41
ORB + SoftMAX Truncated	15.07	85.83

Accuracy: Lost only 0.002%

Compression ratio of ~ 25,06%

Mid-level Representation Compactness Evaluation (Precision X CR)



"SoftMAX Truncated using Ranges" has a High Compression Ratio, but Low Precision

Low-Cost Representation for Mobile Image Search

- [Pessoa et al., 2015a] **A Study on Low-Cost Representations for Image Feature Extraction on Mobile Devices**
  - Mid-level Representation Analysis
    - Efficiency Evaluation (Fast?)
    - Effectiveness Evaluation (Accurate?)
    - Compactness Evaluation (Compact ?)

Publication:

XX Iberoamerican Congress on Pattern Recognition (CIARP 2015)

Contributions

1. A comparative study of Low-Cost Representation for Mobile Image Search

2. We propose two new bag of visual words representations that include spatial information

Low-Cost Representation for Mobile Image Search

- Experiments
  1. Mid-level Representation Analysis (Experiment 1)
     - Efficiency Evaluation (Fast?)
     - Effectiveness Evaluation (Accurate?)
     - Compactness Evaluation (Compact ?)
  2. Low-level Global Representation Analysis (Experiment 2)
     - Efficiency Evaluation (Fast?)
     - Effectiveness Evaluation (Accurate?)
     - Compactness Evaluation (Compact ?)

All Results Analyzed Statistically with 95% confidence

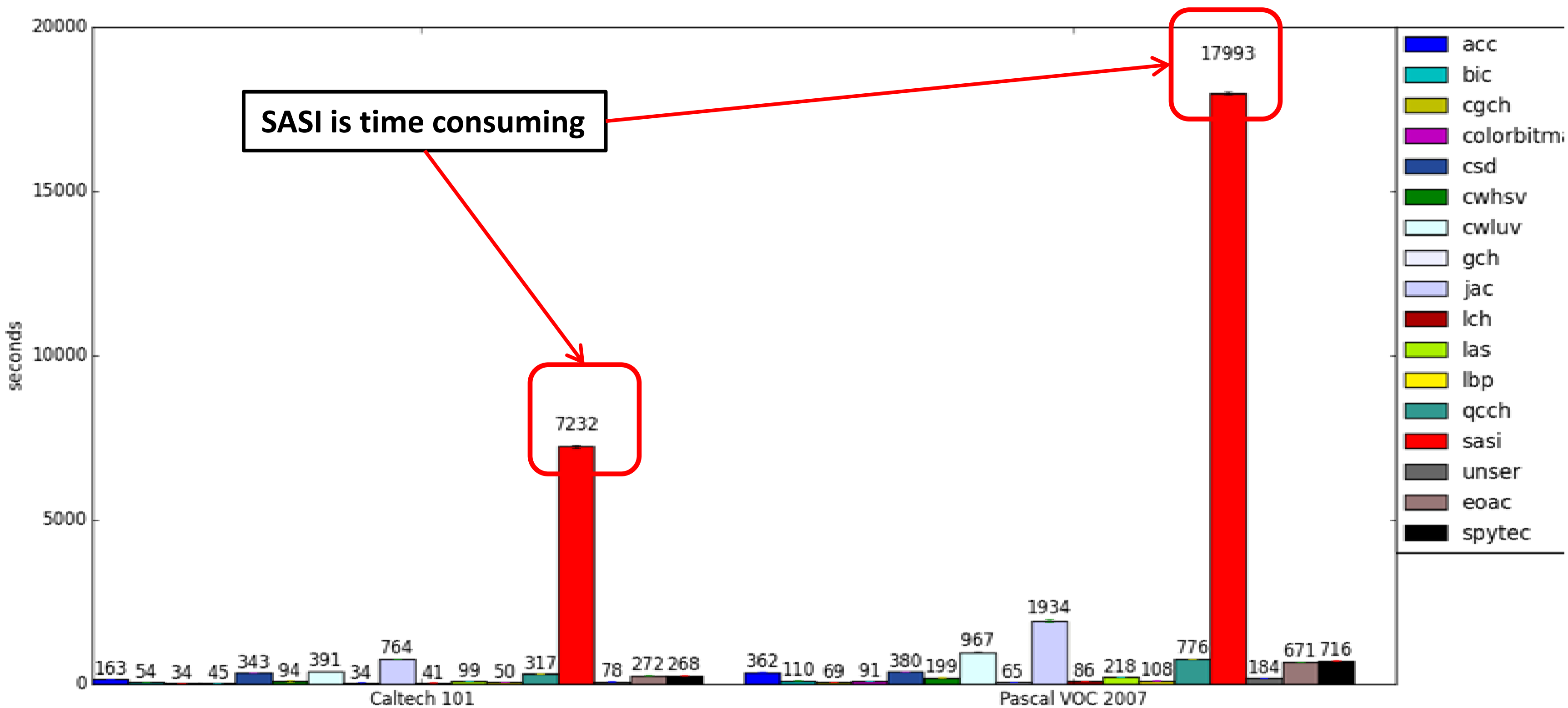
Low-Cost Representation for Mobile Image Search

- Experiment 2

  1. Low-level Global Representation Analysis

     - Efficiency Evaluation (Fast?)
       - Time spent for global feature extraction (Caltech101 and VOC2007 datasets)
     - Effectiveness Evaluation (Accurate?)
       - Precision: P@10/P@5 and MAP
       - Time X Precision
     - Compactness Evaluation (Compact ?)
       - Size of Global Representations
       - Compresion Ratio X Precision

Low-level Global Representation Efficiency Evaluation (Time)



Low-Cost Representation for Mobile Image Search

- Experiment 2

1. Low-level Global Representation Analysis
   - Efficiency Evaluation (Fast?)
     - Time spent for global feature extraction (Caltech101 and VOC2007 datasets)
   - Effectiveness Evaluation (Accurate?)
     - Precision: P@10/P@5 and MAP
     - Time X Precision
   - Compactness Evaluation (Compact ?)
     - Size of Global Representations
     - Compresion Ratio X Precision

Low-level Global Representation Effectiveness Evaluation (Precision)

- In five of ten datasets
  - BIC (Border/Interior Pixel Classification) descriptor achieved the best results
- Other good results
  - LCH (Local Color Histogram)
  - LAS (Local Activity Spectrum)
  - SASI (Statistical Analysis of Structural Information)

Dataset	Descriptor		Descriptor	
WANG (P@10)	BIC	**77.73 +/- 0.1**	QCCH	48.97 +/- 0.14
	LCH	64.13 +/- 0.13	SASI	64.62 +/- 0.12
	LAS	59.84 +/- 0.14	EOAC	60.44 +/- 0.14

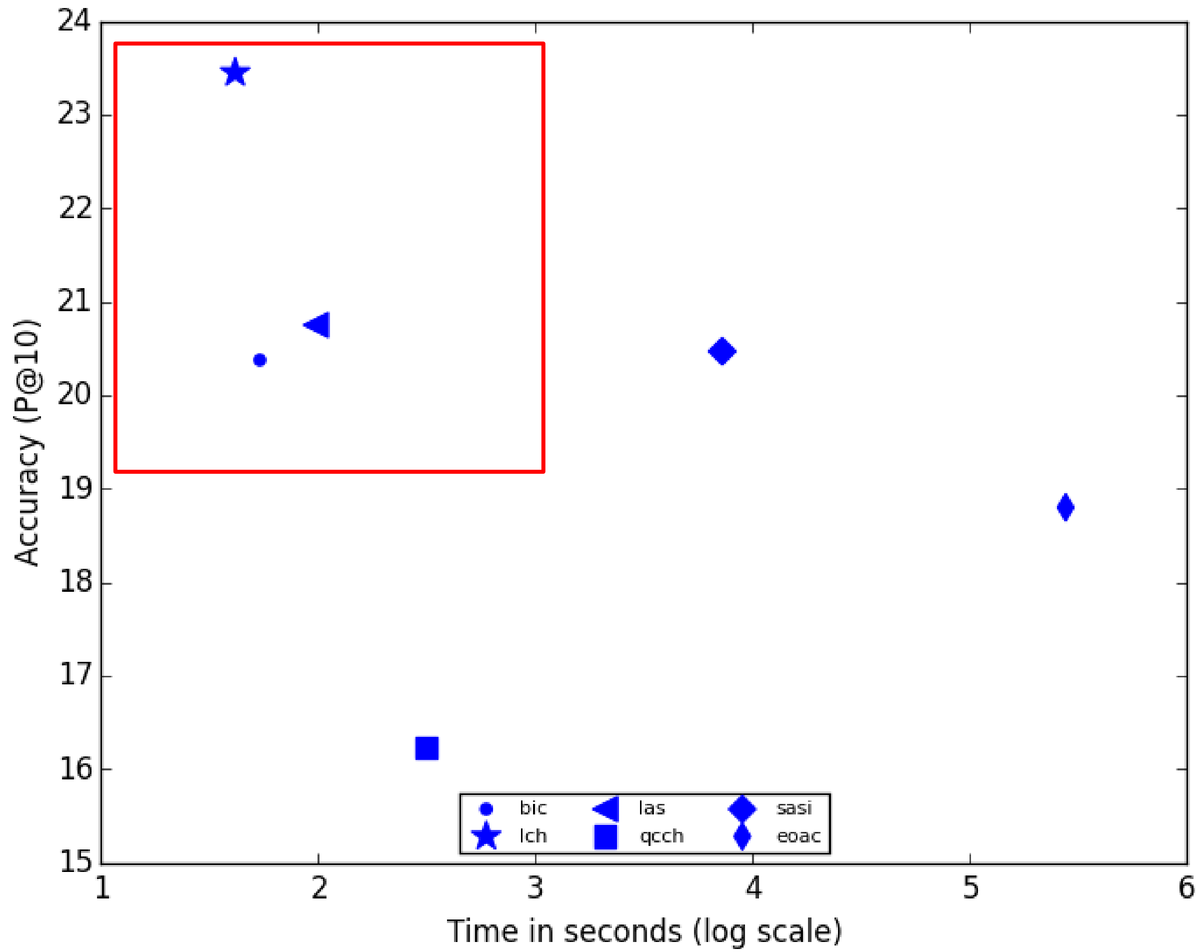
Low-Cost Representation for Mobile Image Search

- Experiment 2

  1. Low-level Global Representation Analysis

     - Efficiency Evaluation (Fast?)
       - Time spent for global feature extraction (Caltech101 and VOC2007 datasets)
     - Effectiveness Evaluation (Accurate?)
       - Precision: P@10/P@5 and MAP
       - Time X Precision
     - Compactness Evaluation (Compact ?)
       - Size of Global Representations
       - Compresion Ratio X Precision

Low-level Global Representation Effectiveness Evaluation (Precision X Time)



Best: BIC (Color), LCH (Color), LAS (Texture)

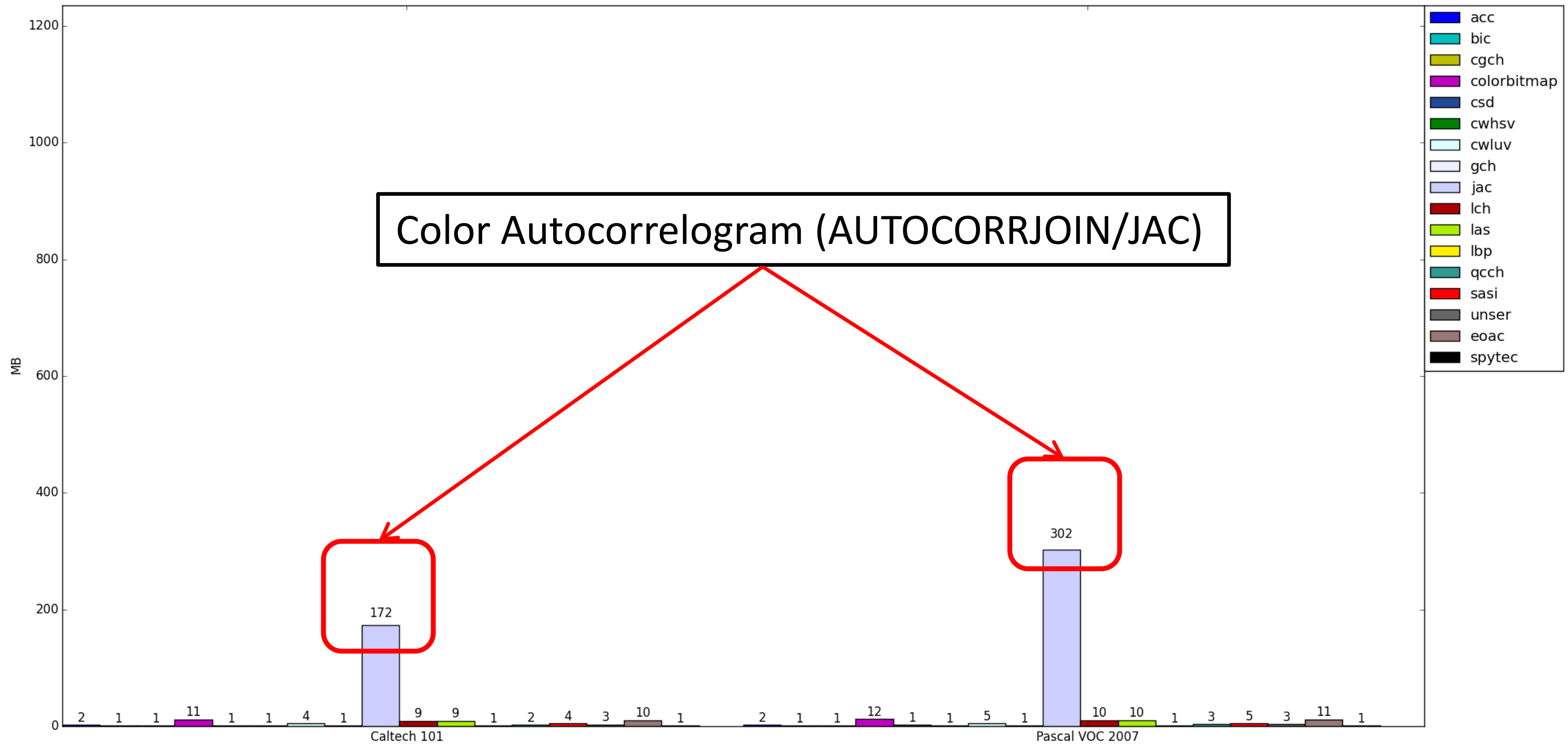
Low-level Global Representation Effectiveness Evaluation (Precision X Time)

- Why BIC and LCH performed well in our experiments?
  - Several of the analyzed datasets have a lot of color information
  - This is a great advantage for the BIC and LCH descriptors
- What happened in datasets with grayscale images?
  - Texture or shape descriptors are the best options, such as LAS (Texture) and SASI (Texture)

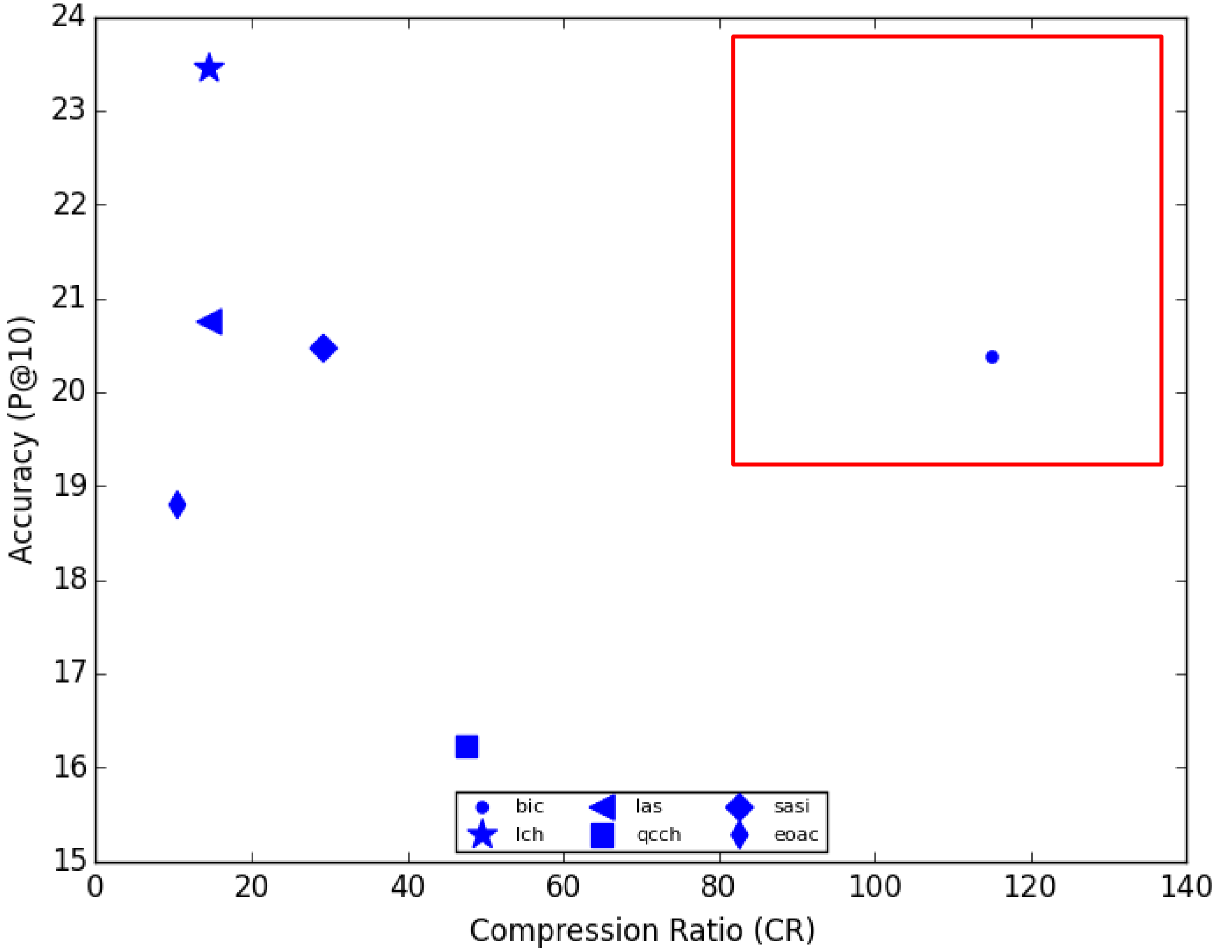
Low-Cost Representation for Mobile Image Search

- Experiment 2
  1. Low-level Global Representation Analysis
     - Efficiency Evaluation (Fast?)
       - Time spent for global feature extraction (Caltech101 and VOC2007 datasets)
     - Effectiveness Evaluation (Accurate?)
       - Precision: P@10/P@5 and MAP
       - Time X Precision
     - Compactness Evaluation (Compact ?)
       - Size of Global Representations
       - Compresion Ratio X Precision

Mid-level Representation Compactness Evaluation



Mid-level Representation Compactness Evaluation



Best: BIC (Color)

Low-Cost Representation

- [Pessoa et al., 2015b] **An Experimental Comparison of Feature Extraction and Distance Metrics for Image Retrieval**
  - Global Descriptors
  - Distance metrics
  - Impact of these factors in the CBIR process
  - Statistical Analysis

Publication:

Conference on Graphics, Patterns and Images (SIBGRAPI WIP 2015)

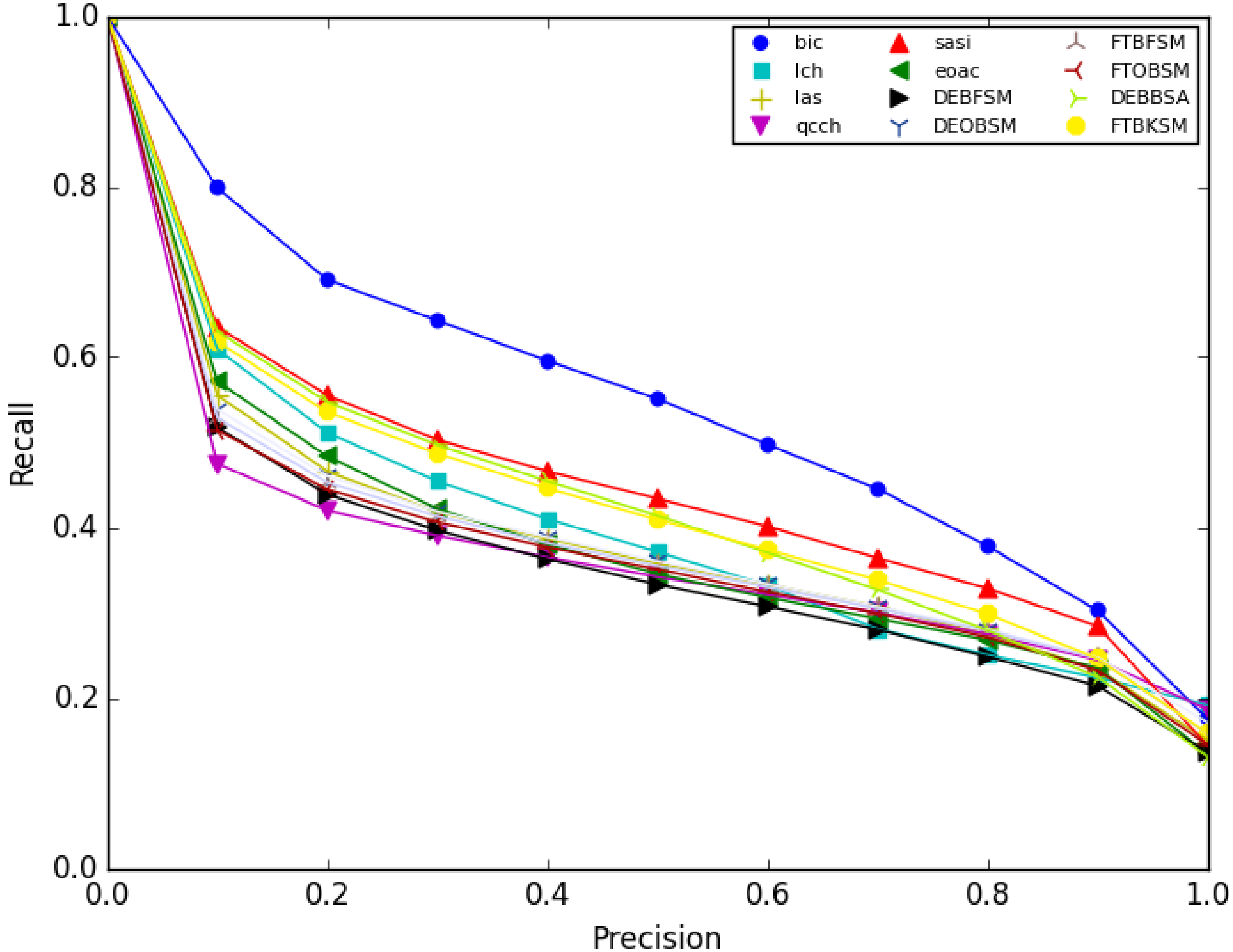
Low-Cost Representation for Mobile Image Search

- Analysis of effectiveness, efficiency and compactness
  - More suitable global descriptor
    - BIC (Border/Interior Pixel Classification)
  - More suitable global descriptor
    - DEOBSM (Bag of Words using Dense Sampling, ORB descriptor, Soft assignment and Maximum pooling)
- **Paired statistical tests**
  - BIC can be considered better then DEOBSM
- For mobile image retrieval, we may consider use BIC descriptor as the best option considering the triple trade-off problem

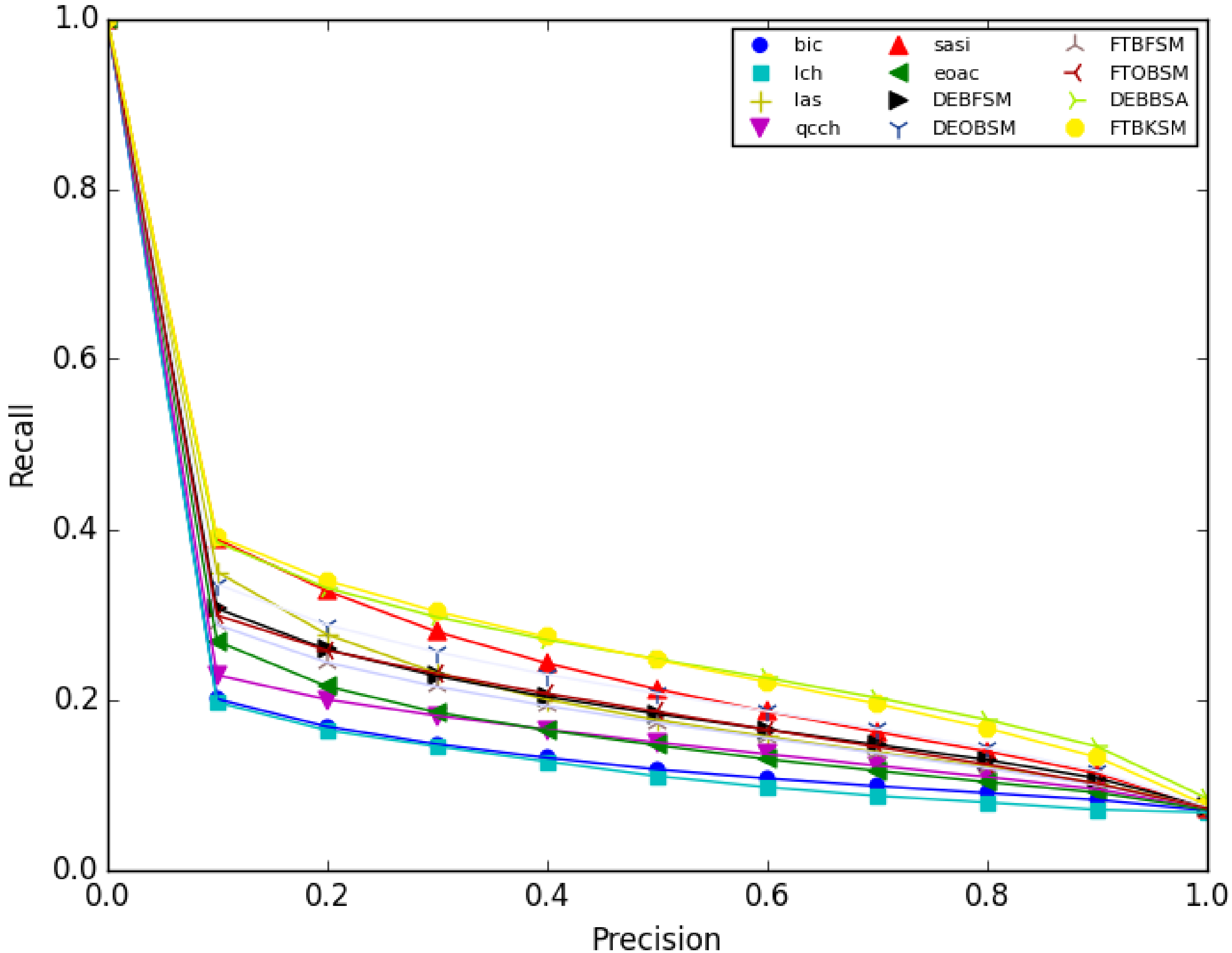
BIC Vs Dense+ORB+SoftMAX (DEOBSM)

Dataset	BIC	DEOBSM	Best, IC (95%)
1: 15Scenes (P@10)	30.2 +/- 0.05	**43.88 +/- 0.08**	DEOBSM
2: caltech101 (P@10)	20.37 +/- 0.03	**26.51 +/- 0.08**	DEOBSM
3: caltech256 (P@10)	**15.31 +/- 0.01**	13.99 +/- 0.08	BIC
4: OxBuild11 (P@10)	28.43 +/- 0.1	**46.29 +/- 0.08**	DEOBSM
5: Paris (P@10)	32.79 +/- 0.07	32.90 +/- 0.08	Not different
6: SMVS692 (P@5)	23.08 +/- 0.01	**34.27 +/- 0.08**	DEOBSM
7: UWdataset (P@10)	**59.74 +/- 0.1**	35.05 +/- 0.08	BIC
8: VOC2007 (P@10)	**21.05 +/- 0.03**	20.83 +/- 0.08	BIC
9: WANG (P@10)	**77.73 +/- 0.1**	56.12 +/- 0.08	BIC
10: ZuBuD (P@5)	**72.62 +/- 0.02**	70.01 +/- 0.08	BIC

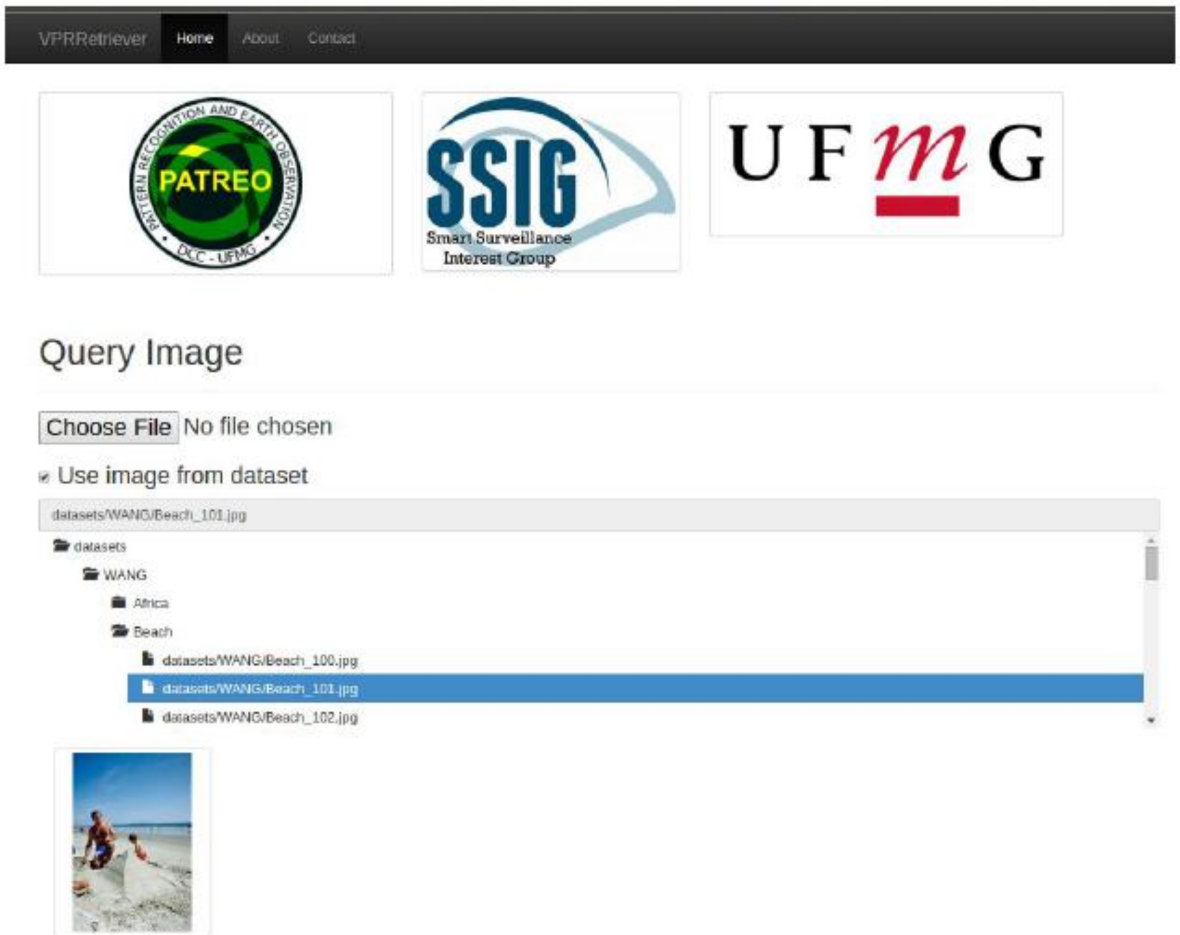
Twelve best descriptors found out in our analysis in Precision Vs Recall Curves (WANG)



Twelve best descriptors found out in our analysis in Precision Vs Recall Curves (15Scenes)



VPR Retriever System
http://goldenretriever.dcc.ufmg.br



VPR Retriever System
http://goldenretriever.dcc.ufmg.br

Dataset

WANG

Distance Metric

Manhattan (L1)

Representation

bic (Global and Color)

Search Similar Images

Retrieved Images

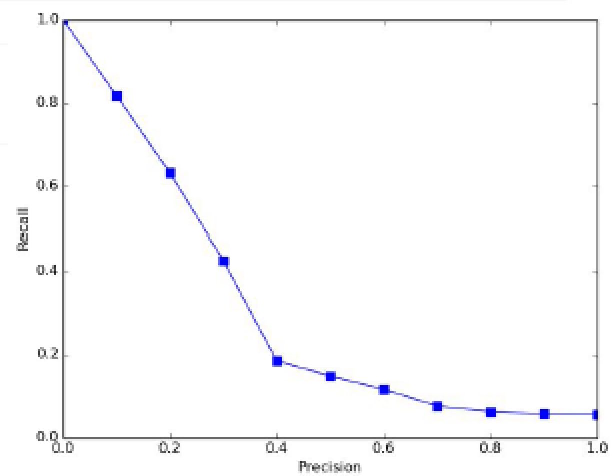
Show 20 **images**

Statistics (last selected image dataset)

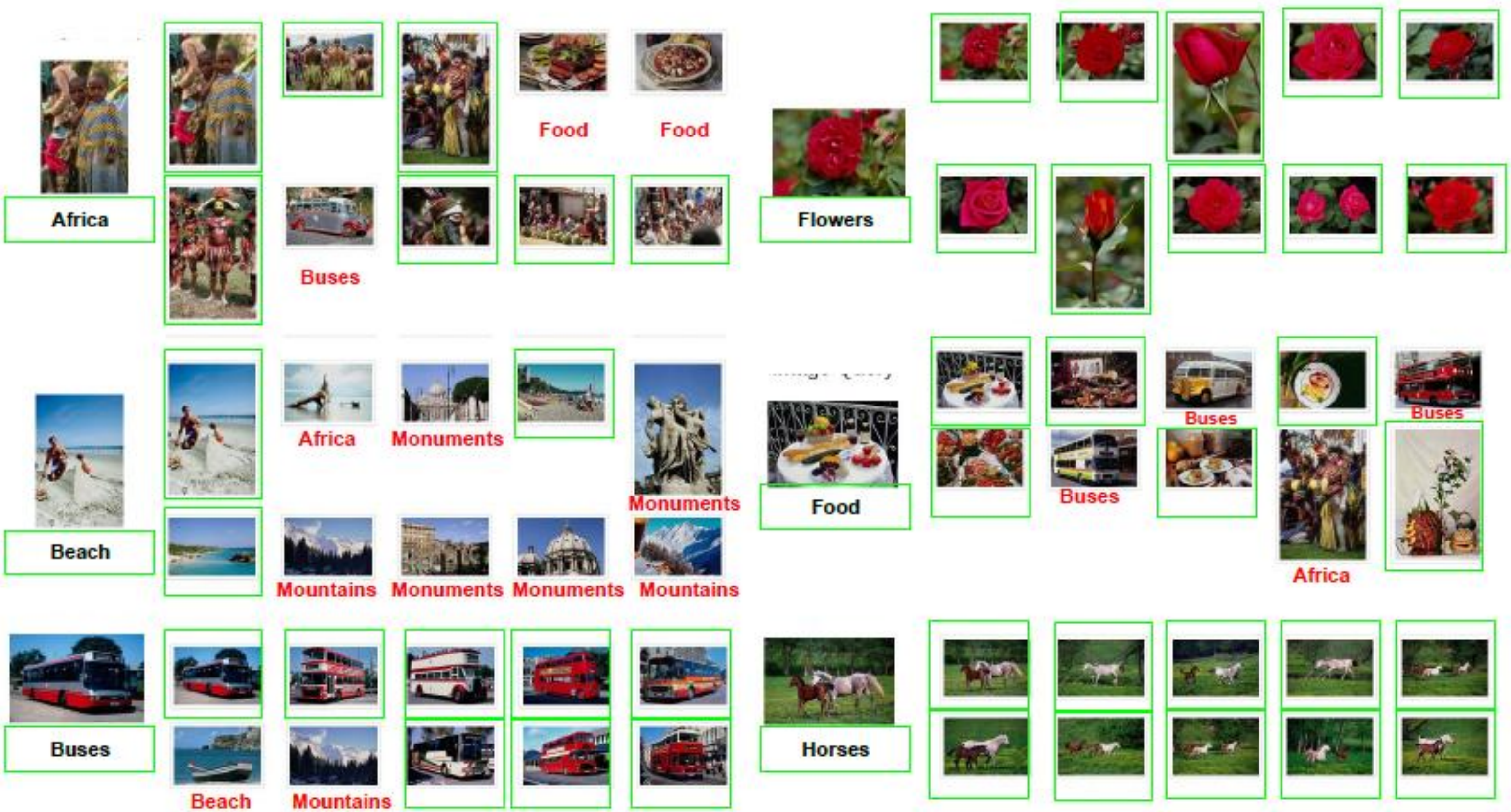
Show 10 Search:

Metric	Value
Mean Average Precision (MAP)	0.1882
Utility (a,b,c,d) Coefficients 1.0_-1.0_0.0_0.0	null
Precision after 10 images retrieved (P10)	0.6
Precision after 15 images retrieved (P15)	0.4
Precision after 20 images retrieved (P20)	0.3
Precision after 30 images retrieved (P30)	0.2667
Precision after 100 images retrieved (P100)	0.13
Precision after 200 images retrieved (P200)	0.085
Precision after 500 images retrieved (P500)	0.062
Precision after 1000 images retrieved (P1000)	0.055

Showing 1 to 10 of 148 records Pages: Previous 1 2 3 ... 15 Next



VPR Retriever System
http://goldenretriever.dcc.ufmg.br



Contributions

1. A comparative study of Low-Cost Representation for Mobile Image Search

2. We propose two new bag of visual words representations that include spatial information

Spatial Feature Representation for Mobile Image Search

- Objective:
  - Extracting spatial information from images to improve the quality of image representation on mobile devices
  - We proposed two new bag-of-visual words-based approaches to encode spatial information
- Bag-of-visual words representations
  - Traditional pooling methods usually discard the spatial configuration
  - This information is important to distinguish types of object and arrangements in the image

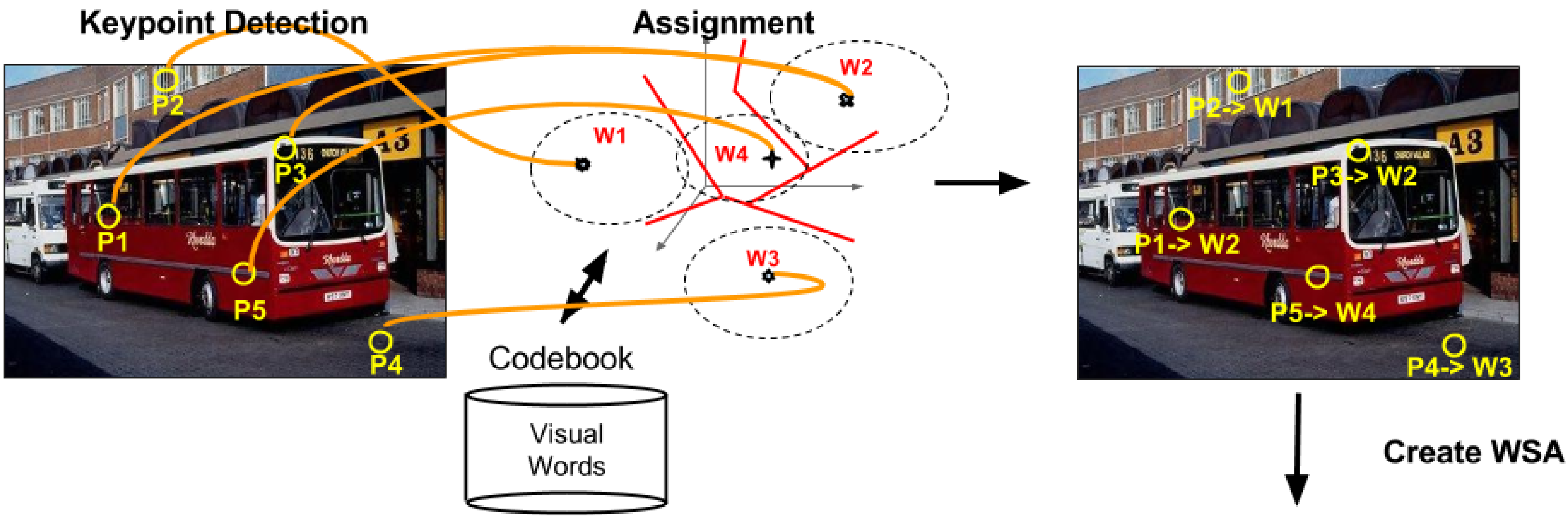
Spatial Feature Representation for Mobile Image Search

- We point out the BIC as one of the best descriptors analyzed
  - This descriptor is used to create representations of part of images
  - BIC representations are used on a mid-level strategy
- We have proposed two approaches
  - BOBGraph encodes relationships of nine fixed quadrants on the image
  - BOBSlic encodes relationships of homogenious regions on the images
- Baselines:
  - Word spatial arrangement (WSA) [Penatti et al., 2011a]
  - Bag Of Statistical Sampling Analysis (BOSSANova) [Avila et al., 2013]

Spatial Feature Representation for Mobile Image Search

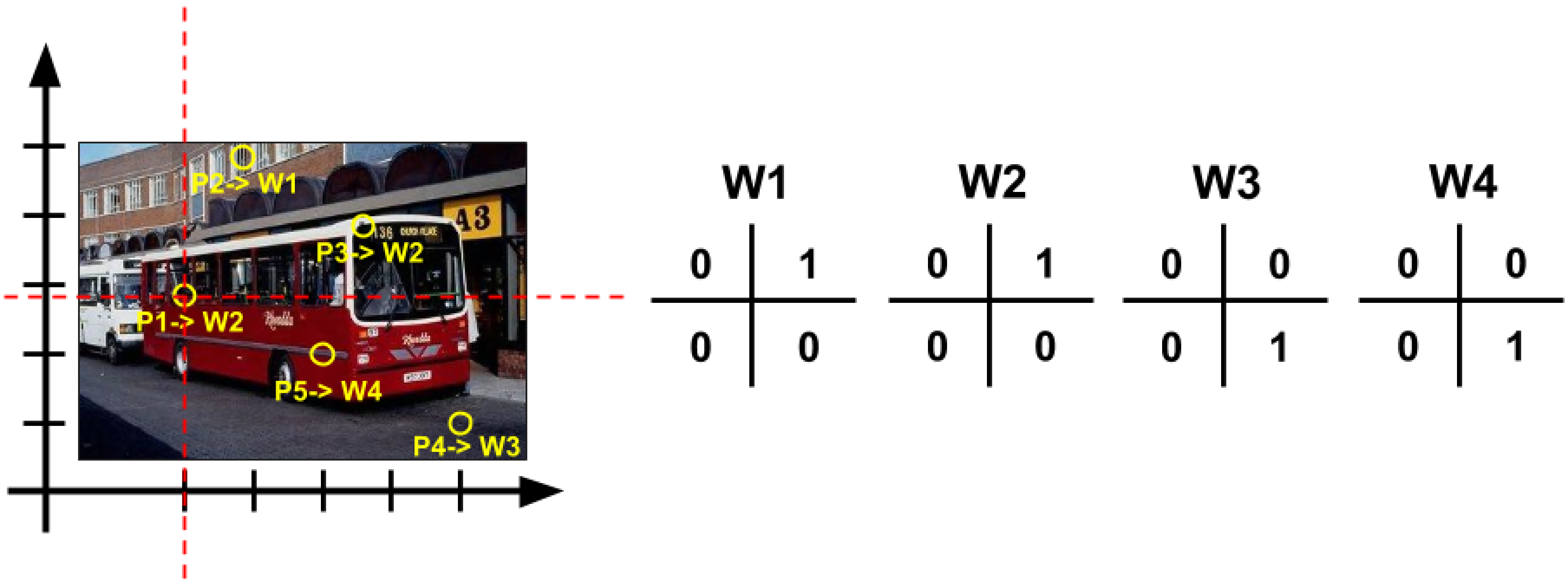
- BOBGraph and BOBSlic – Objectives
  - Using the location of objects in an image to improve their mid-level representation
  - We use the relationships of objects in the image
  - Word spatial arrangement (WSA) is our main baseline
  - **Generic to apply any descriptor in the regions**
- Baselines:
  – Word spatial arrangement (WSA) [Penatti et al., 2011a]
  – Bag Of Statistical Sampling Analysis (BOSSANova) [Avila et al., 2013]

Word spatial arrangement (WSA)

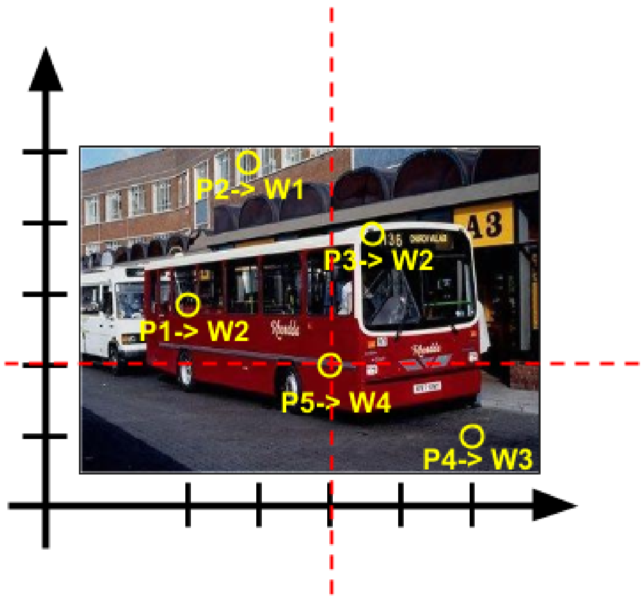


- Our baseline of Spatial Feature Representation

Word spatial arrangement (WSA)



Word spatial arrangement (WSA)



W1

3	1
0	0

sum(W1) = 4

W2

3	2
2	1

sum(W1) = 8

W3

0	0
0	4

sum(W1) = 4

W4

1	0
1	2

sum(W1) = 4

$\underbrace{\hspace{10em}}_{}$ **Normalization**

W1

0.75	0.25
0	0

W2

0.375	0.25
0.25	0.125

W3

0	0
0	1.0

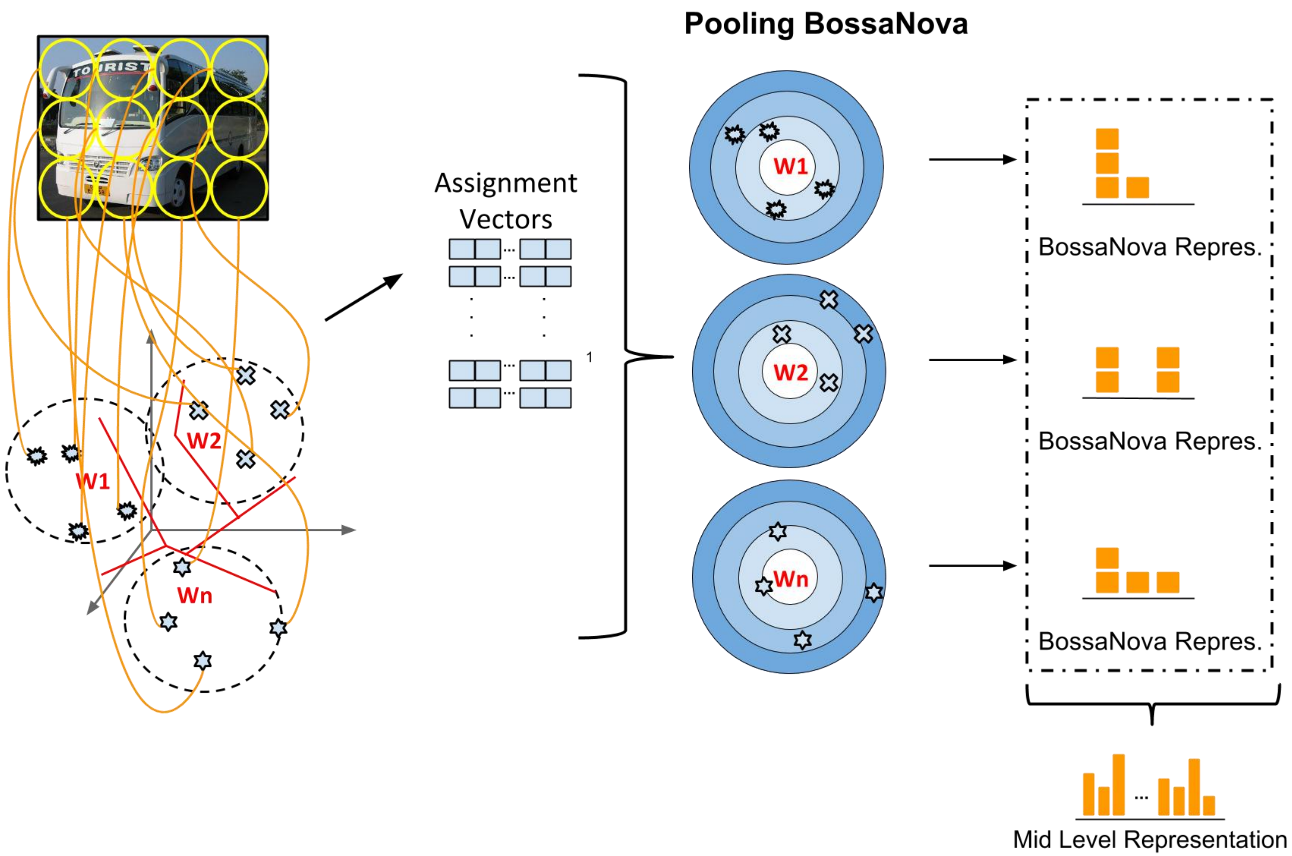
W4

0.25	0
0.25	0.5

WSA Representation

0.75	0.25	0.0	0.0	0.375	0.25	...	0.0	1.0	0.25	0.0	0.25	0.5

Bag Of Statistical Sampling Analysis (BOSSANova)

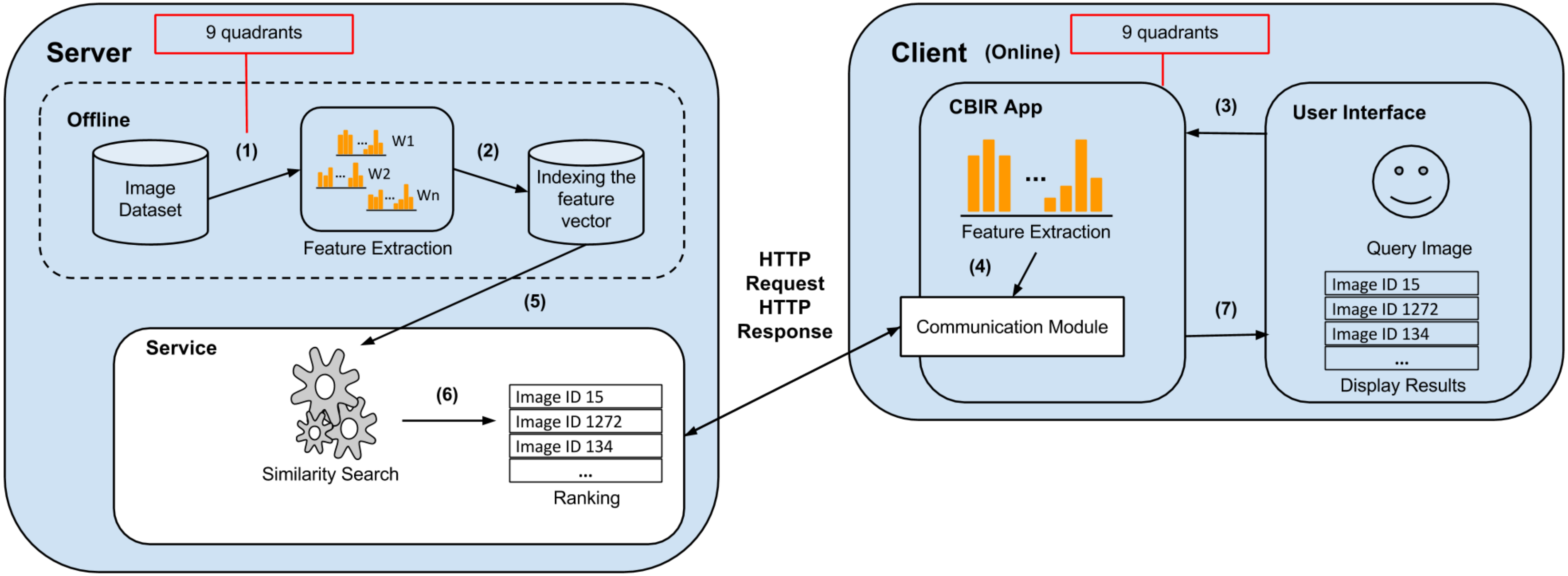


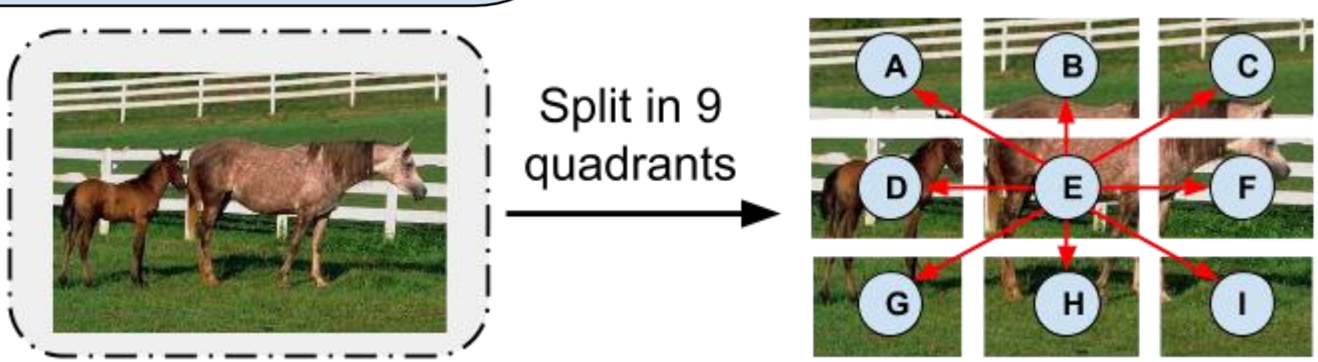
- BossaNova is not Spatial Feature Representation
- BossaNova is a baseline with an improvement in the bag of words

Spatial Feature Representation for Mobile Image Search

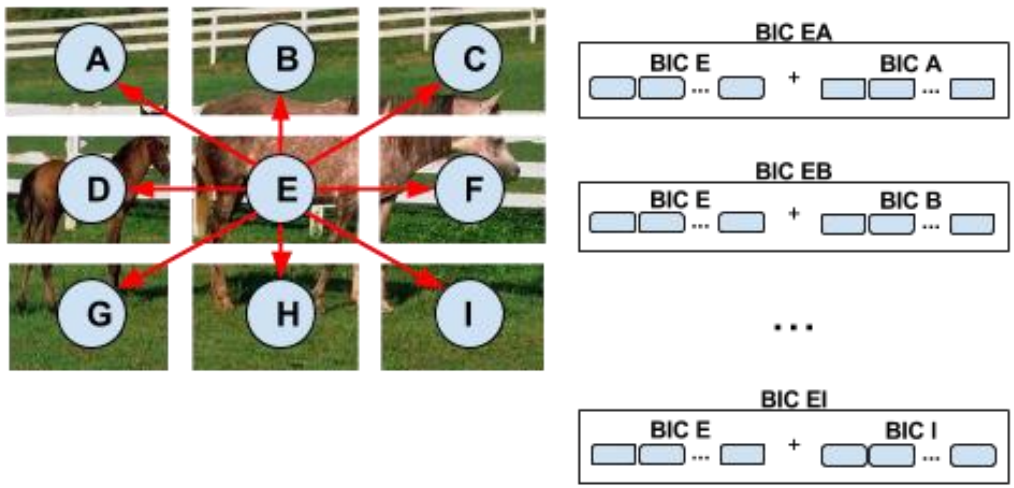
- **BoBGraph (Bag Of BIC Graph) Representation**
  - Split the image in nine similar quadrants
  - BIC descriptor (Most suitable descriptor)
  - Create eight edges
  - BIC Representations of Edges
  - Offline step
    - Nine splited parts of all images on the dataset to create a dictionary (or codebook) using **K-means algorithm**
  - Online step
    - Use the dictionary created to generate BoBGraph representations
  - Dictionary of **128 visual words**
    - Randomly selecting points in the feature space
  - WANG dataset

BoBGraph (Bag Of BIC Graph) Representation

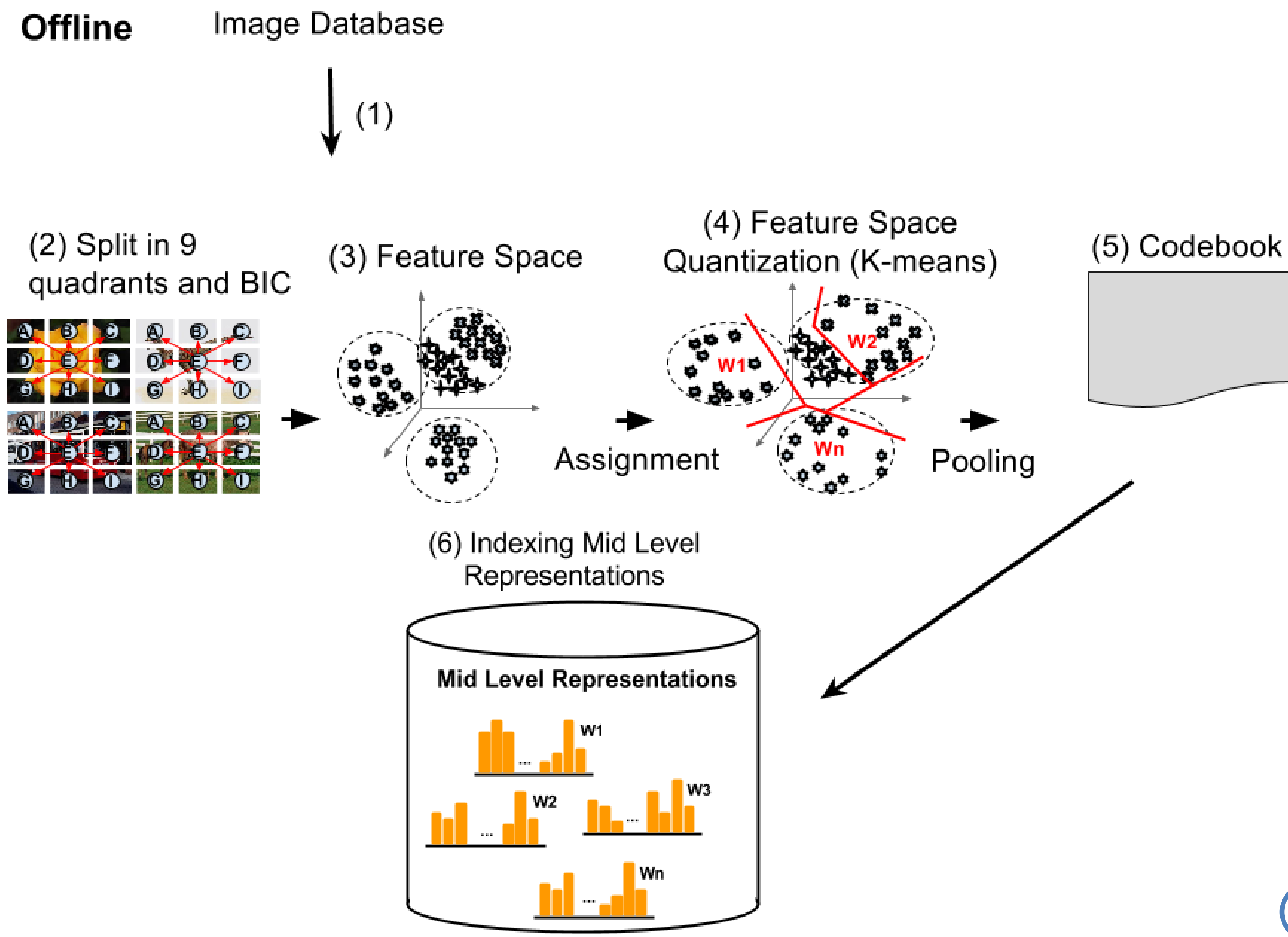




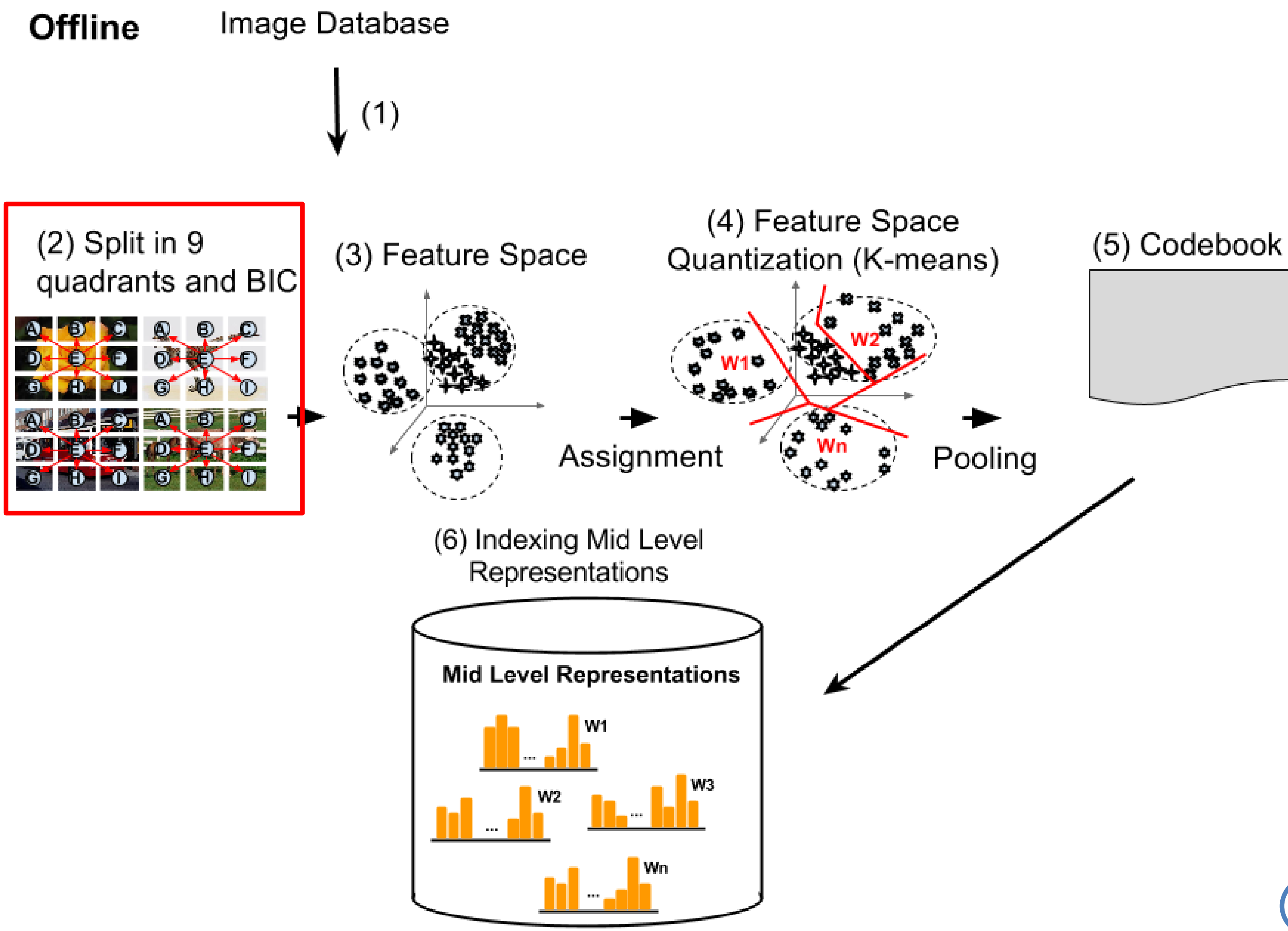
BoBGraph (Bag Of BIC Graph) Representation



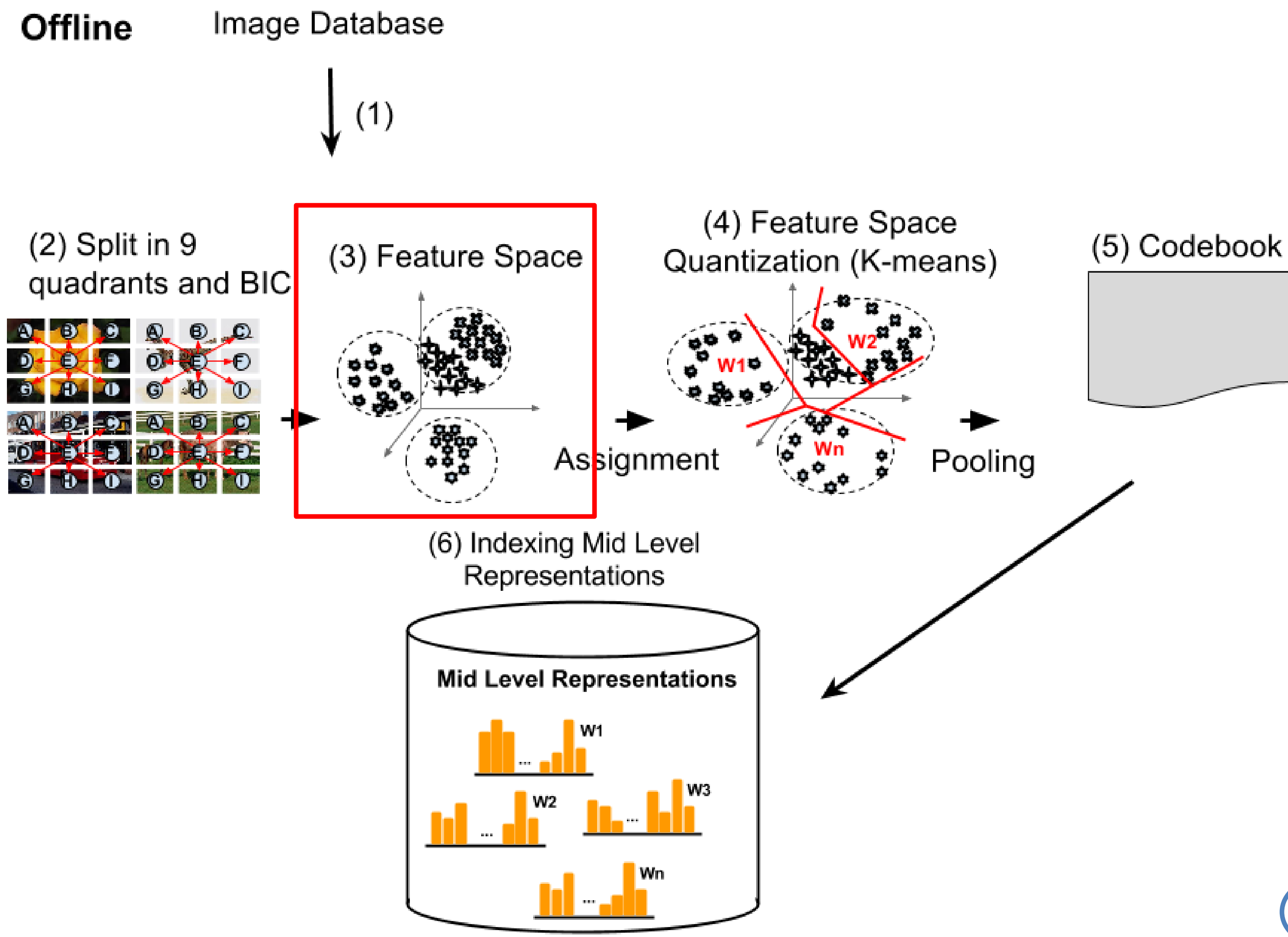
BoBGraph (Bag Of BIC Graph) Representation



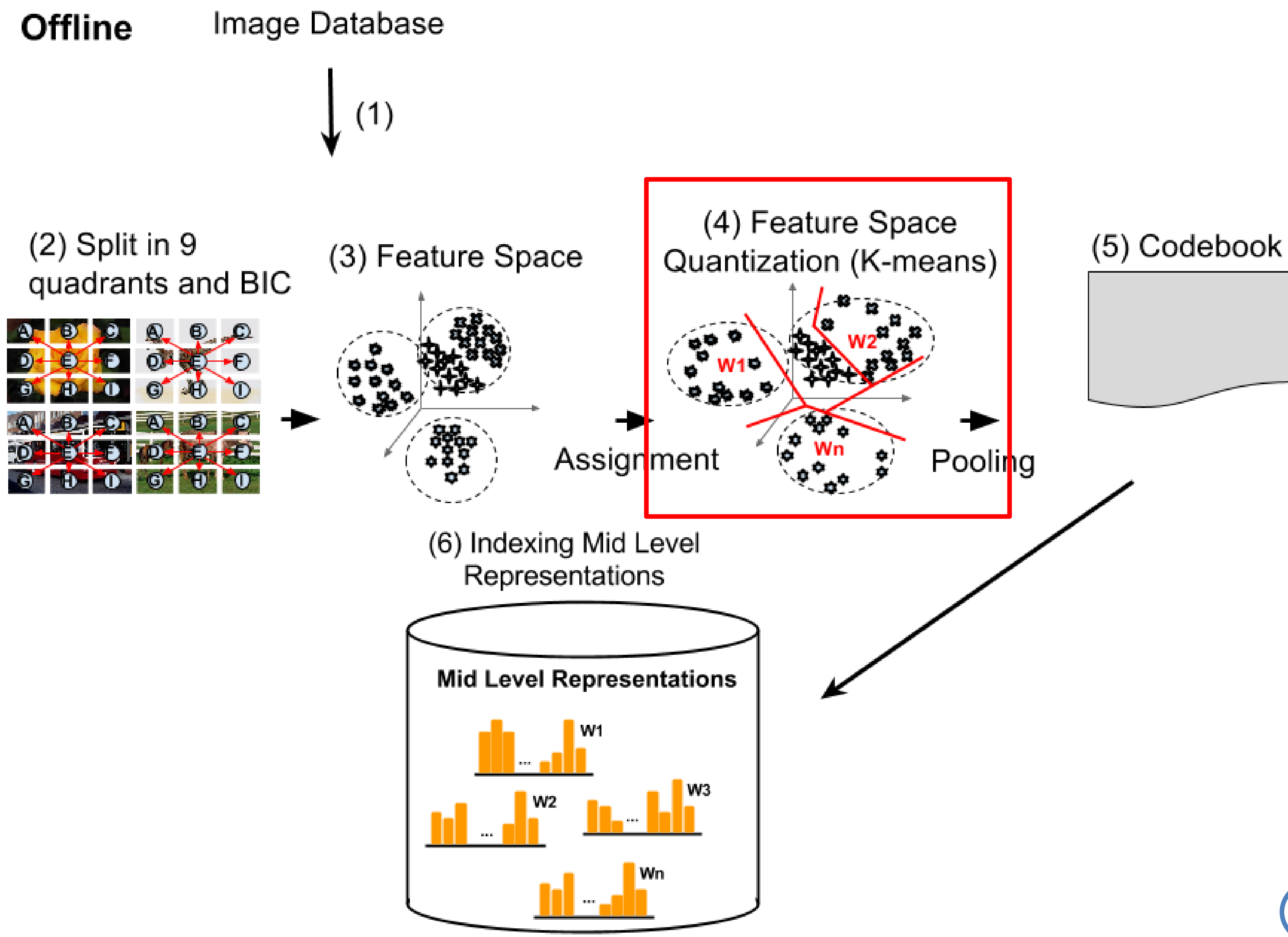
BoBGraph (Bag Of BIC Graph) Representation



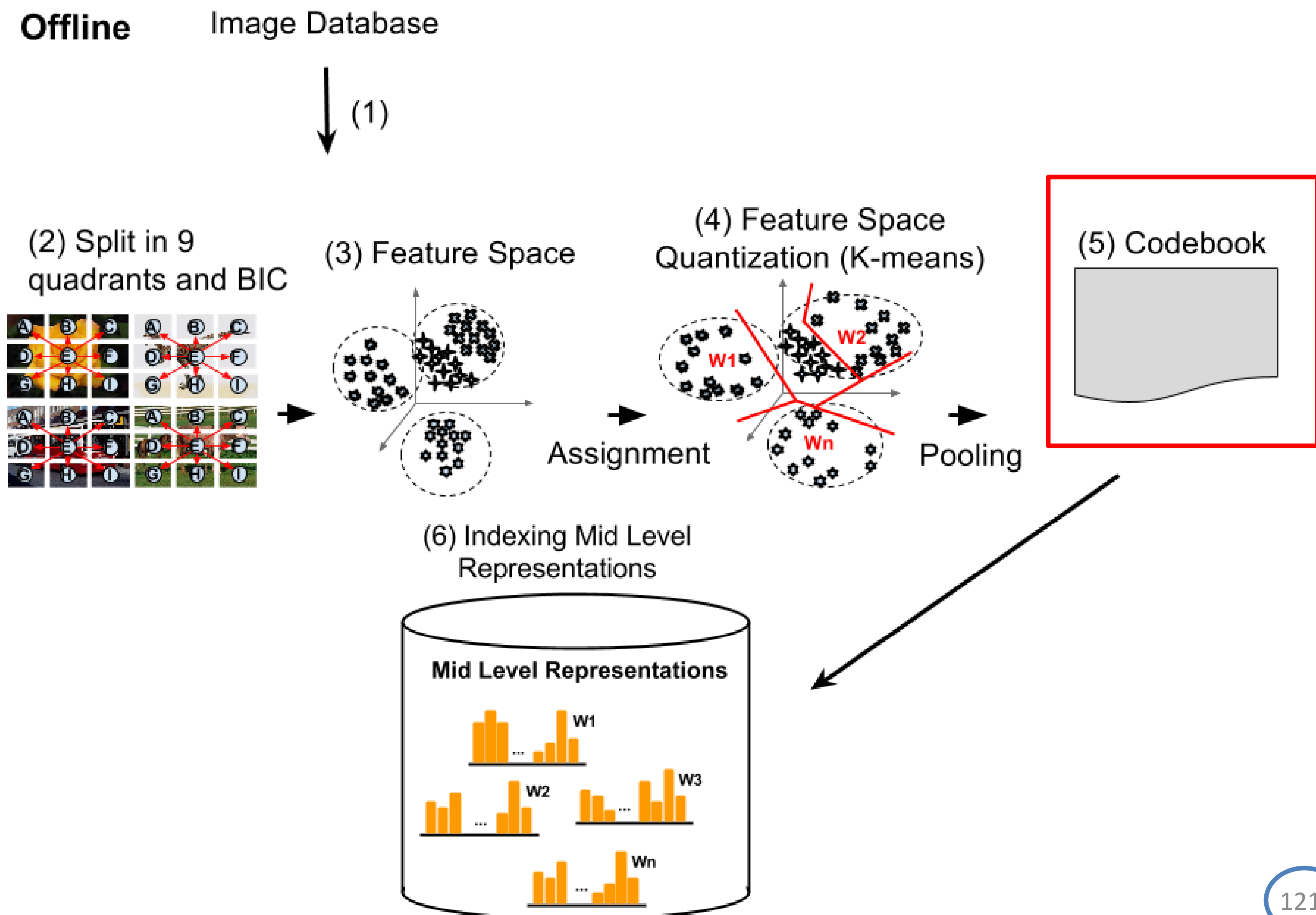
BoBGraph (Bag Of BIC Graph) Representation



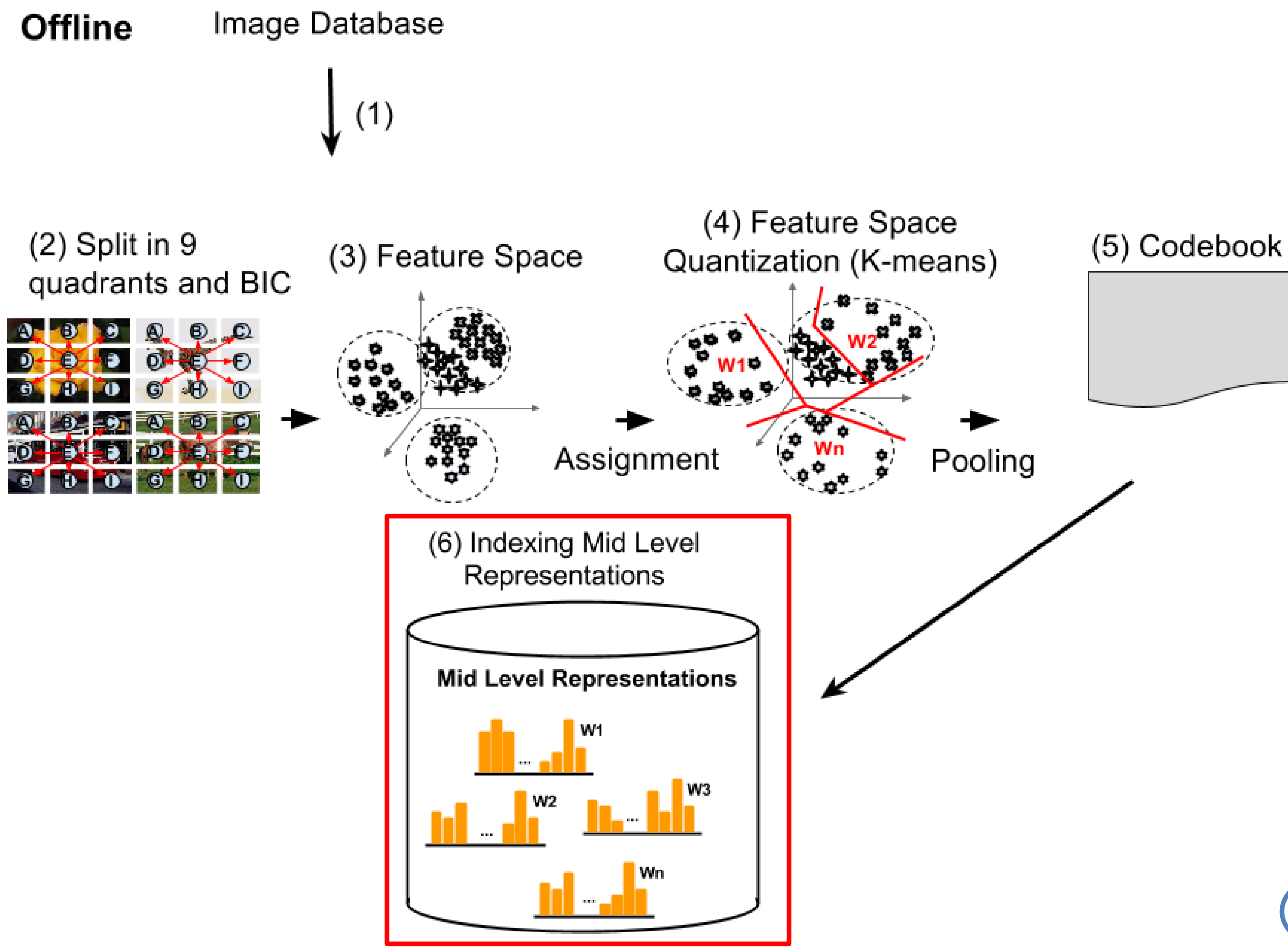
BoBGraph (Bag Of BIC Graph) Representation



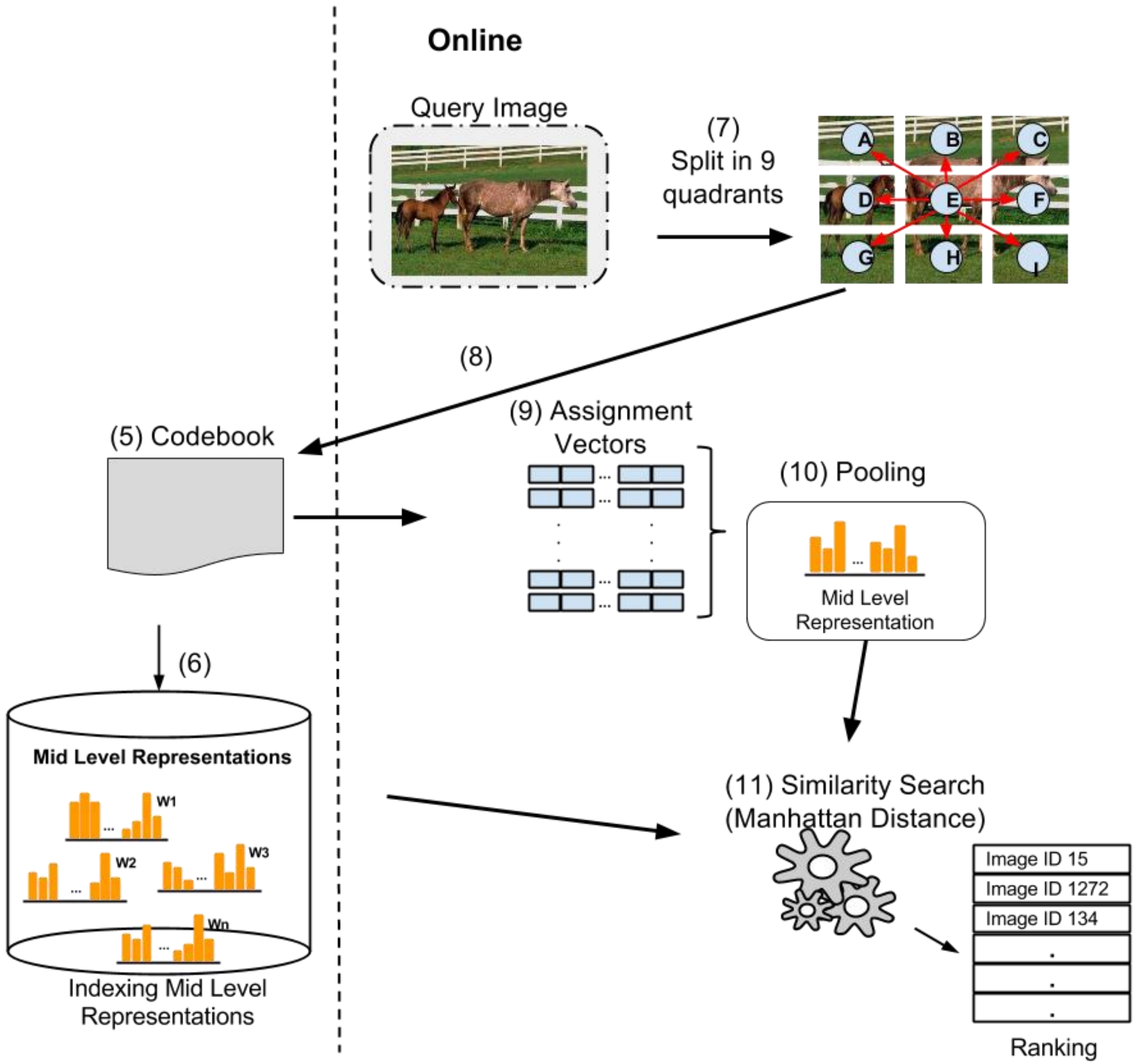
BoBGraph (Bag Of BIC Graph) Representation



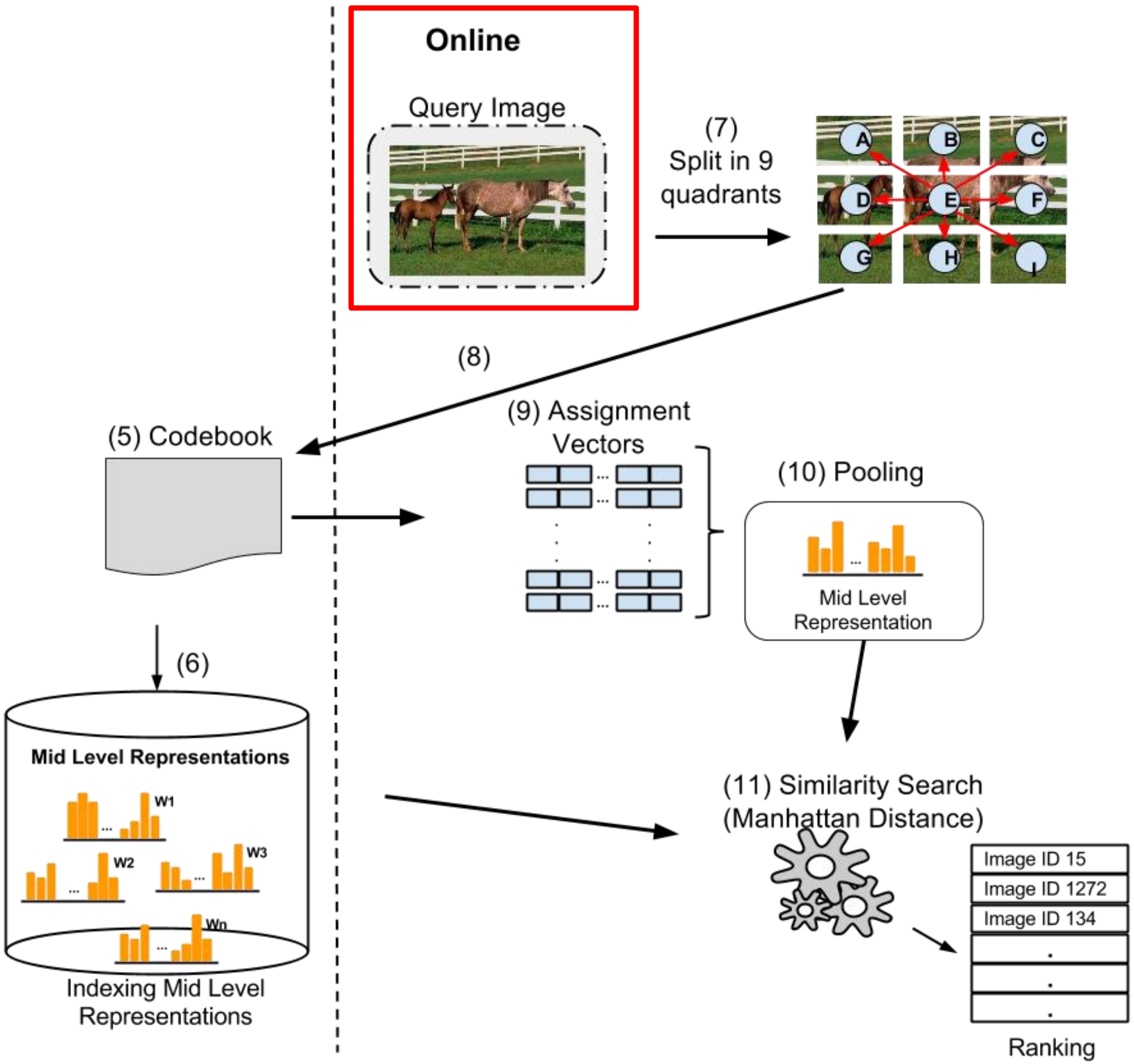
BoBGraph (Bag Of BIC Graph) Representation



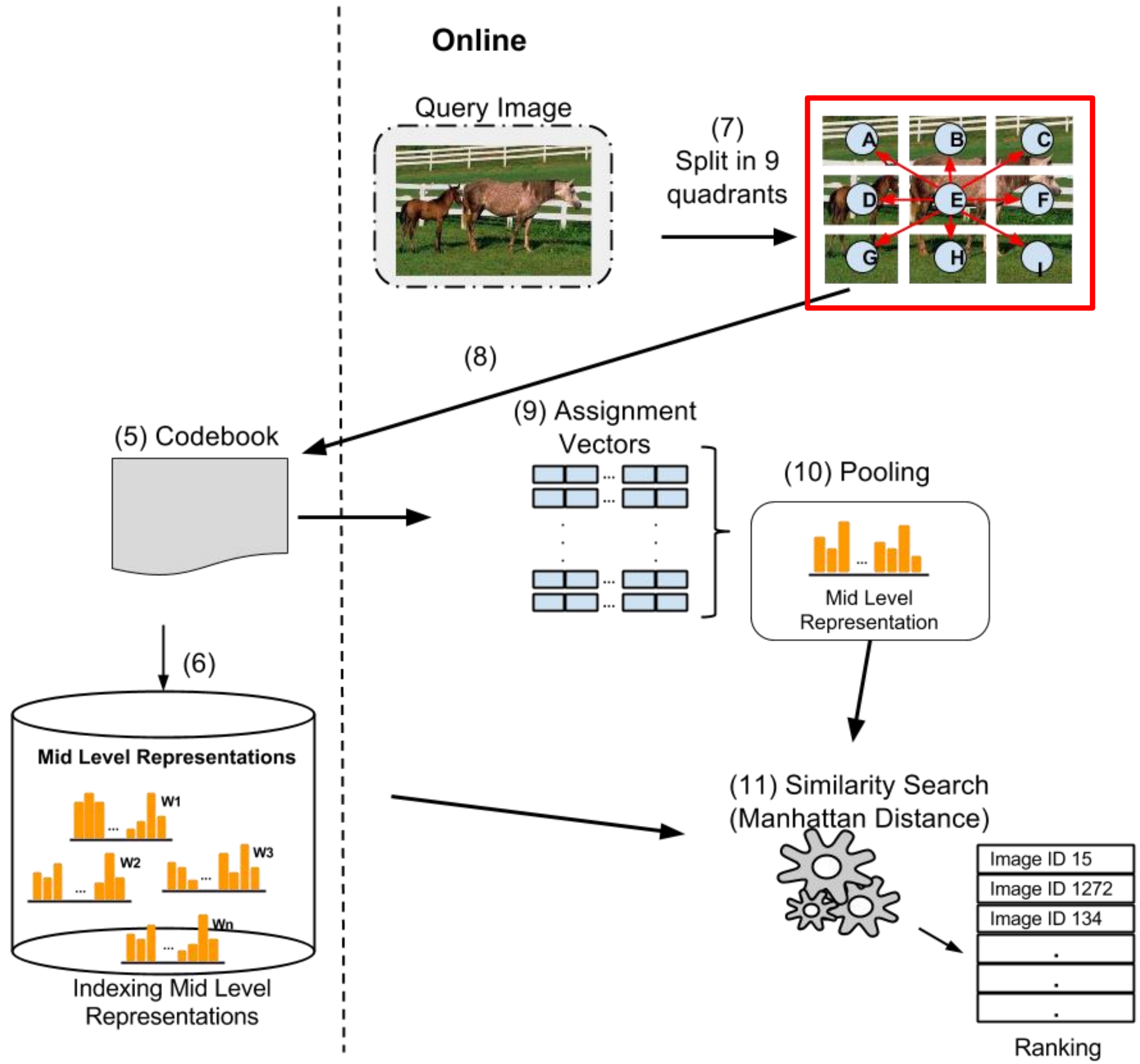
BoBGraph (Bag Of BIC Graph) Representation



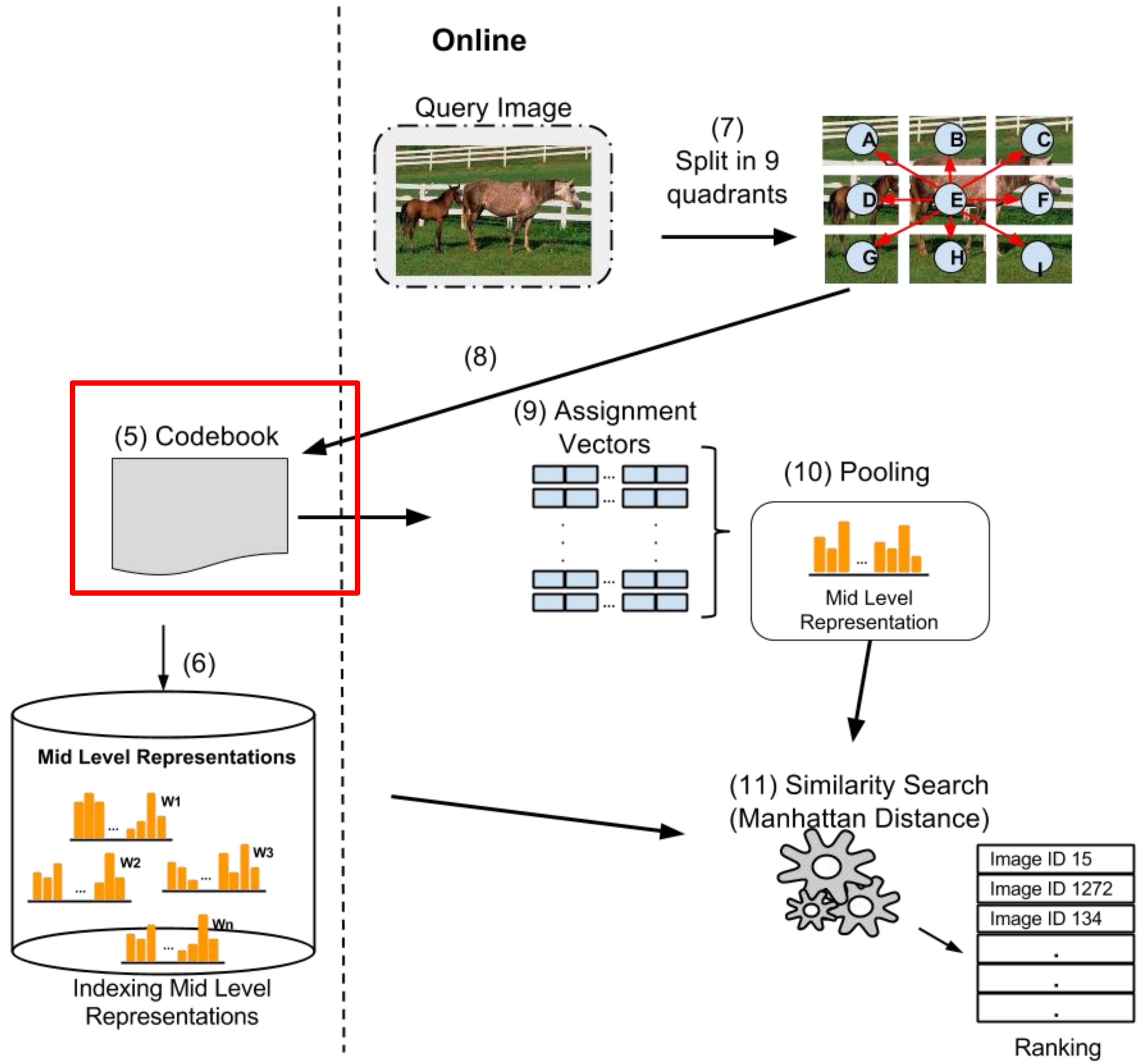
BoBGraph (Bag Of BIC Graph) Representation



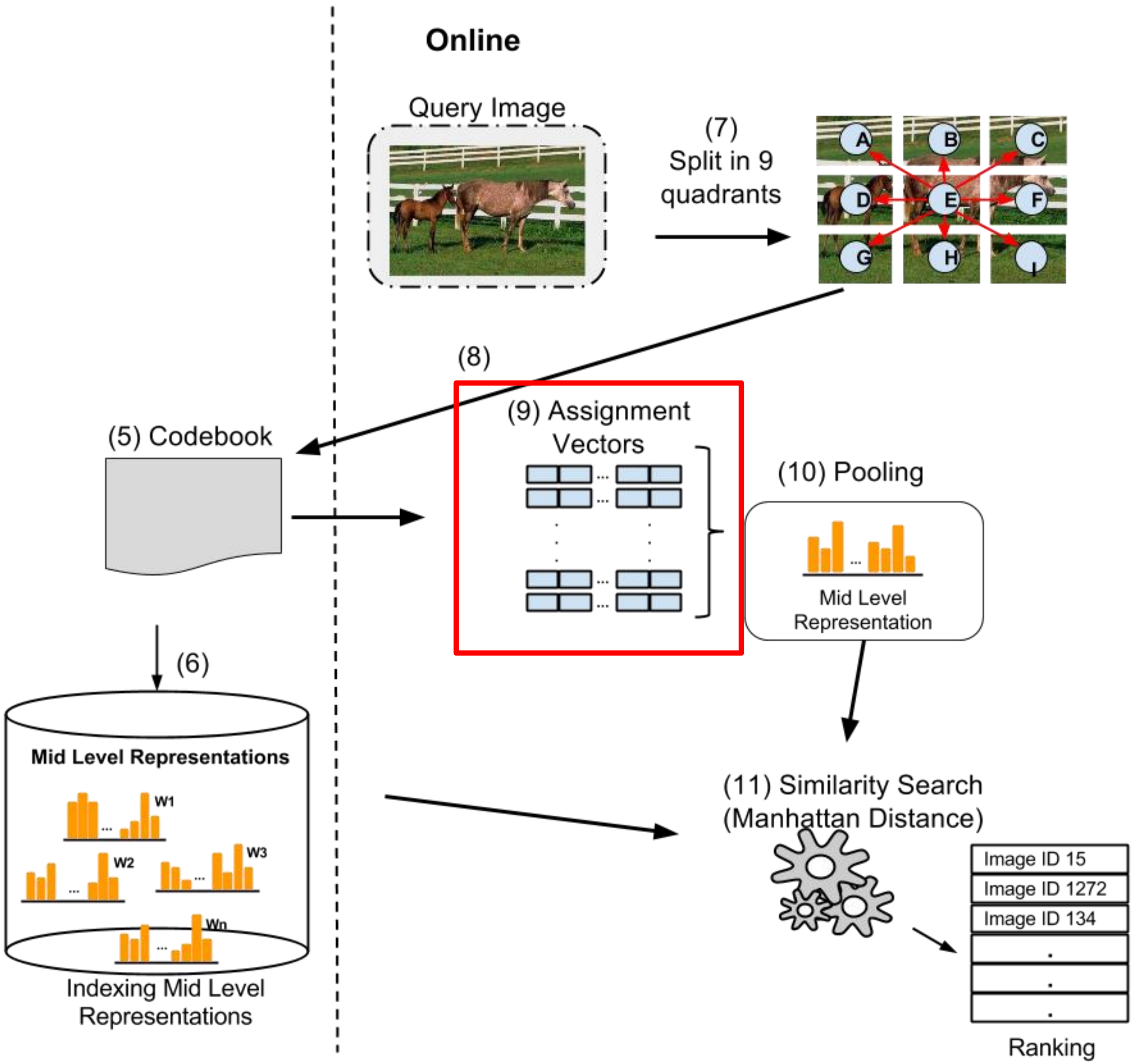
BoBGraph (Bag Of BIC Graph) Representation



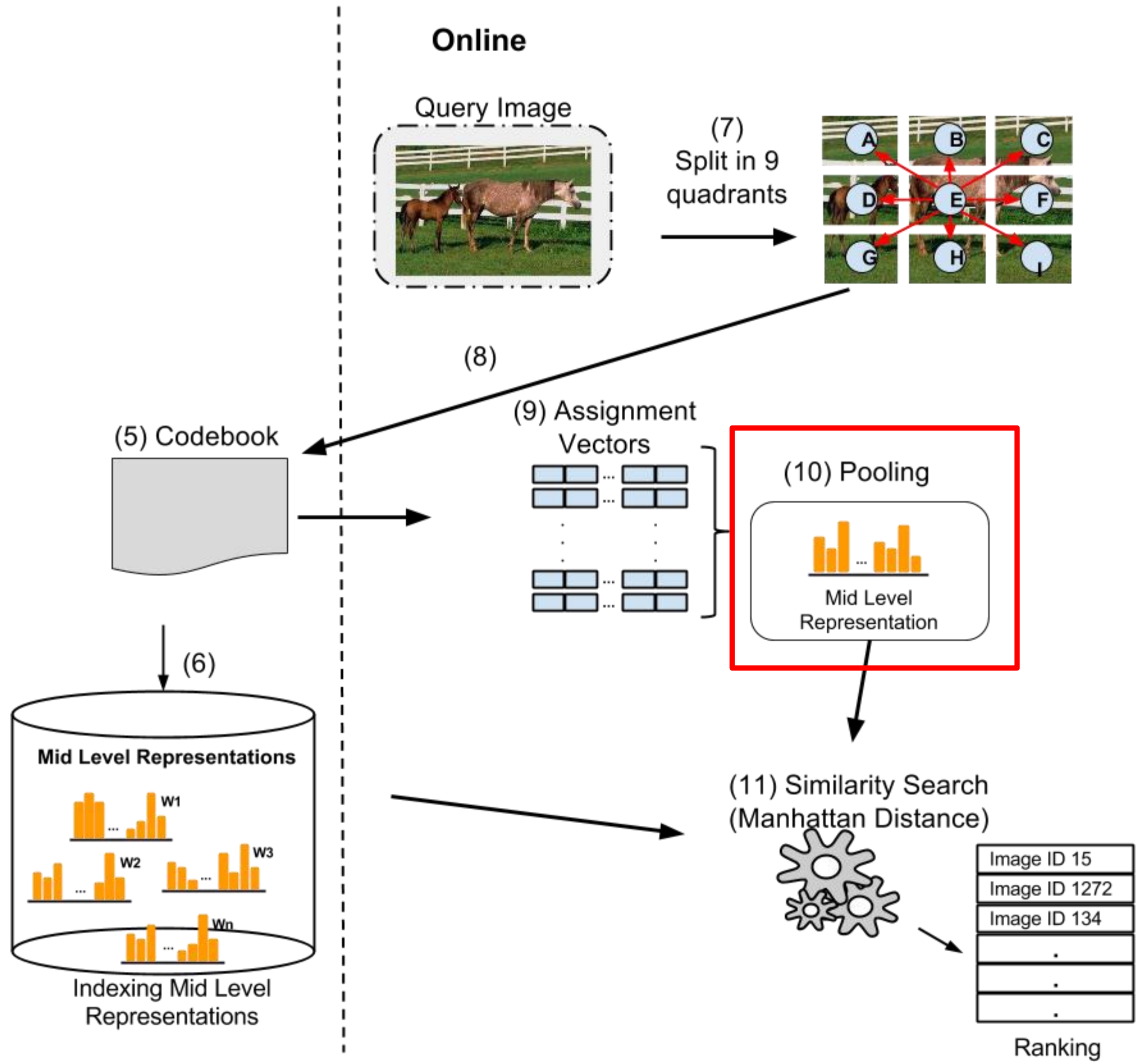
BoBGraph (Bag Of BIC Graph) Representation



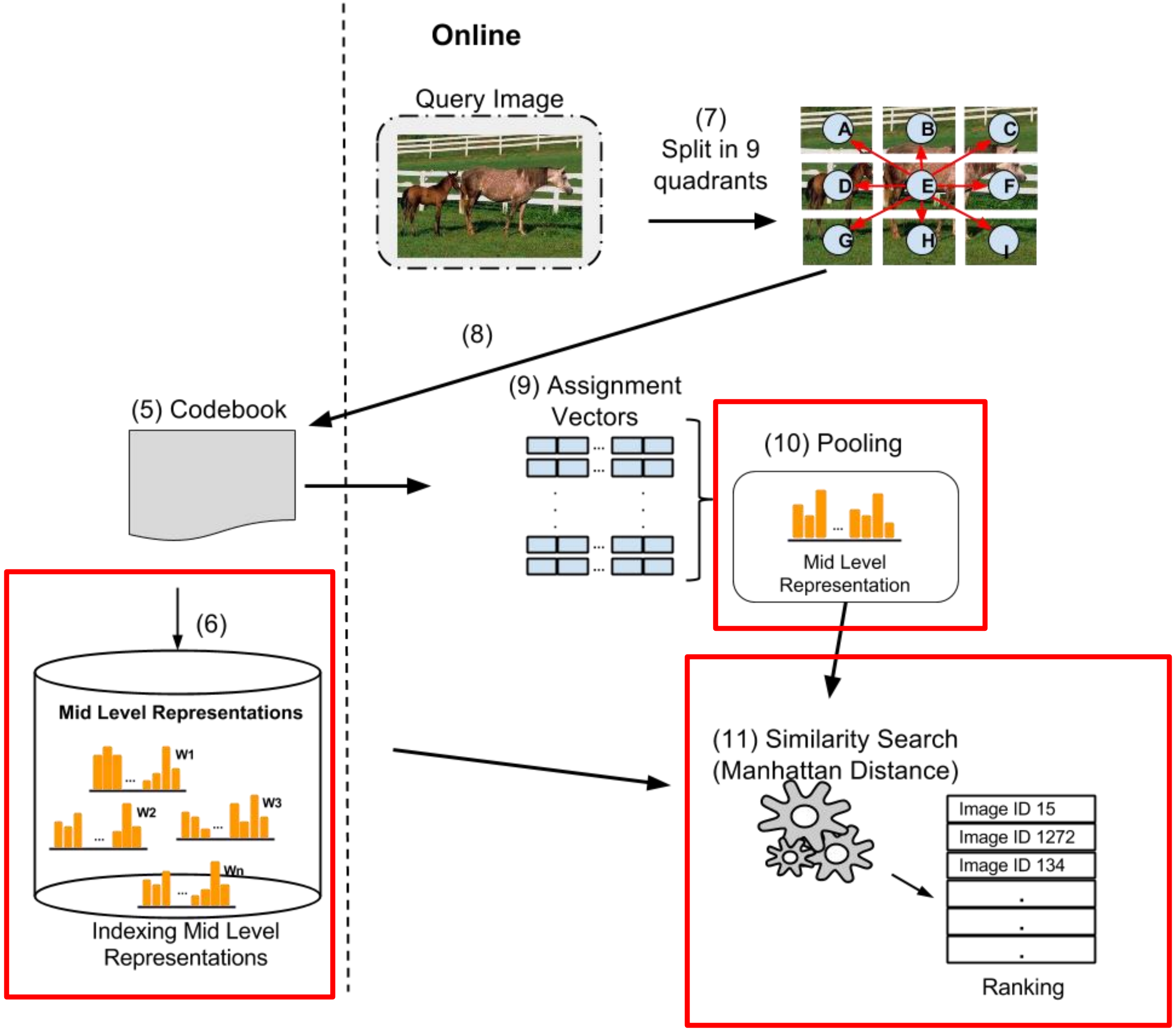
BoBGraph (Bag Of BIC Graph) Representation



BoBGraph (Bag Of BIC Graph) Representation



BoBGraph (Bag Of BIC Graph) Representation



Spatial Feature Representation for Mobile Image Search

- **BoBSlic (Bag Of Slic) Representation**
  - Use the SLIC algorithm [Achanta et al., 2012] to segment the image
  - BIC descriptor (Most suitable descriptor)
  - We tested several values for the parameters
    - Number of superpixels n = 10, 20 or 30
    - Compactness c = 0, 5, 20, 50
  - The best results were obtained with n = 10 and c = 50
  - After SLIC, we compute BIC Representations of Edges
  - Offline step
    - Segmented regions of all images on the dataset to create a dictionary (or codebook) using **K-means algorithm**
  - Online step
    - Use the dictionary created to generate BoBSlic representations
  - Dictionary of **128 visual words**
    - Randomly selecting points in the feature space
  - WANG dataset

Image segmentation

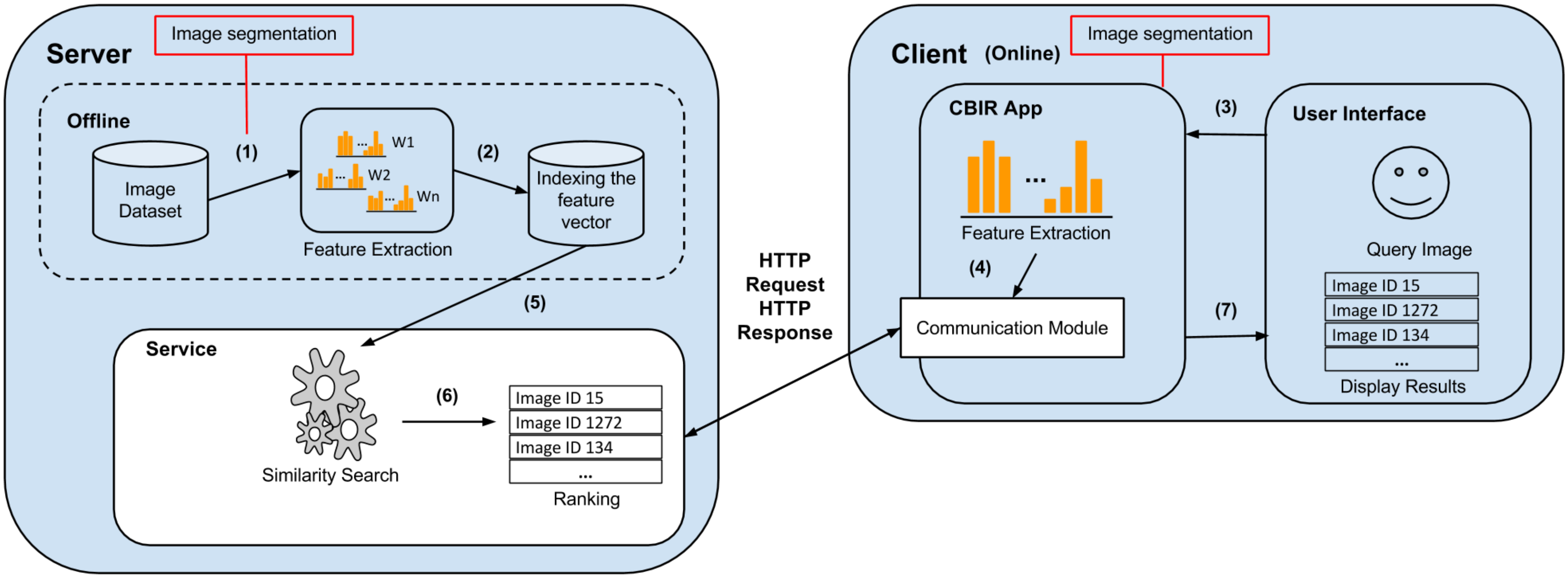
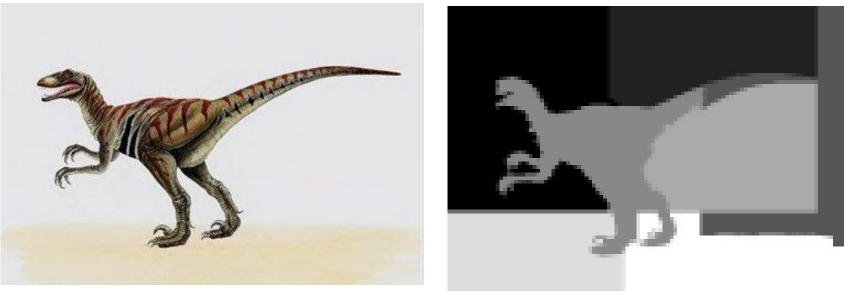


Image segmentation

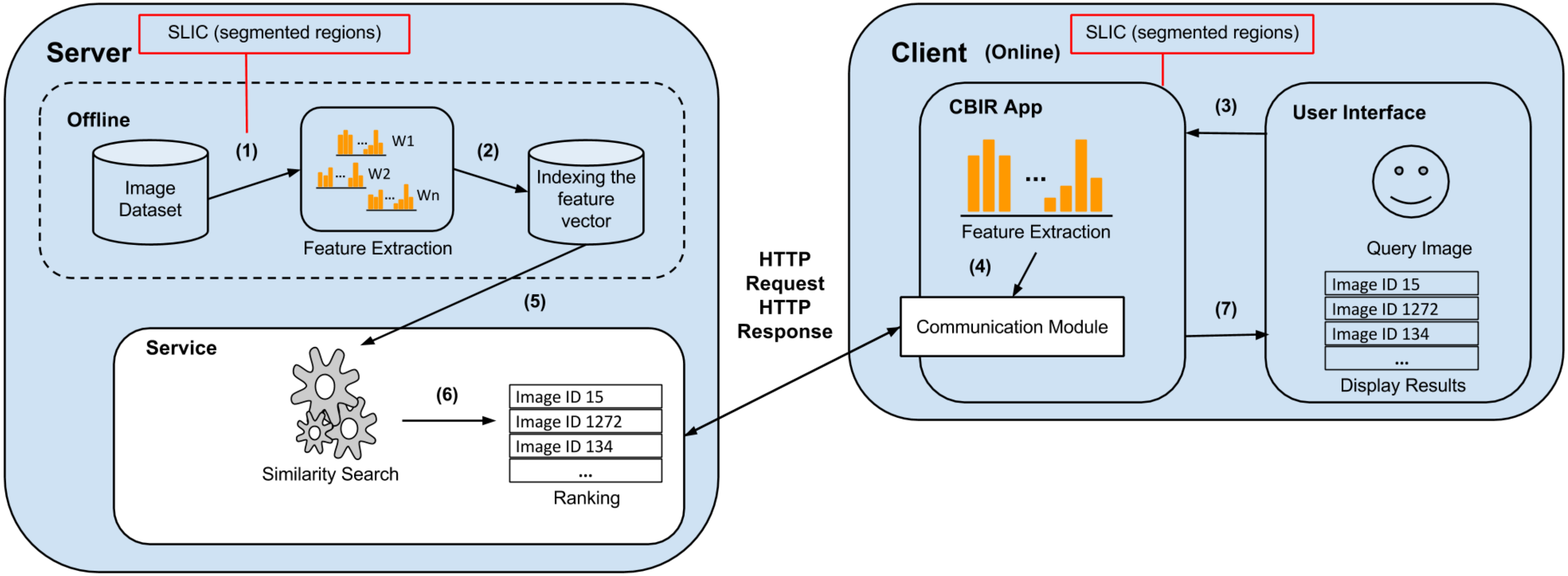
- Simple Linear Iterative Clustering (SLIC) [Achanta et al., 2012]
  - To efficiently generate superpixels
  - Performs K-means in the 5d space of color information and image location
  - SLIC has two parameters
    - n = Number of superpixels
    - c = Compactness of superpixels

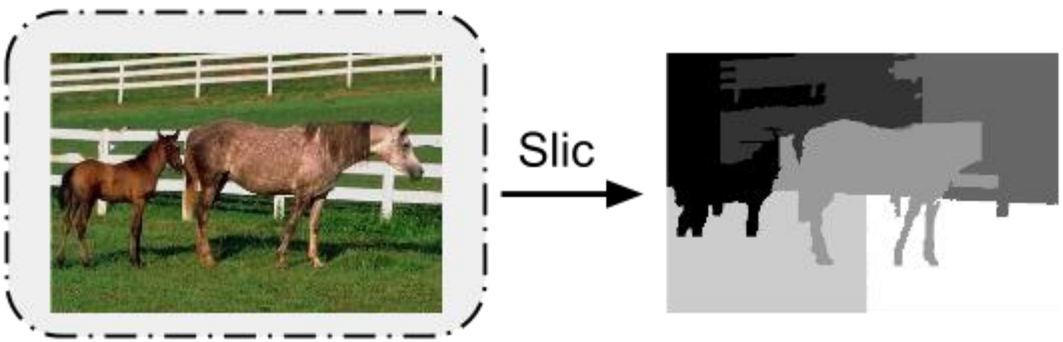


Simple Linear Iterative Clustering (SLIC)

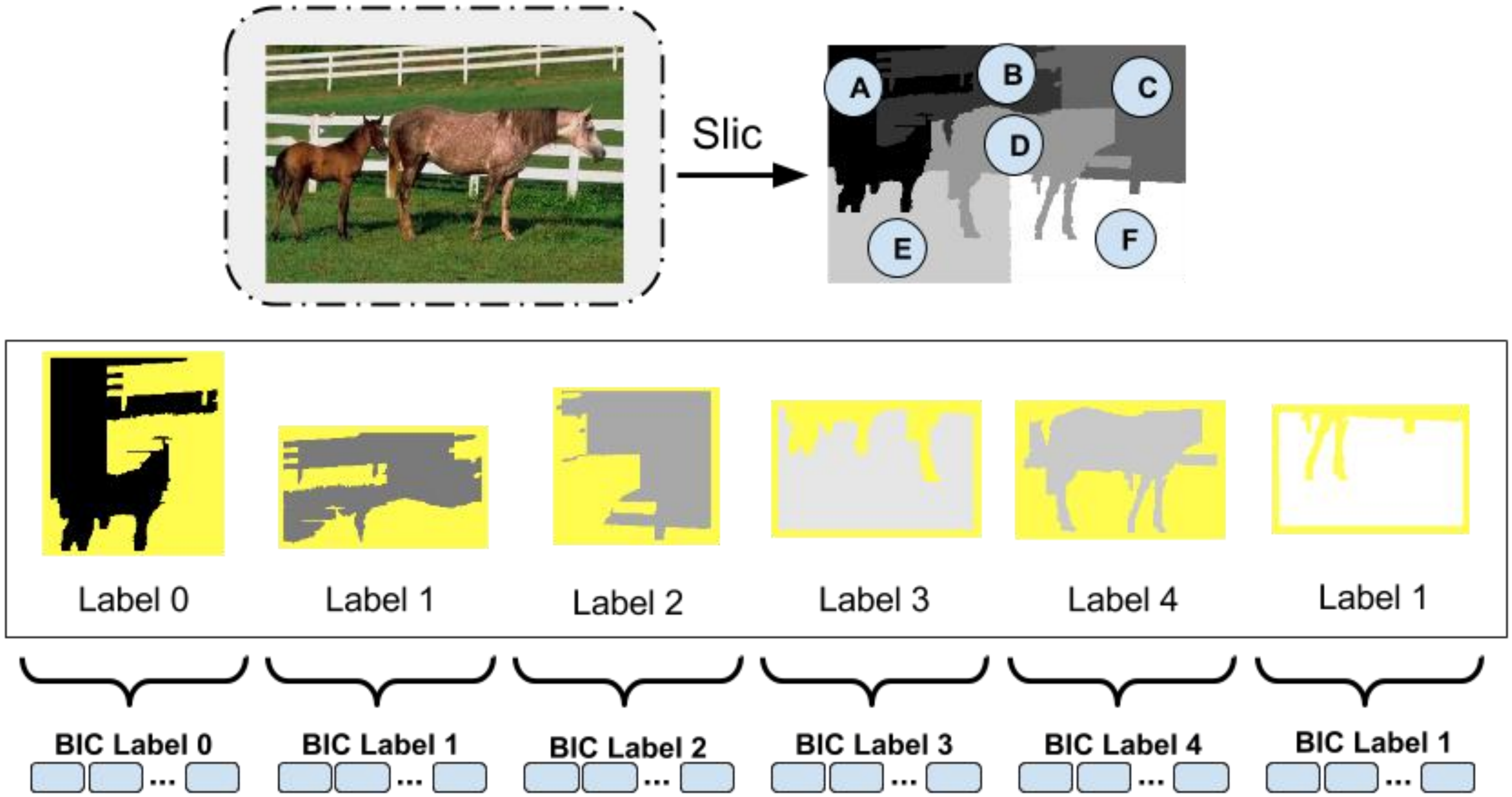
Simple Linear Iterative Clustering (SLIC)

BoBSlic (Bag Of Slic) Representation

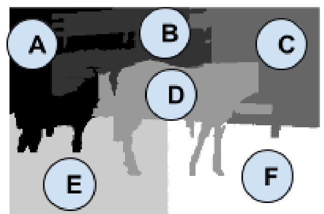


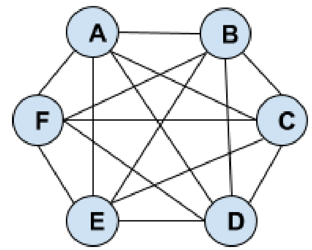


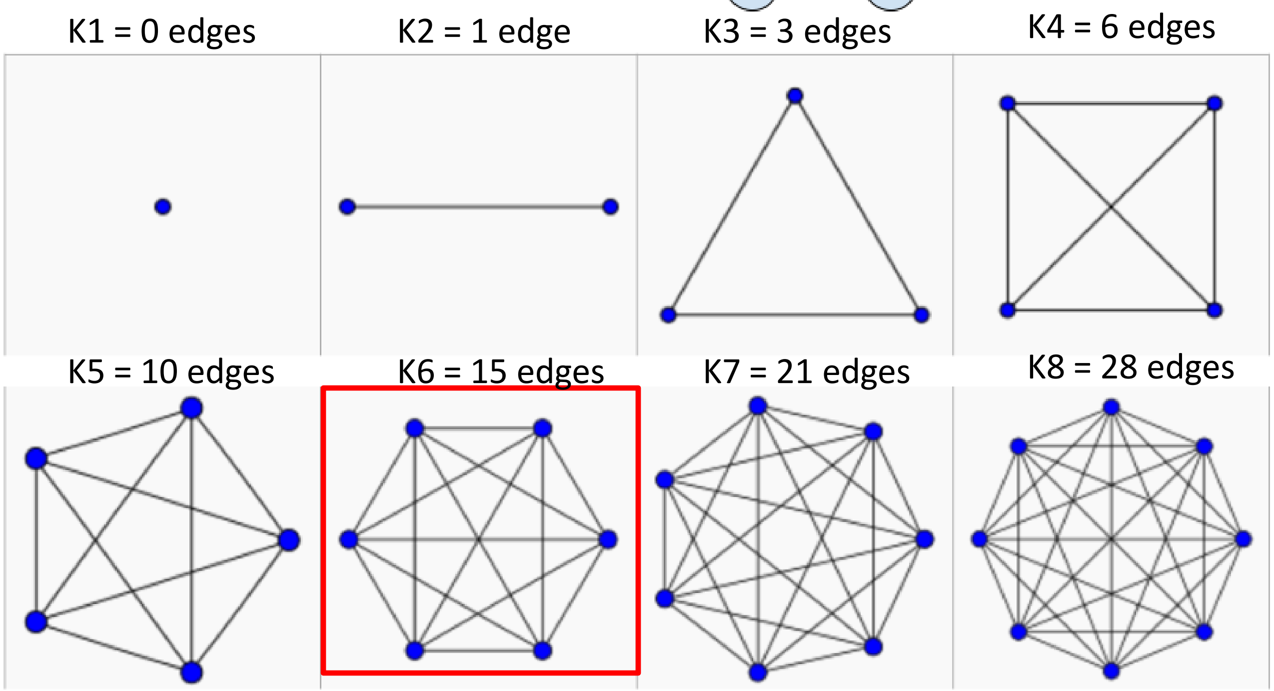
BoBSlic (Bag Of Slic) Representation



BoBSlic (Bag Of Slic) Representation

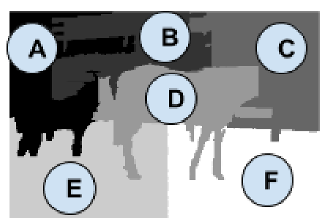


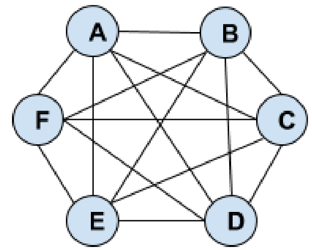


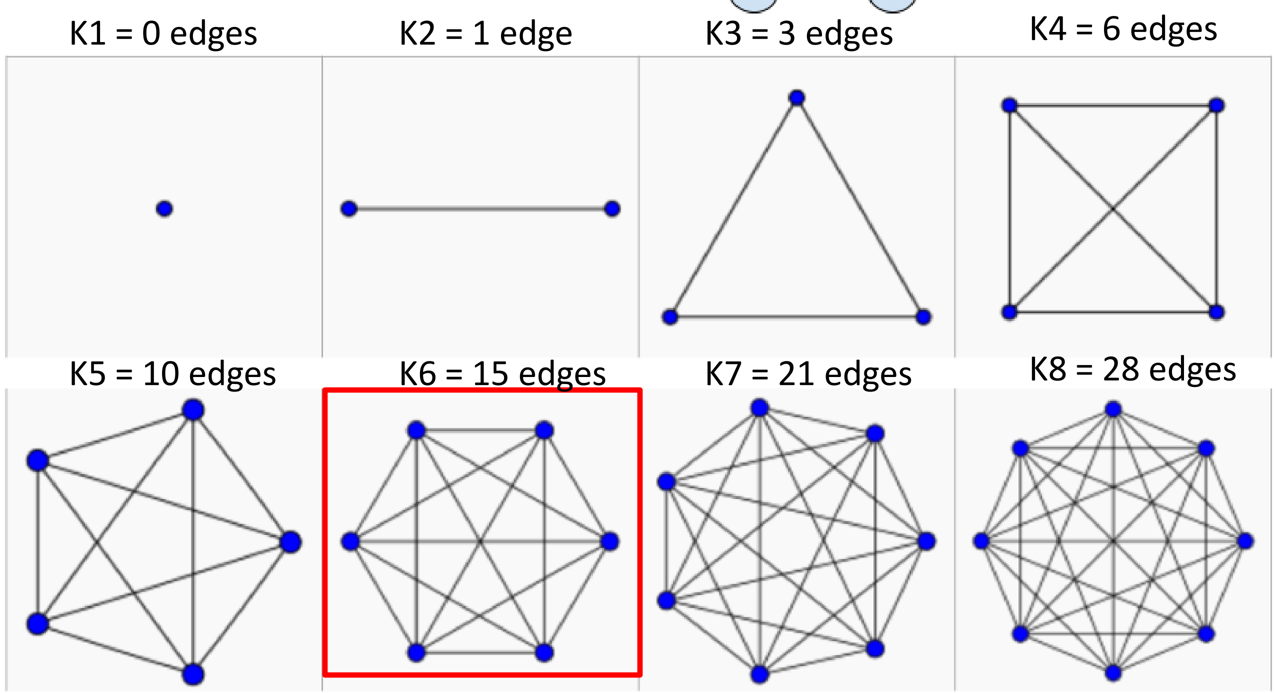


Number of edges: $K_n = \binom{n}{2} = \frac{n(n-1)}{2}$

BoBSlic (Bag Of Slic) Representation

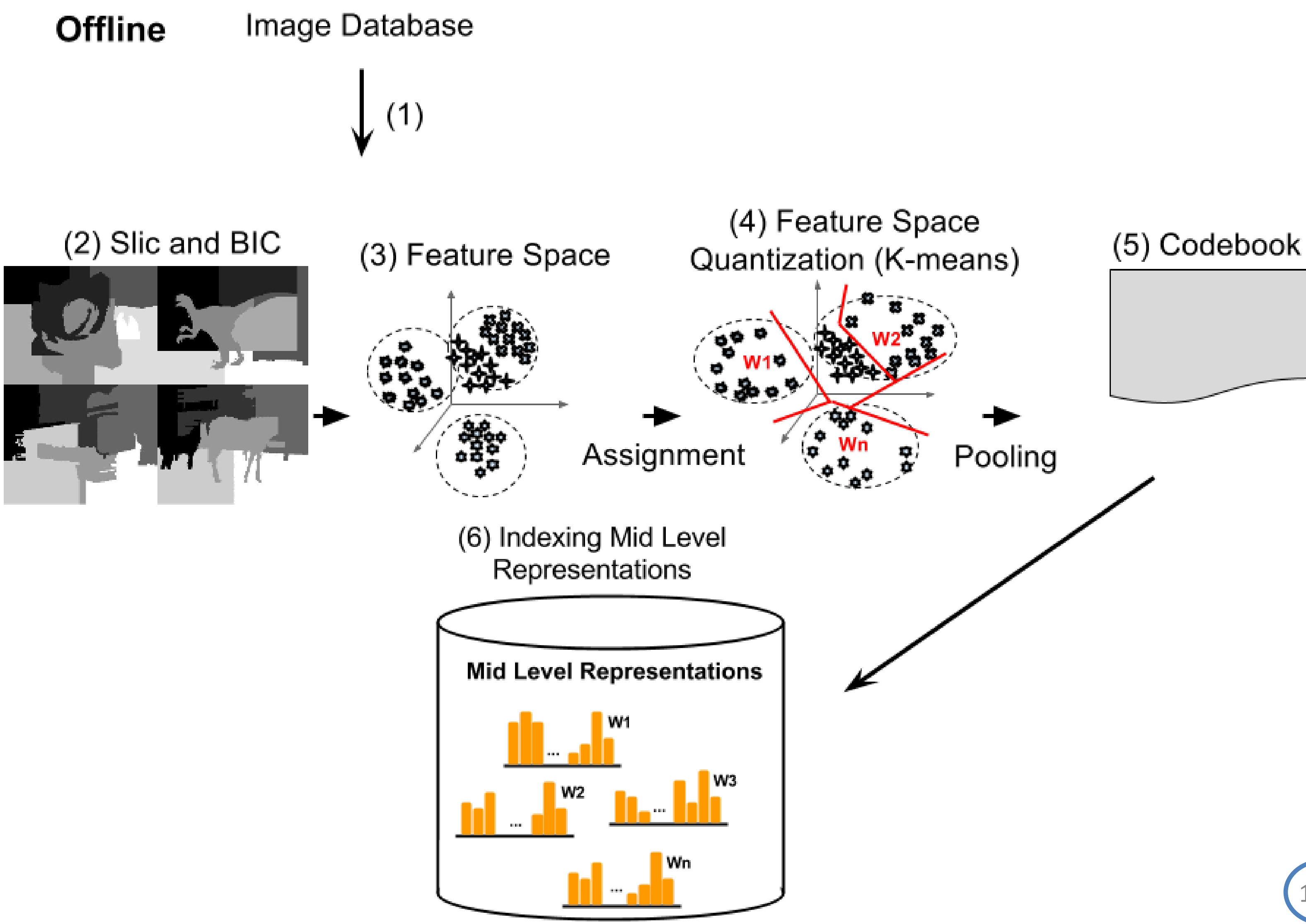




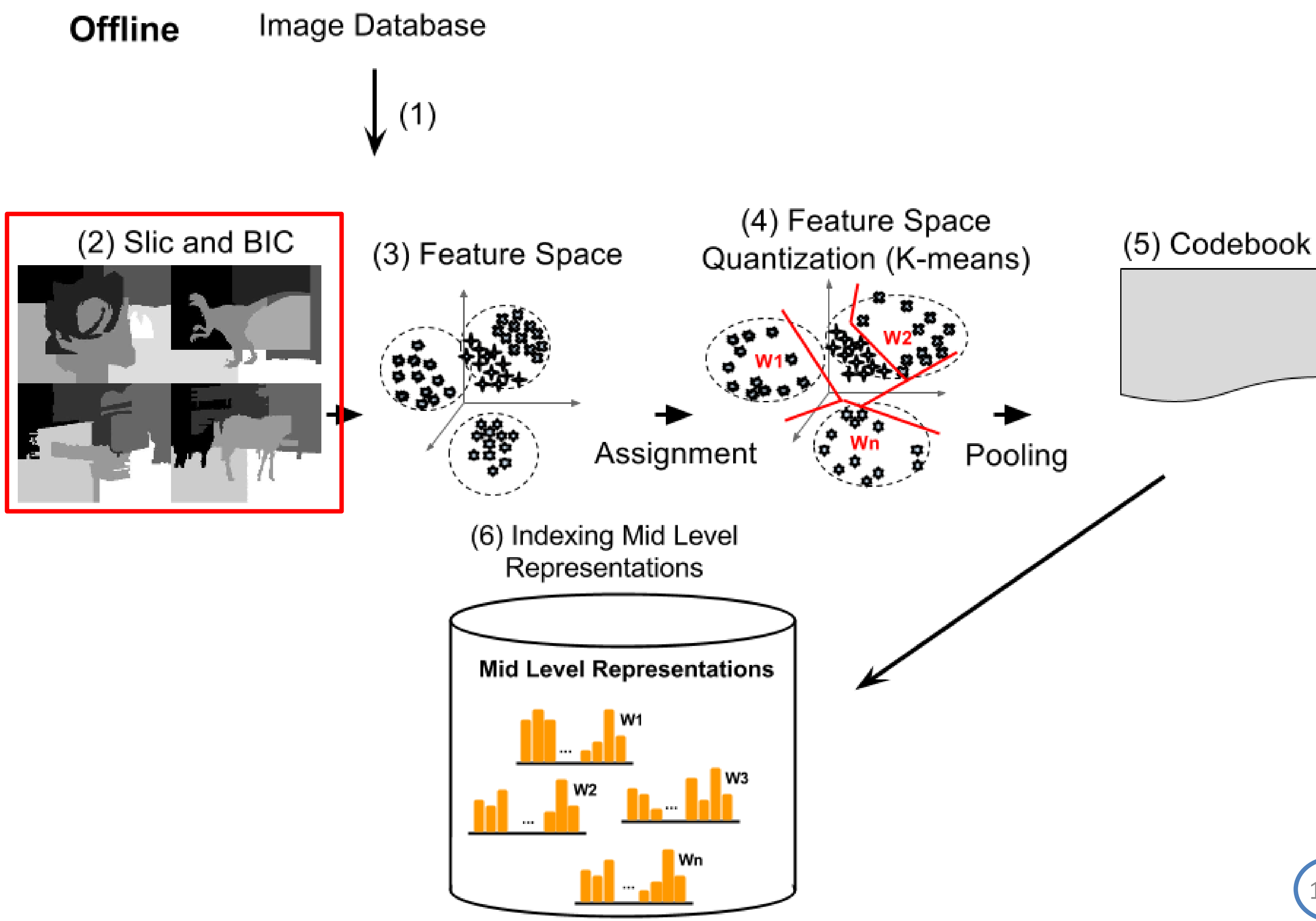


Number of edges: $K_n = \binom{n}{2} = \dfrac{n(n-1)}{2}$

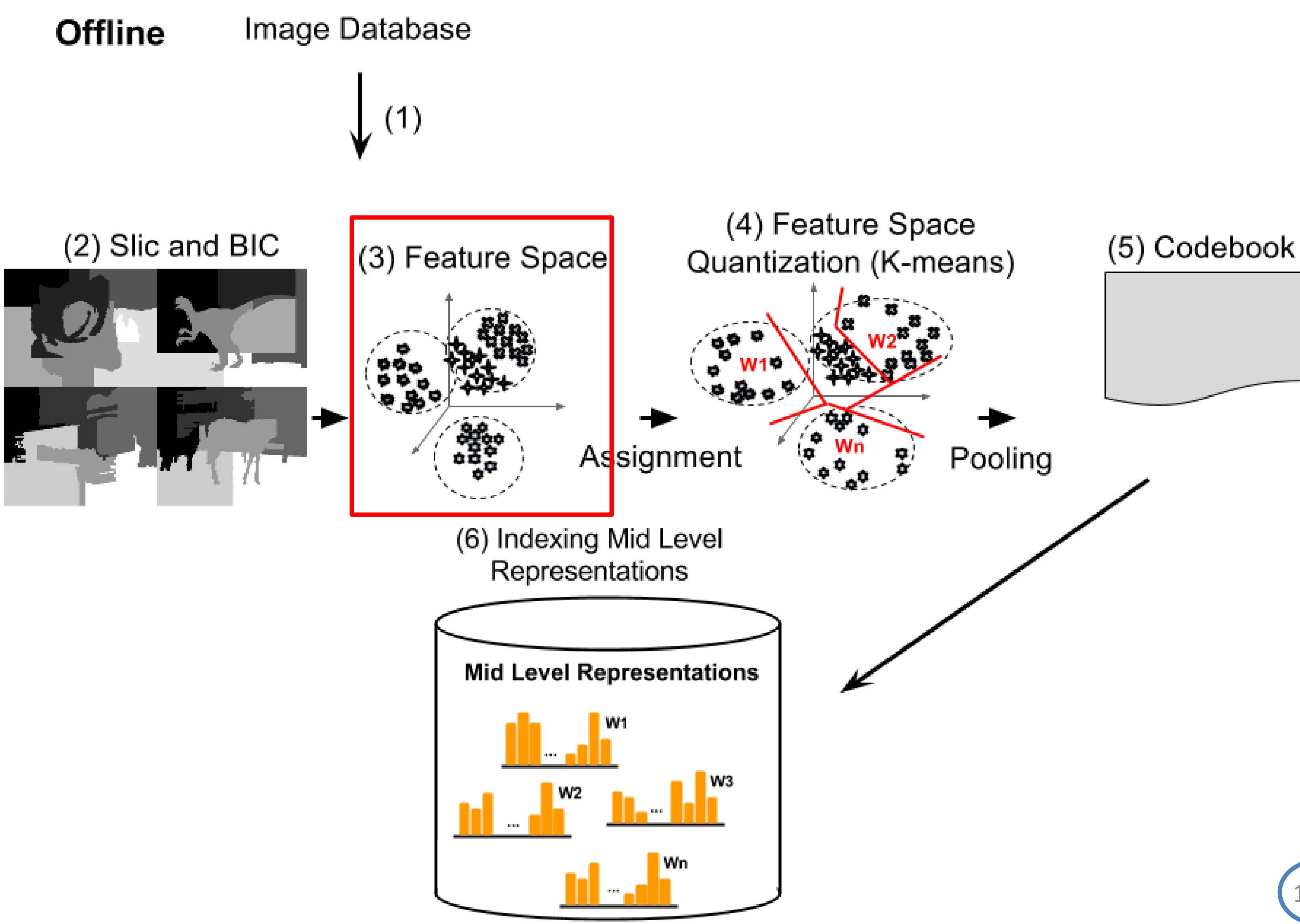
BoBGraph (Bag Of Slic) Representation



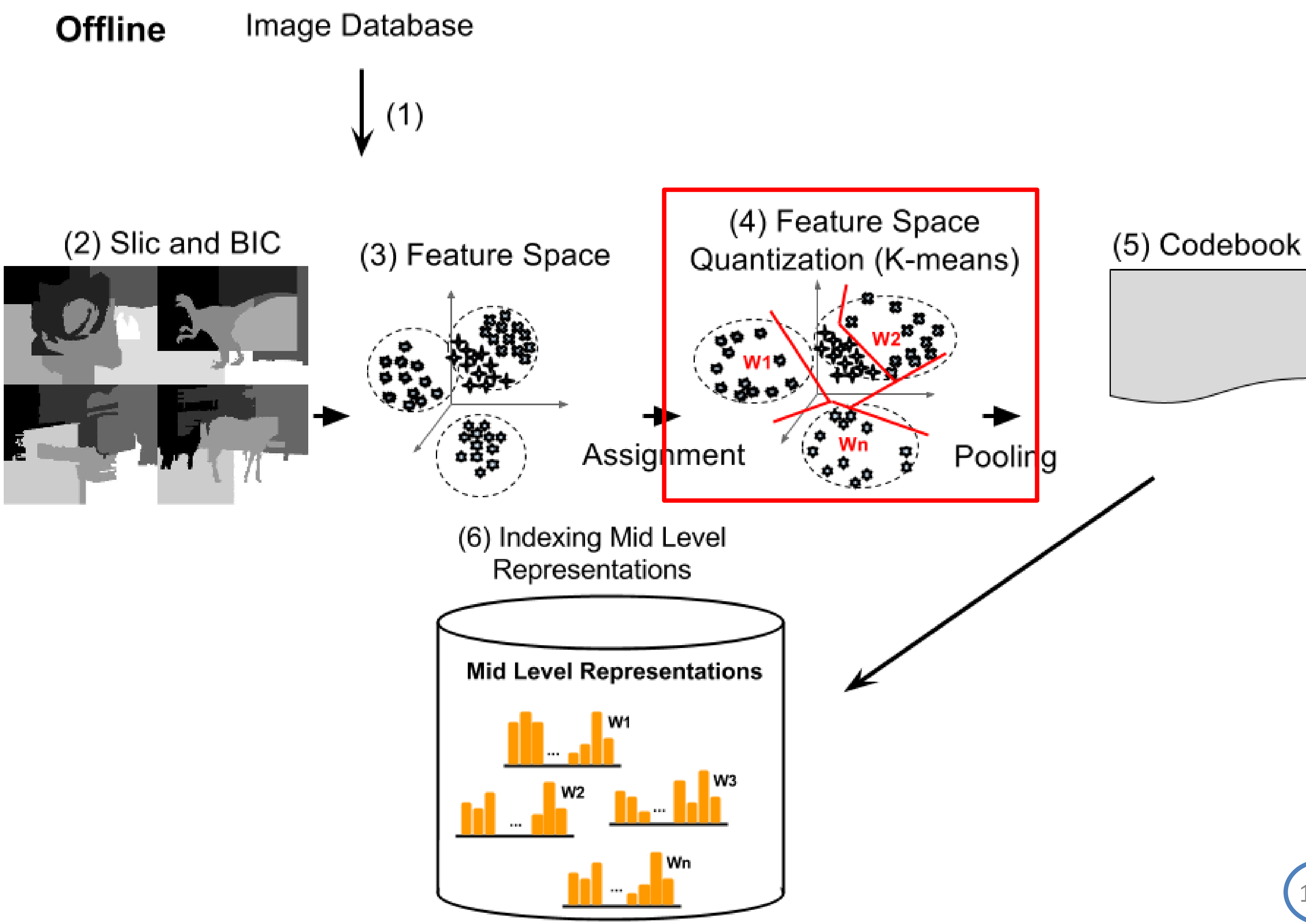
BoBGraph (Bag Of Slic) Representation



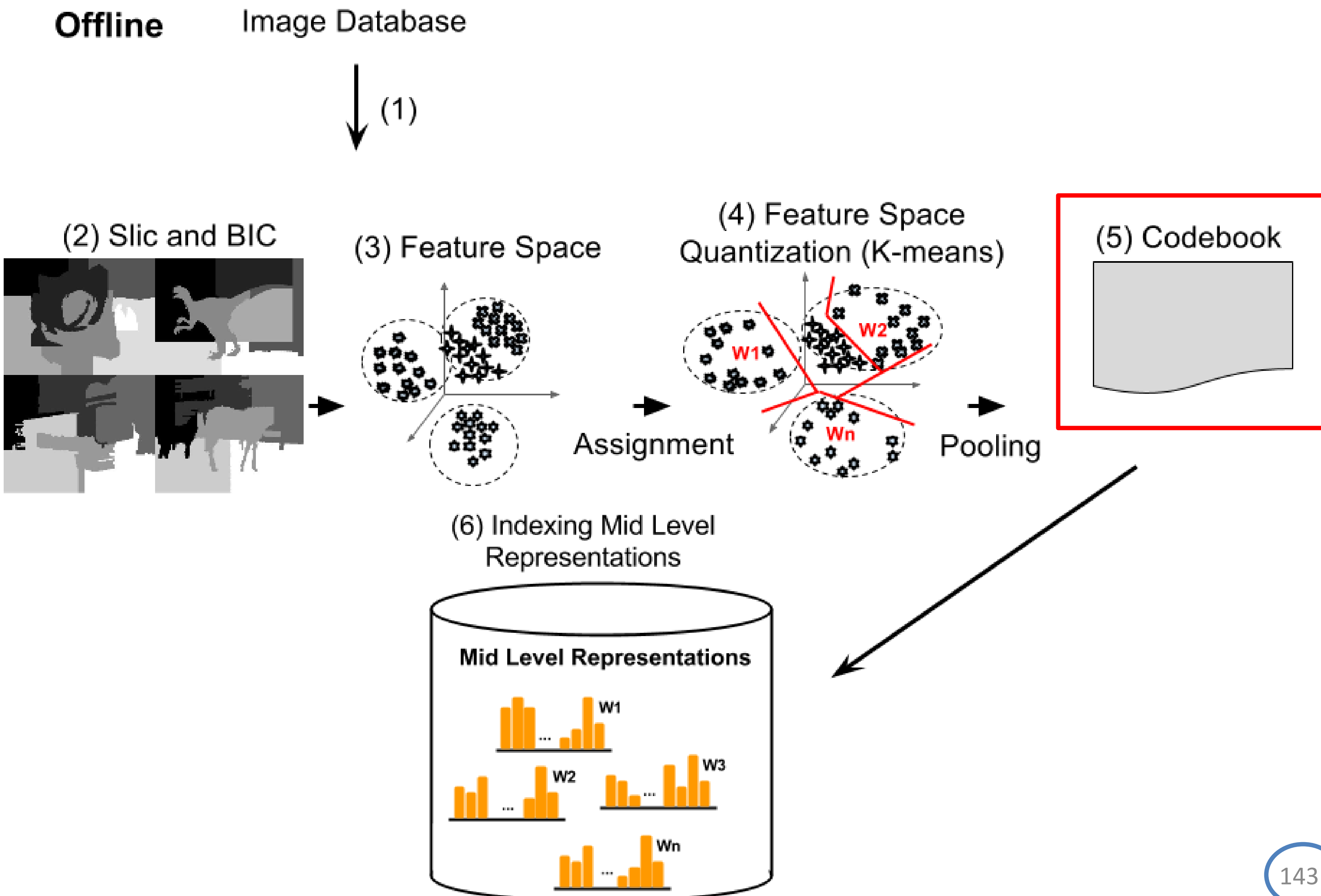
BoBGraph (Bag Of Slic) Representation



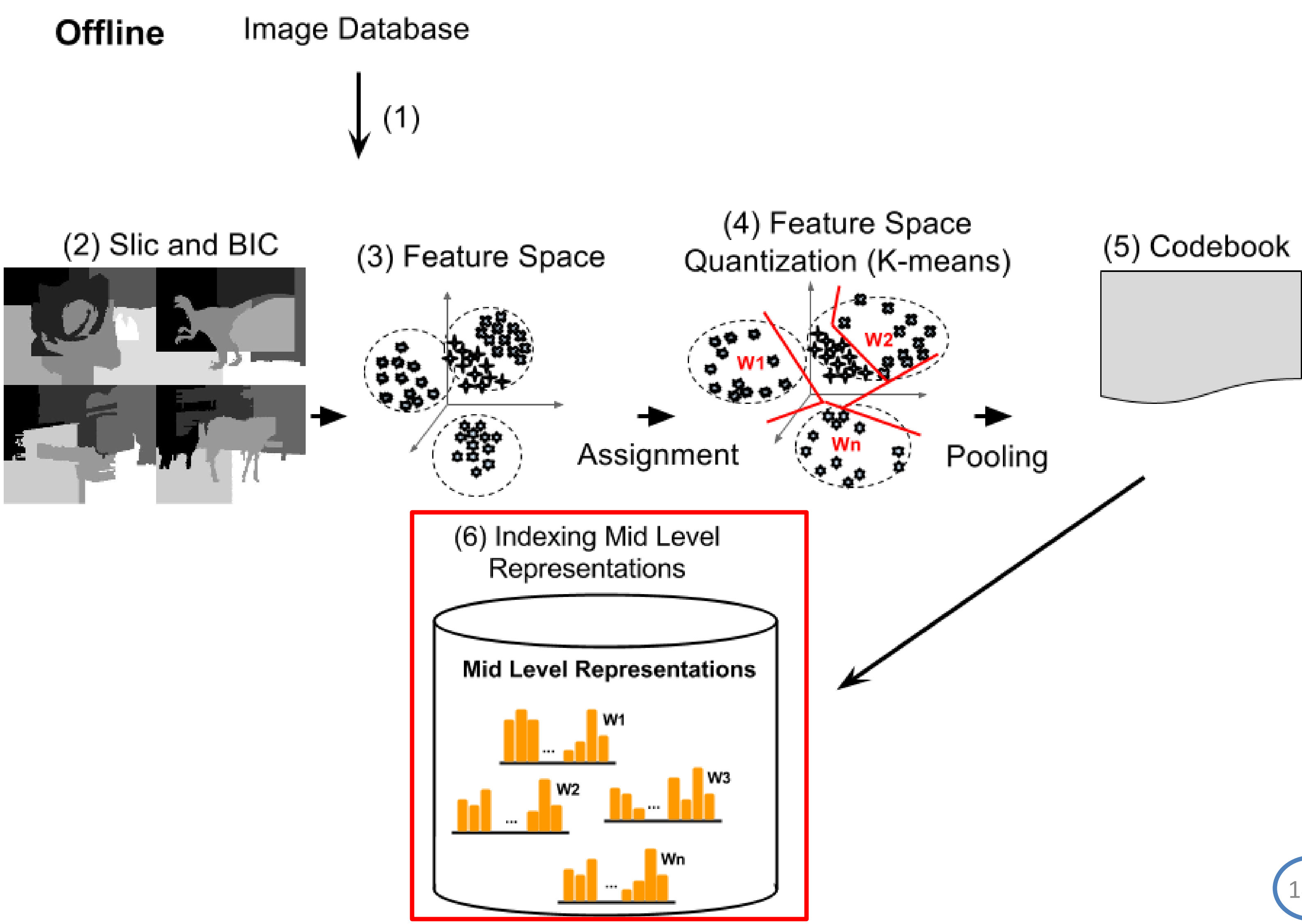
BoBGraph (Bag Of Slic) Representation



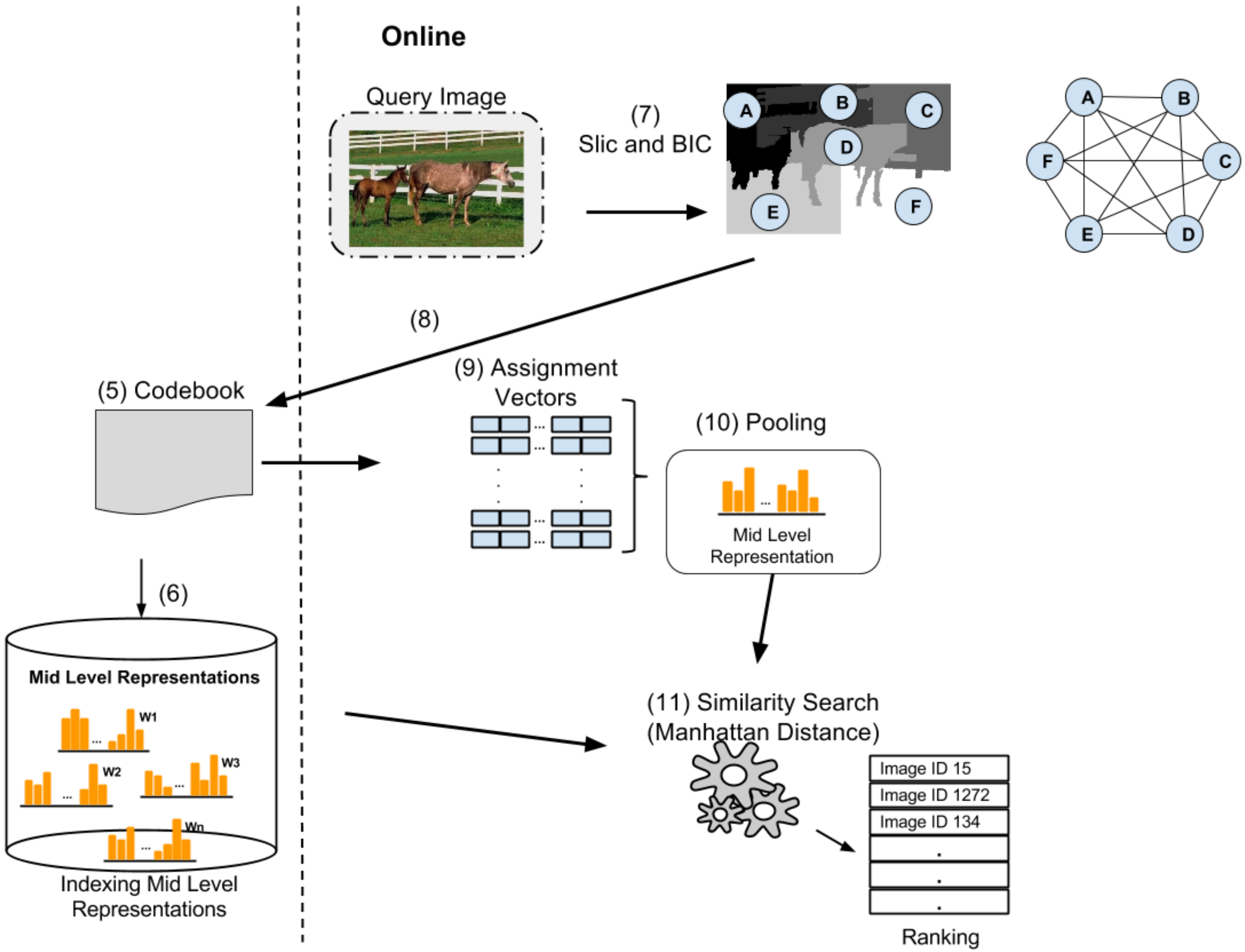
BoBGraph (Bag Of Slic) Representation



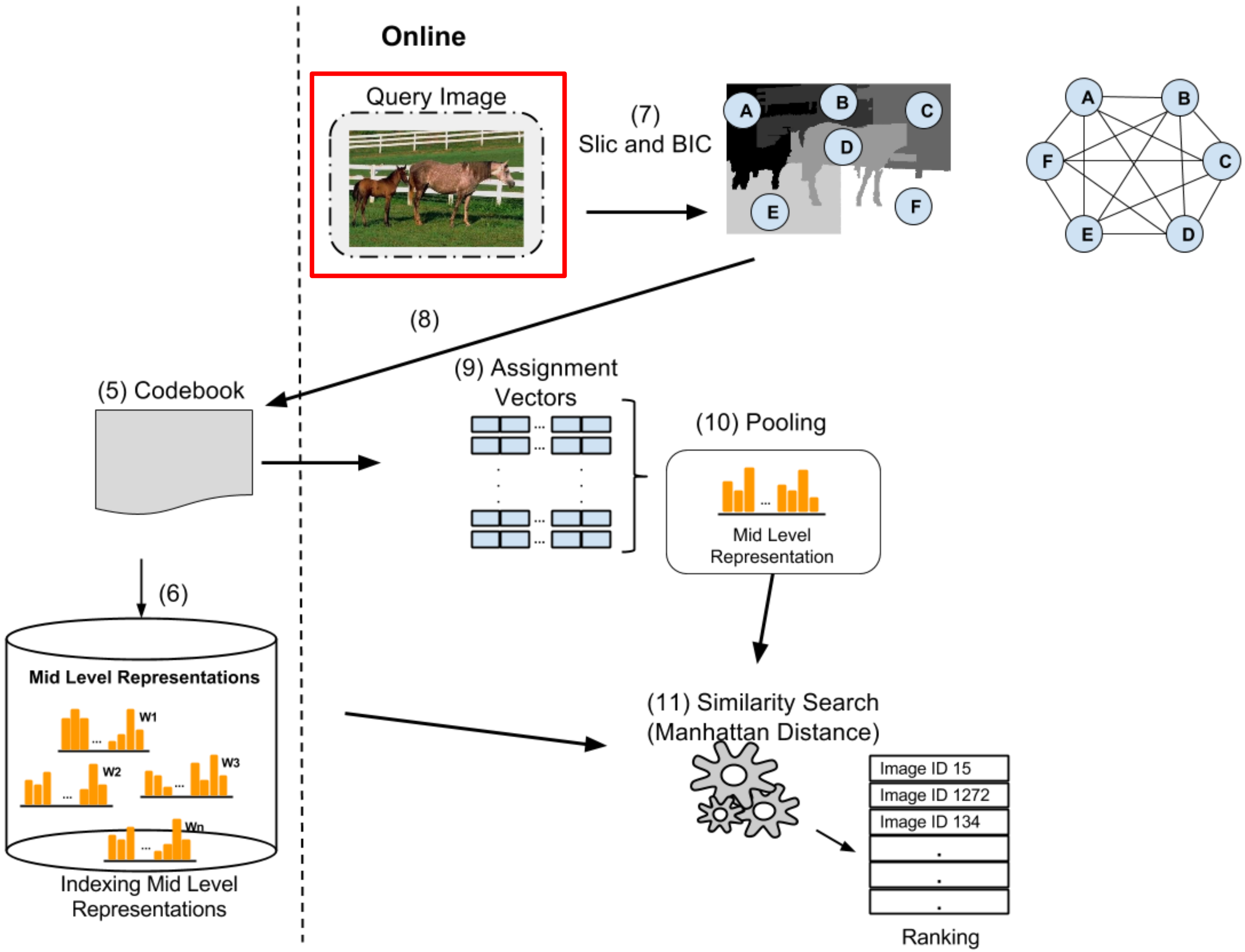
BoBGraph (Bag Of Slic) Representation



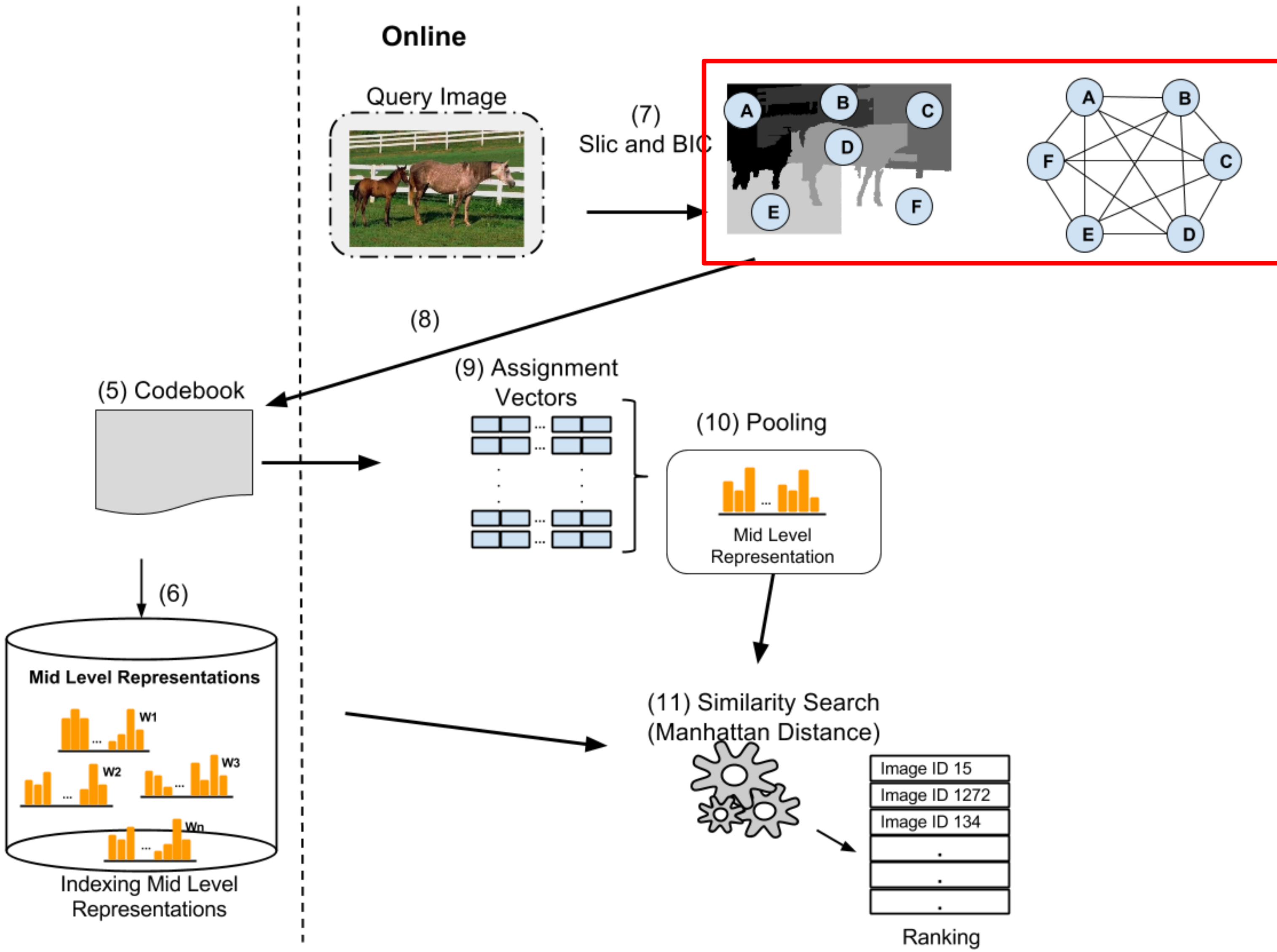
BoBGraph (Bag Of Slic) Representation



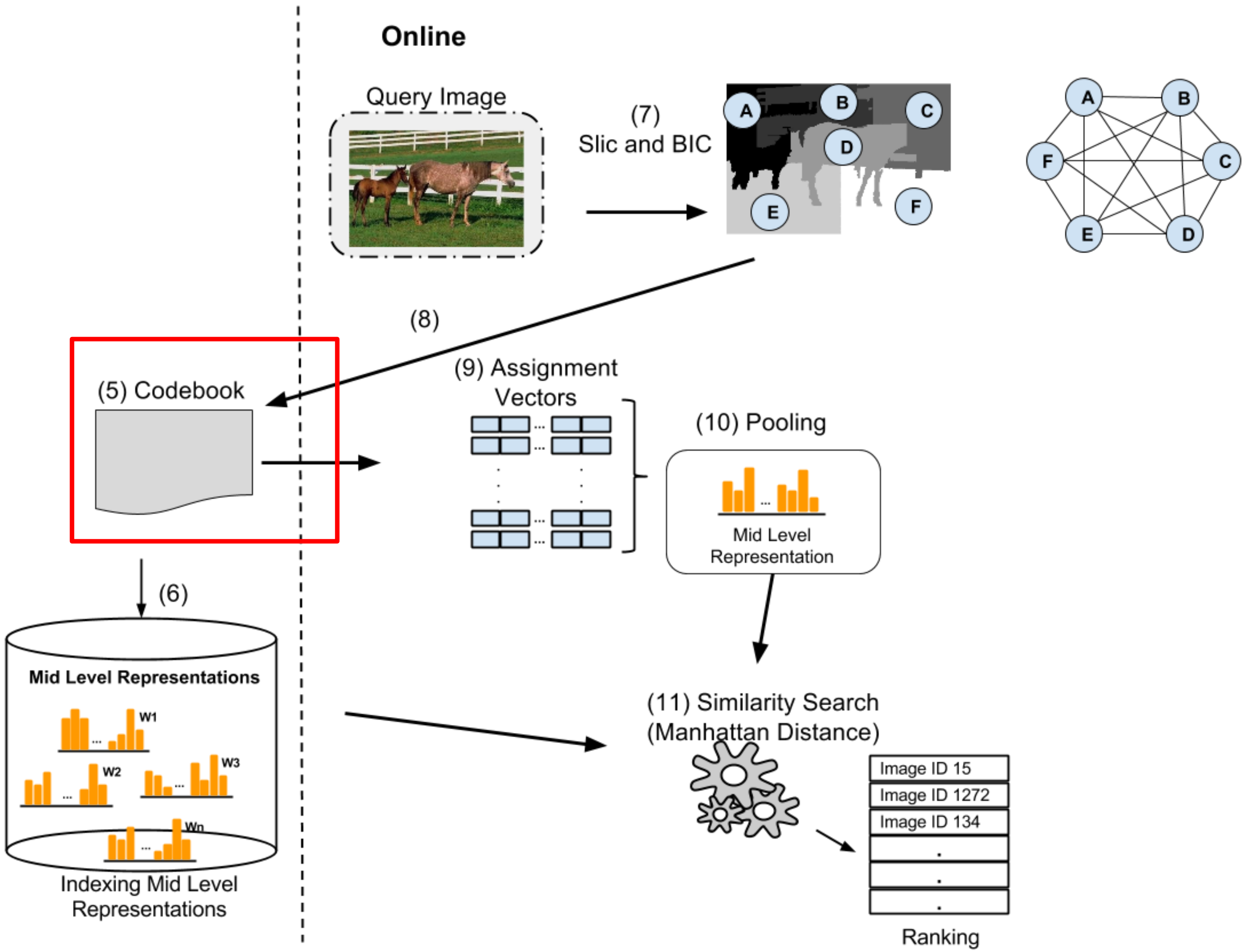
BoBGraph (Bag Of Slic) Representation



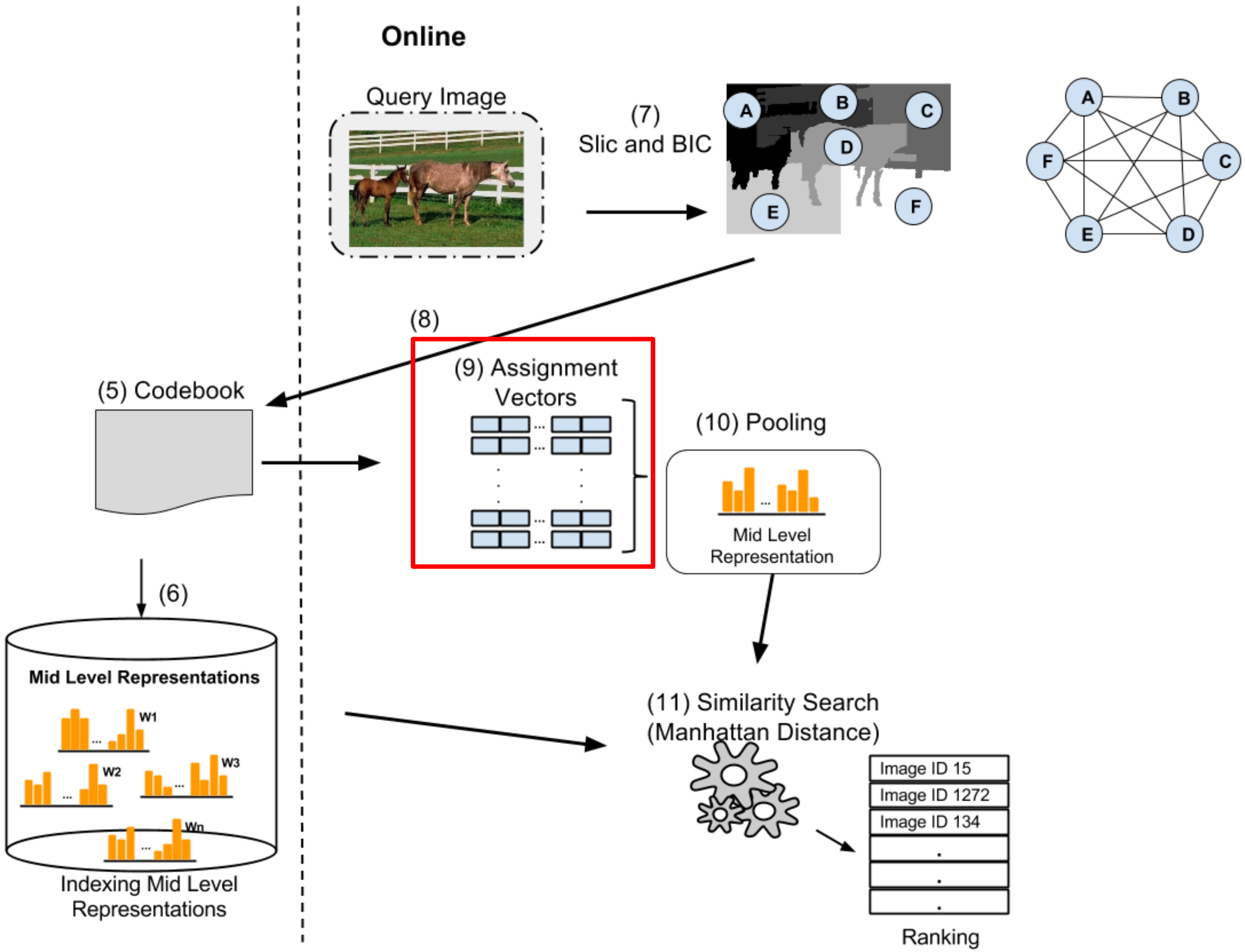
BoBGraph (Bag Of Slic) Representation



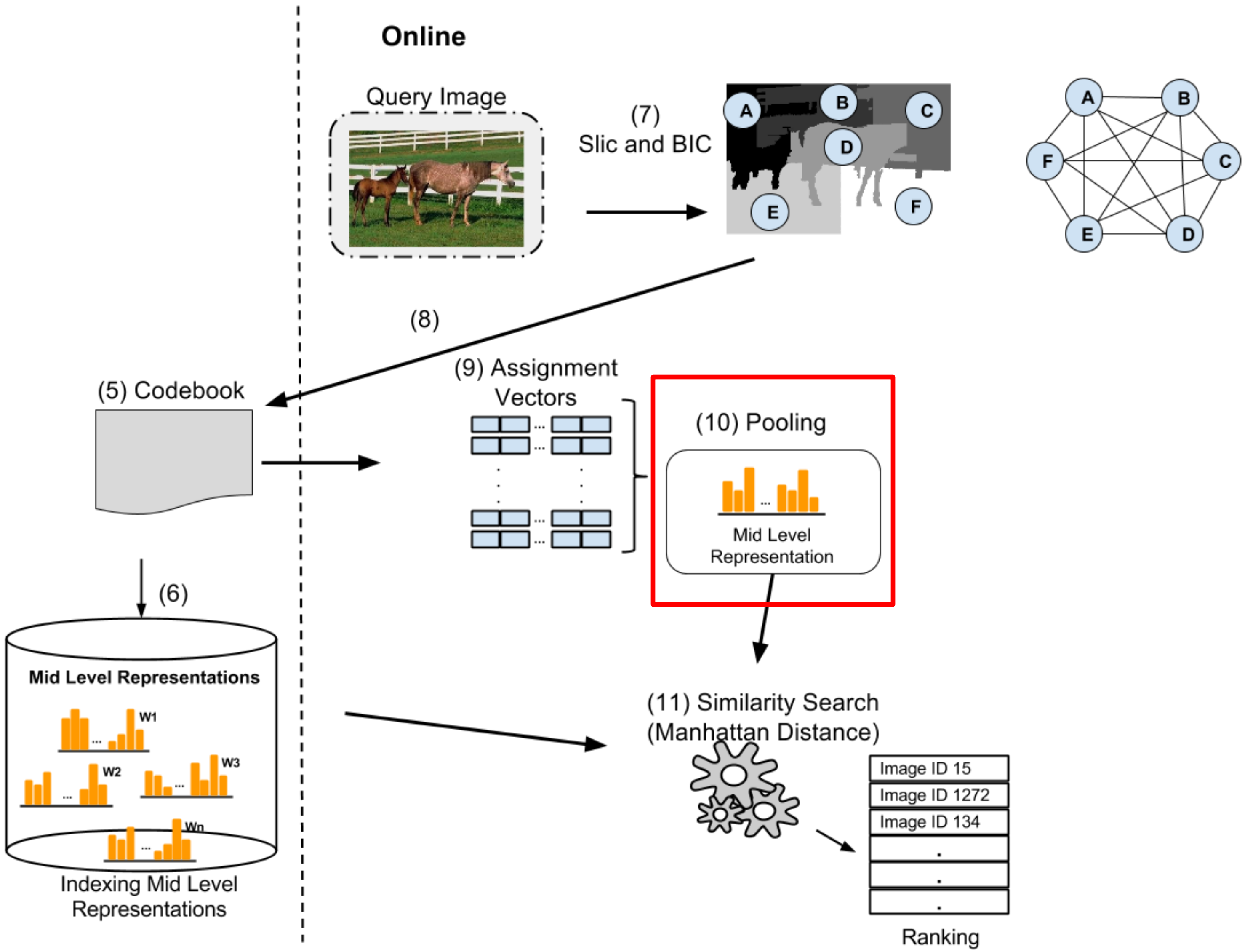
BoBGraph (Bag Of Slic) Representation



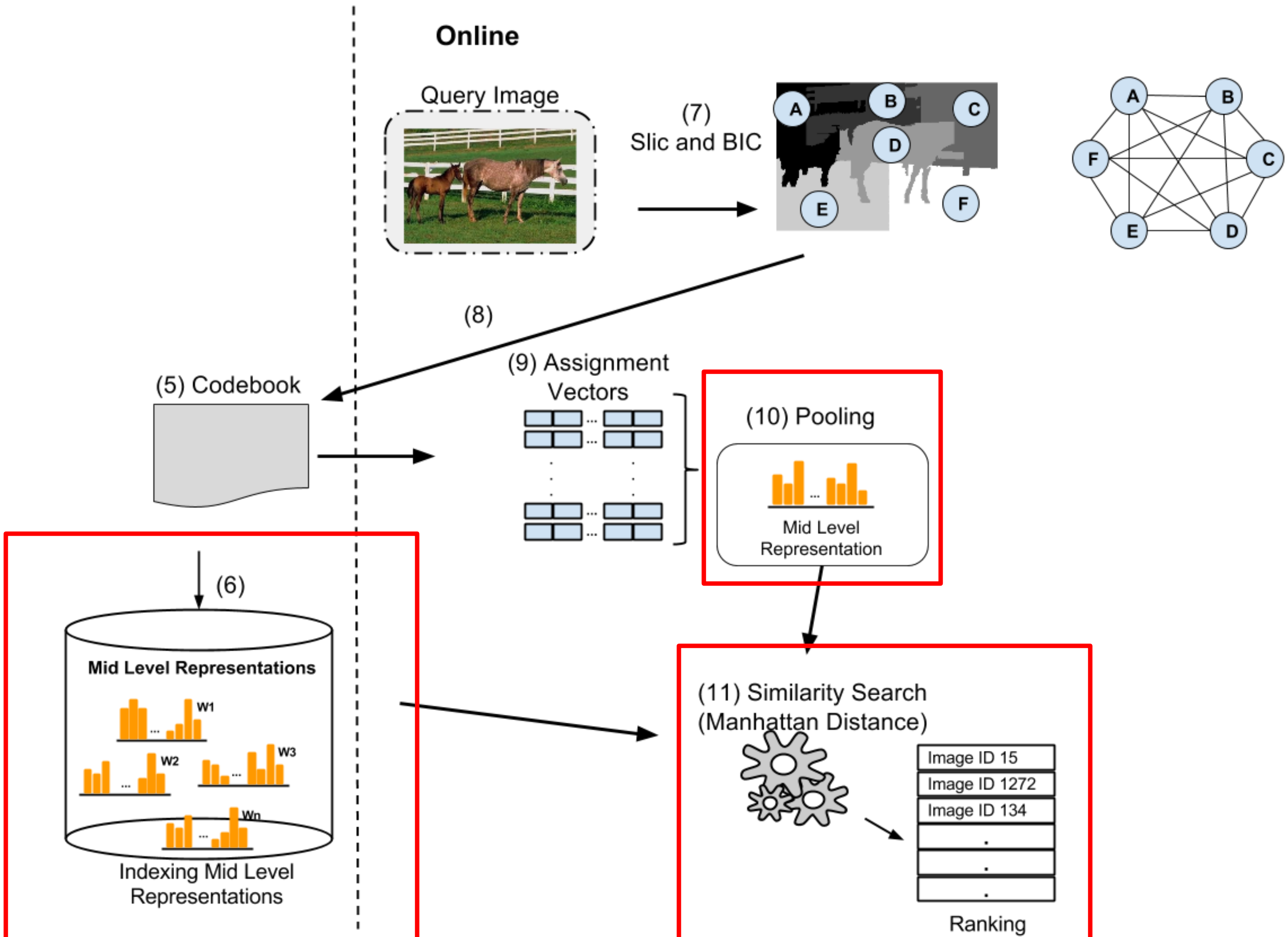
BoBGraph (Bag Of Slic) Representation



BoBGraph (Bag Of Slic) Representation



BoBGraph (Bag Of Slic) Representation



Our Results Vs Baselines: Precision Results

	Approach	P@10	MAP (%)
Baselines	Dense + ORB + BossaNova	58.67	38.93
	GFTTHarris + ORB + BossaNova	46.95	29.33
	ORB + ORB + BossaNova	40.31	27.43
	GFTTHarris + Orb + HardMAX + WSA	18.60	11.96
	GFTTHarris + Orb + SoftMAX + WSA	23.16	13.67
	ORB + ORB + HardMAX + WSA	18.45	11.90
	ORB + ORB + SoftMAX + WSA	23.16	17.58
Our Spatial BoW's Results	**BOBGraph**	**71.57**	**48.06**
	BOBSlic	**71.43**	**48.81**

Our Results Vs Baselines: Feature Vector Size

Method	Feature Vector Size
Traditional BoW's	K
Word Spatial Arrangement (WSA)	4K
BossaNova	K * (B + 1)
BOBGraph	128
BOBSlic	128

- K is the dictionary size
- B is a BossaNova parameter
  - Indicates the local histogram number of bins (Default: B=2)

Our Results Vs Baselines:

Statistical test paired t-test, with 95% of confidence

Method	P@10	CI (95%)
1) BOBGraph	**71.57**	
2) GFTTHarris + ORB + BossaNova	58.67	(1 and 2) X
3) ORB + ORB + BossaNova	46.95	(1 and 3) X
4) GFTTHarris + Orb + HardMAX + WSA	40.31	(1 and 4) X
5) GFTTHarris + Orb + SoftMAX + WSA	18.60	(1 and 5) X
6) ORB + ORB + HardMAX + WSA	23.16	(1 and 6) X
7) ORB + ORB + SoftMAX + WSA	18.45	(1 and 7) X
8) GFTTHarris + ORB + BossaNova	20.54	(1 and 8) X

- Statistical test paired t-test between the **BOBGraph** approach using the ORB descriptor and other descriptors
- The difference is considered significant if it is marked with X

Our Results Vs Baselines:

Statistical test paired t-test, with 95% of confidence

Method	P@10	CI (95%)
1) BOBSlic	**71.43**	
2) GFTTHarris + ORB + BossaNova	58.67	(1 and 2) X
3) ORB + ORB + BossaNova	46.95	(1 and 3) X
4) GFTTHarris + Orb + HardMAX + WSA	40.31	(1 and 4) X
5) GFTTHarris + Orb + SoftMAX + WSA	18.60	(1 and 5) X
6) ORB + ORB + HardMAX + WSA	23.16	(1 and 6) X
7) ORB + ORB + SoftMAX + WSA	18.45	(1 and 7) X
8) GFTTHarris + ORB + BossaNova	20.54	(1 and 8) X

- Statistical test paired t-test between the **BOBSlic** approach using the ORB descriptor and other descriptors
- The difference is considered significant if it is marked with X

Conclusion

- Triple trade-off problem in mobile devices
- Efforts in four main fronts:
  1. Binary low-level descriptor selection
  2. Mid-level representation Analysis
  3. Low-level global representation Analysis
  4. Feasibility analysis of data compression techniques
- The user decides the best triple trade-off configuration
  – Using more or less resources on the mobile devices

Conclusion

- We achieved our goal
  - comparing global descriptors and mid-level representation
    - Some approaches save energy consumption in mobile devices
    - They send more compact feature vector to be processed on the server side
  - Proposing two new bag-of-visual words-based approaches to encode spatial information

Conclusion

- More suitable descriptors to be used in the mobile image search scenario
  - BIC (Border/Interior Pixel Classification) [Stehling et al., 2002]
  - Bag of Words using Dense Sampling, ORB descriptor, Soft assignment and Maximum pooling
- Several experiments that indicates that dense sampling is more accurate than interest point detection

Conclusion

- BOBGraph (Spatial Bag of BIC graph)
- BOBSlic (Spatial Bag of Slic BIC)
- WSA (Visual Word Spatial Arrangement)
- BossaNova (Bag Of Statistical Sampling Analysis)
  - Both BOBGraph and BOBSlic outperform WSA and BossaNova algorithms
- Our descriptors:
  - More compact (just 128 bins)
  - More suitable for mobile devices applications

Future Work

1. To perform experimental analysis on mobile devices
2. To evaluate algorithms of text processing and analyze a multimodal approach to improve the similarity search
3. To exploit more algorithms which use semantic or spatial information
4. To propose a method based to select the best descriptors to use in an average rank aggregation approach
5. To develop deep learning algorithms on mobile devices in a client-server architecture

Publications

- Some results obtained in this work were published in

  1. **XX Iberoamerican Congress on Pattern Recognition (CIARP 2015)**

     [Pessoa et al., 2015a] performed an extensive study on low-cost representations for image feature extraction on mobile devices

  2. **XXVIII Conference on Graphics, Patterns and Images (SIBGRAPI WIP 2015)**

     Pessoa et al. [2015b] performed an experimental comparison of feature extraction and distance metrics for image retrieval

Publication not related to image retrieval on mobile devices

- **SBSEG 2014, Simpósio Brasileiro em Segurança da Informação e de Sistemas Computacionais**

  Pessoa et al. [2014c]

  - Show a Tool for Pornographic Content Detection
  - Use Binary descriptors and Mid-level representation

References

[Ascenso and Pereira, 2013] Ascenso, J. and Pereira, F. (2013). Lossless compression of binary image descriptors for visual sensor networks. In 18th International Conference on Digital Signal Processing (DSP), 2013, pages 1–8.

[Avila et al., 2013] Avila, S., Thome, N., Cord, M., Valle, E., and De A. AraúJo, A. (2013). Pooling in image representation: The visual codeword point of view. Computer Vision and Image Understanding (CVIU), 117(5):453-465

[Cisco, 2015] Cisco (2015). Cisco visual networking index: Global mobile data traffic forecast update. Technical report.

[Girod et al., 2011] Girod, B., Chandrasekhar, V., Chen, D. M., Cheung, N.-M., Grzeszczuk, R., Reznik, Y., Takacs, G., Tsai, S. S., and Vedantham, R. (2011). Mobile visual search. Signal Processing Magazine, IEEE, 28(4):61-76.

[Kumar and Lu, 2010] Kumar, K. and Lu, Y.-H. (2010). Cloud computing for mobile users: Can offloading computation save energy? Computer, 43(4):51-56.

References

[Penatti et al., 2011a] Penatti, O. A., Valle, E., and Torres, R. d. S. (2011a). Encoding spatial arrangement of visual words. In San Martin, C. and Kim, S.-W., editors, Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications, volume 7042 of Lecture Notes in Computer Science, pages 240-247. Springer.

[Pessoa et al., 2015a] Pessoa, R. F., Schwartz, W. R., and dos Santos, J. A. (2015a). A study on lowcost representations for image feature extraction on mobile devices. CIARP 2015, Proceedings, volume 9423, pages 424-431. Springer International Publishing. xx, xxi.

[Pessoa et al., 2015b] Pessoa, R. F., Schwartz, W. R., and Santos, J. A. d. (2015b). An experimental comparison of feature extraction and distance metrics for image retrieval. XXVIII, SIBGRAPI, August 26-29, 2015.

[Zhuang et al., 2014] Zhuang, D., Zhang, D., Li, J., and Tian, Q. (2014). Binary feature from intensity quantization and weakly spatial contextual coding for image search. Information Sciences, 302:94-107.

Thank You

Any Questions

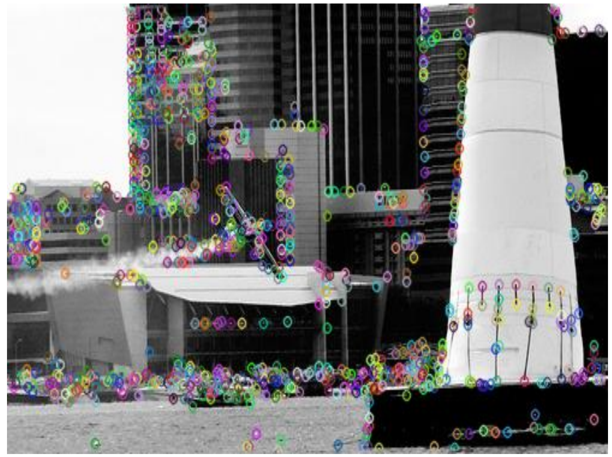
Low-Cost Visual Feature Representations for Image Retrieval

PATREO · UFMG · SSIG Smart Surveillance Interest Group

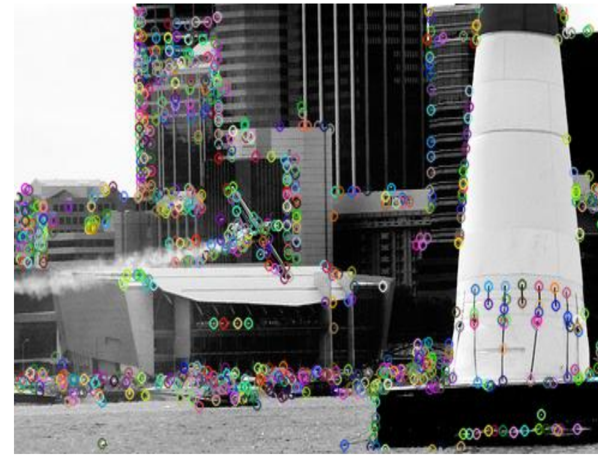
Slides backup

Sampling Strategies

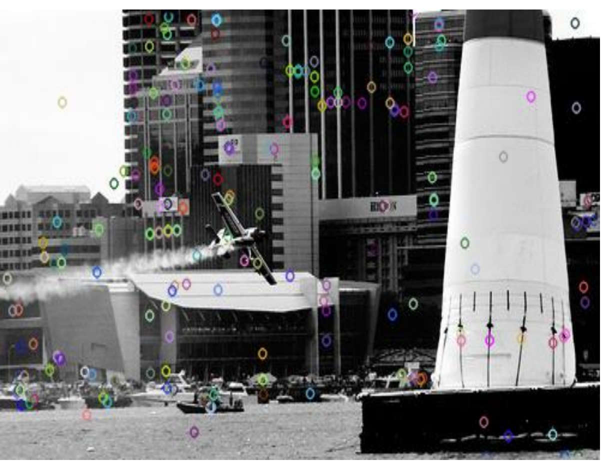
FAST



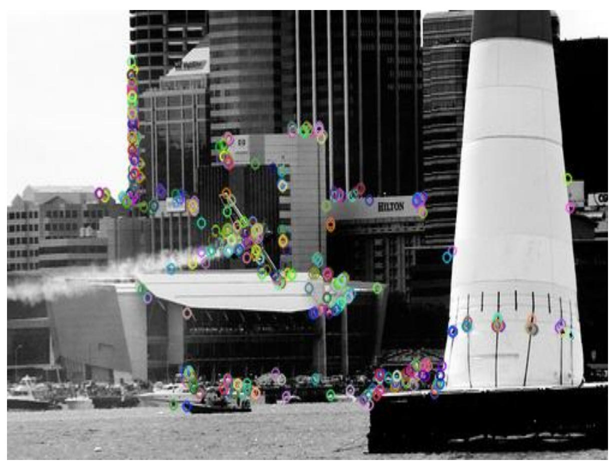
GFTT



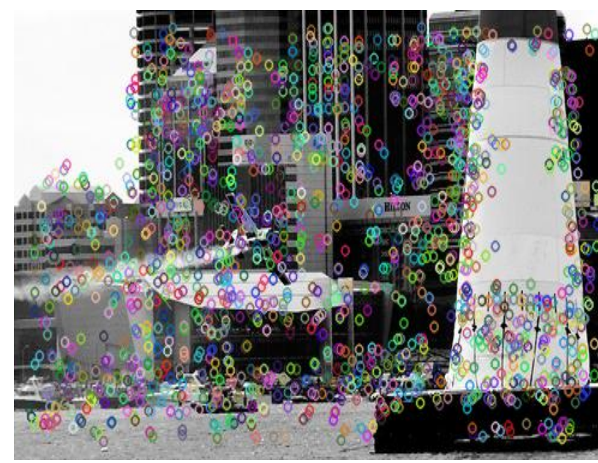
GFTTHarris



MSER

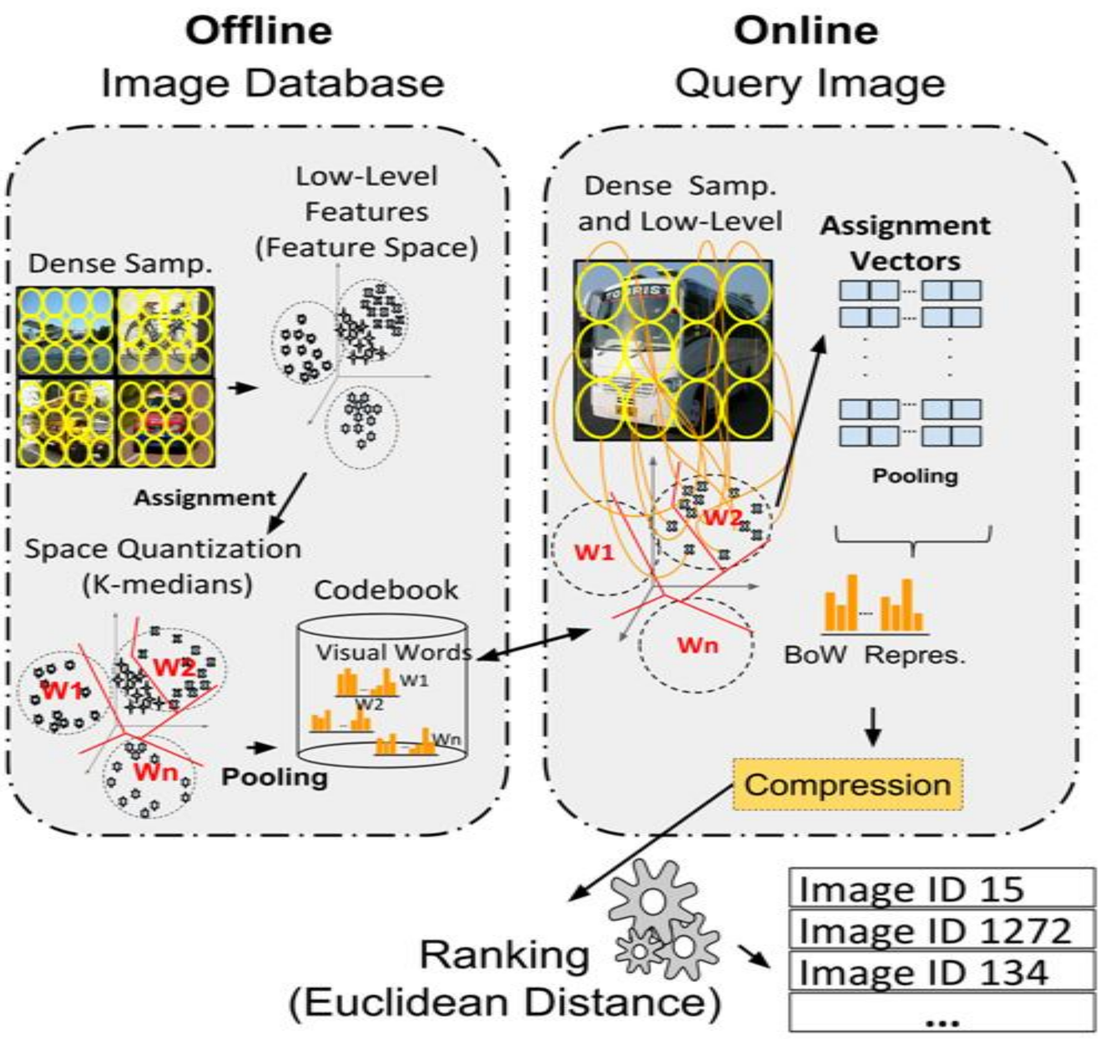


ORBDetector

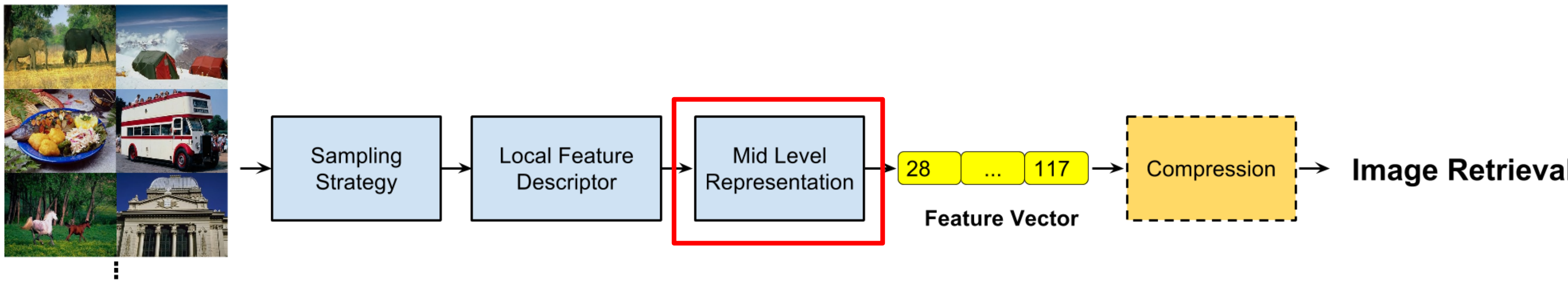


SURF Detector

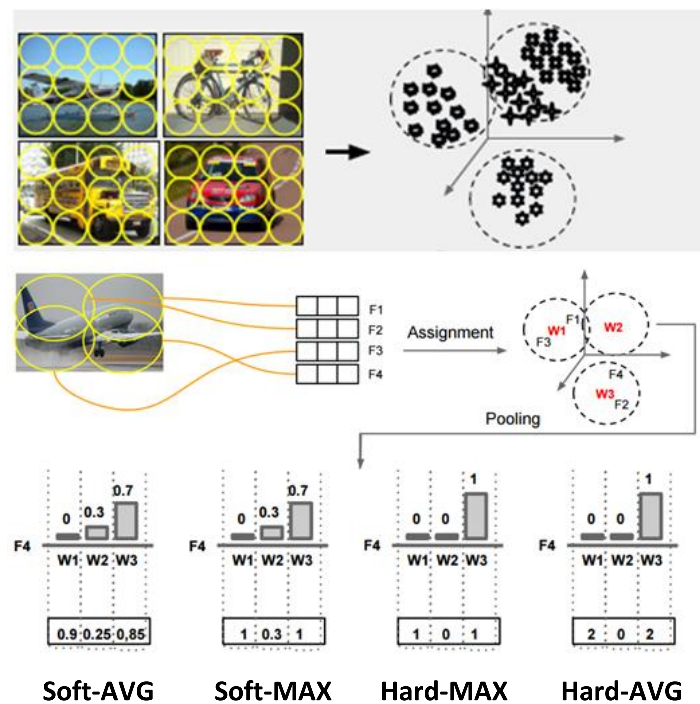
Bag of Visual Words



Background: Mid-Level Representations



- Bag of Visual Words (BoW)
  - Assignment
    - Hard or Soft
  - Pooling
    - Average or Maximum



- 20 BoWs Representations
- We use
  - K-medians
  - Five Binary Descriptors

Benchmark Datasets

Dataset
1: 15Scences
2: OxBuild11
3: ParisLandmarks
4: ZuBuD
5: SMVS692
6: catech101
7: caltech256
8: WANG
9: VOC2007
10: UWdataset

15Scenes

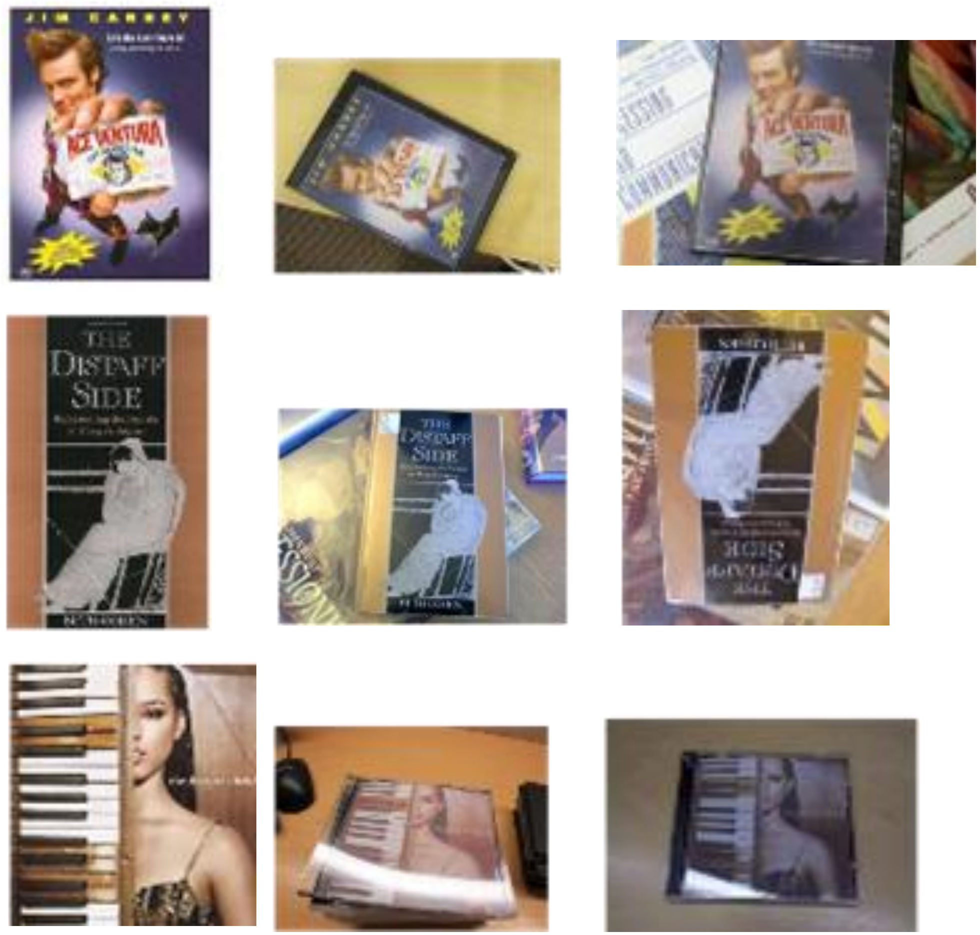
Paris

OxBuild11

Zurich (ZuBuD)

Benchmark Datasets

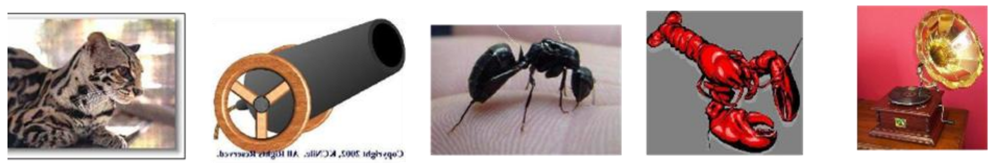
Dataset
1: 15Scences
2: OxBuild11
3: ParisLandmarks
4: ZuBuD
5: SMVS692
6: catech101
7: caltech256
8: WANG
9: VOC2007
10: UWdataset



SMVS692

Benchmark Datasets

Dataset
1: 15Scences
2: OxBuild11
3: ParisLandmarks
4: ZuBuD
5: SMVS692
6: catech101
7: caltech256
8: WANG
9: VOC2007
10: UWdataset



Caltech101

Caltech256

WANG

Benchmark Datasets

Dataset
1: 15Scences
2: OxBuild11
3: ParisLandmarks
4: ZuBuD
5: SMVS692
6: catech101
7: caltech256
8: WANG
9: VOC2007
10: UWdataset

bus, car

car, motorbike, person

cat

VOC2007

australia

Tags: trees beach ocean sky cliffs

geneva

Tags: Buildings Leafless Trees Clear Sky

greenlake

Tags: clear sky building cars street people

UWdataset