Part-based Object Retrieval with Binary Partition Trees

Phd thesis dissertation by Xavier Giró-i-Nieto

Supervised by Professor Ferran Marqués Acosta (UPC) and Professor Shih-Fu Chang (Columbia)
Acknowledgements

Jury

GPI @ UPC

DVMM @ Columbia

Students @ UPC

X. Giró-i-Nieto, “Part-based Object Retrieval with Binary Partition Trees” 31/5/2012 @ UPC
Outline

1. Introduction

2. Hierarchical Image Partitions

3. Interactive segmentation

4. Object Tree

5. Region Features

Part II

6. Codebooks

7. Fusion of Visual Modalities

8. Partition Tree Matching

Part III

9. Part-based Object Model

10. Conclusions
Problem statement

Content-Based Image Retrieval
Problem statement

- Semantic gap between user query and machine data.
Challenges

- Visual **diversity** within the same semantic class

[Anchor image with visual examples of diversity]
Challenges

- Visual diversity within the same instance

This anchor
Challenges

- Semantics located at multiple scales

- News
- Head
- TV studio
- Anchor
- Jacket
- Button
Problem statement

This thesis focuses in the **object retrieval** with one example...

- **Query object**
- **Target dataset**
- **Ranked list**
Problem statement

...and multiple examples.

Query objects

Target dataset

Ranked list
Outline

1. Introduction

2. Hierarchical Image Partitions

Part I

3. Interactive segmentation
4. Object Tree
5. Region Features

Part II

6. Codebooks
7. Fusion of Visual Modalities
8. Partition Tree Matching

Part III

9. Part-based Object Model
10. Conclusions
2. Hierarchical Image Partition

Object analysis

Local scale

Sliding windows  Interest points  Image partition

[Viola-Jones'01]  SIFT [Lowe'04]  SURF [Bay'06]

Object Segmentation?

False  False  True
2. Hierarchical Image Partition

Semantics at multiple scales

Multi-scale analysis

Spatial Pyramid of points

[Lazebnik'06]
[Bosch'07]

Multiple segmentations

[Ravinovich'07]
[Vieux'10]

Hierarchical partition

[Adamek'06]
[Gu'09]
2. Hierarchical Image Partition

**Adopted Solution:** Binary Partition Trees (BPTs) [Garrido, Salembier 2000].

(a)

(b)
2. Hierarchical Image Partition

Semantic Object “head”

Perceptual BPT

Thesis topic

Split

Semantic

Object “head”

Perceptual

BPT
Outline

1. Introduction
2. Hierarchical Image Partitions
3. Interactive segmentation
4. Object Tree
5. Region Features
6. Codebooks
7. Fusion of Visual Modalities
8. Partition Tree Matching
9. Part-based Object Model
10. Conclusions
3. Interactive segmentation

Motivations: Ground truth generation and retrieval interface.

Graphical Annotation Tool (GAT)

Graphical Object Searcher (GOS)

Demo @ http://upseek.upc.edu/gat & http://upseek.upc.edu/gos

X. Giró-i-Nieto, “Part-based Object Retrieval with Binary Partition Trees” 31/5/2012 @ UPC
Outline

1. Introduction
2. Hierarchical Image Partitions
3. Interactive segmentation
4. Object Tree
5. Region Features
6. Codebooks
7. Fusion of Visual Modalities
8. Partition Tree Matching
9. Part-based Object Model
10. Conclusions
4. Object Tree


Level 4: Object

Level 3: Object-part

Level 2: Perceptual

Level 1: Region

BPT nodes

BPT leaves

Thesis approach
4. Object Tree

- **Assumption**: All objects are connected

Interactive segmentation  
BPT sub-roots selection
4. Object Tree

The Object Tree (OT) represents an instance of the object as a hierarchy of regions.
Outline

1. Introduction
2. Hierarchical Image Partitions
3. Interactive segmentation
4. Object Tree
5. Region Features
6. Codebooks
7. Fusion of Visual Modalities
8. Partition Tree Matching
9. Part-based Object Model
10. Conclusions

Part I

Part II

Part III
5. Region Features

MPEG-7 and many more, “Visual Descriptors (VD)” (2001)

Region

Visual Descriptors

Shape
Color
Texture

Similar?

Shape
Color
Texture
5. Region Features

→ Region features in BPT

- Dominant Colors
- Contour Shape
- Edge Histogram

NOT Thesis topic
Outline

1. Introduction
2. Hierarchical Image Partitions
3. Interactive segmentation
4. Object Tree
5. Region Features ←
6. Codebooks
7. Fusion of Visual Modalities
8. Partition Tree Matching
9. Part-based Object Model
10. Conclusions

Part I
1. Introduction
2. Hierarchical Image Partitions
3. Interactive segmentation
4. Object Tree
5. Region Features ←

Part II
6. Codebooks
7. Fusion of Visual Modalities
8. Partition Tree Matching

Part III
9. Part-based Object Model
10. Conclusions

Evaluation
Evaluation

How to evaluate an object retrieval system?

Annotated dataset + Metrics

1. Image Retrieval
2. Object Detection
3. Object Segmentation
4. Search time
Evaluation (in Part II)

Annotated Dataset ETHZ: Category #1 “Apple logos”
Evaluation (in Part II)

Annotated Dataset ETHZ: Category #2 “Bottles”
Evaluation (in Part II)

Annotated Dataset ETHZ: Category #3 “Giraffes”
Evaluation (in Part II)

Annotated Dataset ETHZ: Category #4 “Mugs”
Evaluation (in Part II)

Annotated Dataset ETHZ: Category #5 “Swans”
Evaluation

1. Image Retrieval (global scale)

Only global scale is considered.

User Query (local)  Ranked list of images (global)
Evaluation

1. Image Retrieval (global scale)

Mean Average Precision (MAP) combines Precision and Recall for a set of queries.

Precision

\[ P(k) = \frac{|\{\text{relevant}\} \cap \{\text{retrieved}(k)\}|}{k} \]

Recall

\[ R(k) = \frac{|\{\text{relevant}\} \cap \{\text{retrieved}(k)\}|}{|\text{relevant}|} \]
5. Region features

Image Retrieval (global scale)

![Graph showing performance comparison between region features]

- VDContourShape
- VDDominantColor
- VDEdgeHistogram
- Random

Mean Average Precision

Values:
- VDContourShape: 0.284
- VDDominantColor: 0.295
- VDEdgeHistogram: 0.339
- Random: 0.242
Evaluation

2. Object Detection (local scale-rough)

User Query (local)  Retrieved boxes in relevant images
Evaluation

2. Object Detection (local scale-rough)

Pascal criterion requires a minimum of 50% overlapping in the union of detected object box and ground truth box.

Detection rate = \frac{\text{correct detections}}{\text{relevant}}
5. Region features

Object Detection (local scale-rough)

![Bar chart showing detection rates for different features]

- VDContourShape: 0.192
- VDDominantColor: 0.183
- VDEdgeHistogram: 0.194
Evaluation

3. Object Segmentation (local scale-precise)

User Query (local) → Retrieved regions from correct detections
Evaluation

3. Object Segmentation (local scale-precise)

- **Jaccard index**-$J(A,B)$ compares retrieved region and ground truth mask at the pixel level.

\[
J(A, B) = \frac{|A \cap B|}{|A \cup B|}
\]

- **Mean Jaccard Index** combines the J's obtained by a set of correct detections $D$.

\[
\bar{J} = \frac{1}{|D|} \sum_{i \in D} J_i
\]
5. Region features

Segmentation (local scale-precise)

![Graph showing Jaccard Index averages for different features: VDContourShape (0.607), VDDominantColor (0.547), VDEdgeHistogram (0.586).]
Outline

1. Introduction
2. Hierarchical Image Partitions
3. Interactive segmentation
4. Object Tree
5. Region Features
6. Codebooks
7. Fusion of Visual Modalities
8. Partition Tree Matching
9. Part-based Object Model
10. Conclusions
6. Codebooks

**Visual Word (VW):** Vector quantization of observation (point, region...) onto a codebook [Sivic, Zissermann 2003].

- Regions
- Codebook
- Visual Words

\[ \begin{align*}
(a_1, a_2, a_3, a_4) & \quad (b_1, b_2, b_3, b_4) & \quad (c_1, c_2, c_3, c_4)
\end{align*} \]
6. Codebooks

Code Regions (CR) define the basis of the Parts Space.

“Regions described by their sub-parts”
6. Codebooks

(a) Hard quantization

Visual Words

CR1
(1,0,0,0) ?

CR2
(1,0,0,0)

CR3
(0,1,0,0) ?

CR4

Regions

Codebook

[Sivic 2003]
6. Codebooks

(b) Probabilistic quantization [Alpaydin 1998]

Regions

Codebook

Visual Words

- CR1: (0.25, 0.25, 0.25, 0.25)
  - Visual Words: (0.25, 0.25, 0.25, 0.25)
  - Match
- CR2: (1.0, 0.0, 0.0)
  - Visual Words: (1.0, 0.0, 0.0)
  - Match
- CR3: (0.25, 0.25, 0.25, 0.25)
  - Visual Words: (0.25, 0.25, 0.25, 0.25)
  - Match
- CR4: (0.25, 0.25, 0.25, 0.25)
  - Visual Words: (0.25, 0.25, 0.25, 0.25)
  - Match
6. Codebooks

(c) Possibilistic quantization [Bezdek 1995]

Regions

Codebook

Visual Words

CR1

CR2

CR3

CR4

Similarity to CRs

(0,0,0,0)

(1,0,0,0)

(0,0,0,0)
How to define the Parts Space with a given codebook?

Possibilistic quantization alone characterizes regions, but not their sub-parts.

Region-based Possibilistic
6. Codebooks

How to define the Parts Space with a given codebook?

Bottom-up Max Pooling through BPT

[Boureau'10]
6. Codebooks

→ Hierarchy of Bags of Regions (HBoR)

- BPT is exploited to describe sub-trees.
6. Codebooks

How to generate a codebook? (1/2)

Train set

BPT-based decomposition

K-Means Clustering

Object parts

Codebook

CR₁  CR₂  CR₃  CR₄
6. Codebooks

How to generate a codebook? (2/2)

a) Foreground and background regions - “generic”

b) Only object regions - “frgd”
6. Codebooks

1. Image Retrieval (global scale) - Texture

HBoR features outstanding region-based for shape & texture (not dominant colors).

Little differences between generic vs frgd codebooks (5 classes in ETHZ).

Codebook size
Outline

1. Introduction
2. Hierarchical Image Partitions
3. Interactive segmentation
4. Object Tree
5. Region Features
6. Codebooks
7. Fusion of Visual Modalities
8. Partition Tree Matching
9. Part-based Object Model
10. Conclusions
5. Region Features + 7. Fusion

- **Problem:** Similarity scores between descriptors are not comparable.
- **Solution:**

\[
\begin{align*}
  d_{\text{color}} & \quad d_{\text{shape}} & \quad d_{\text{texture}} \\
  \overline{x}_{\text{color}} & \quad \overline{x}_{\text{shape}} & \quad \overline{x}_{\text{texture}} \\
  \overline{\chi} & \\
\end{align*}
\]

Normalization (z-scores, SVM, w-scores)

Fusion (average, product, max/min)
Outline

1. Introduction
2. Hierarchical Image Partitions
3. Interactive segmentation
4. Object Tree
5. Region Features

Part II
6. Codebooks ←
7. Fusion of Visual Modalities ←

Part III
8. Partition Tree Matching
9. Part-based Object Model
10. Conclusions
6. Codebooks + 7. Fusion

When working with codebooks, when to apply fusion?

(a) Fusion after quantization - “nofused”
6. Codebooks + 7. Fusion

When working with codebooks, when to apply fusion?

(b) Fusion during quantization - “fused”

Query region

\[
\begin{bmatrix}
\text{color} \\
\text{shape} \\
\text{texture}
\end{bmatrix}
\]

Target region

\[
\begin{bmatrix}
\text{color} \\
\text{shape} \\
\text{texture}
\end{bmatrix}
\]

Quantization

\[
[ VW ]
\]

Cosine distance

\[
x
\]

Fused codebook
6. Codebooks + 7. Fusion

When working with codebooks, when to apply fusion?

1. Image Retrieval (global scale)

- **z-scores**
- **w-scores**

W-scores outperform z-scores when working with HBoR.
Outline

1. Introduction
2. Hierarchical Image Partitions
3. Interactive segmentation
4. Object Tree
5. Region Features
6. Codebooks
7. Fusion of Visual Modalities
8. Partition Tree Matching
9. Part-based Object Model
10. Conclusions

Part I

Part II

Part III

Evaluation

X. Giró-i-Nieto, “Part-based Object Retrieval with Binary Partition Trees” 31/5/2012 @ UPC
Evaluation (in Part III)

- CCMA News dataset.

- Topology diversity on 11 anchors objects.
Outline

Part I
1. Introduction
2. Hierarchical Image Partitions
3. Interactive segmentation
4. Object Tree
5. Region Features

Part II
6. Codebooks
7. Fusion of Visual Modalities

Part III
8. Partition Tree Matching
9. Part-based Object Model
10. Conclusions
8. Partition Tree Matching

**Reminder:** The Object Retrieval Problem (single instance)

- **Query object**
- **Target dataset**
- **Ranked list**

X.Giró-i-Nieto, “Part-based Object Retrieval with Binary Partition Trees” 31/5/2012 @ UPC
8. Partition Tree Matching

---

Object Tree (OT)

- Q0
  - Q1
  - Q2
    - Q3
    - Q4

Target BPT

Challenges:

(a) Distance for a given OT-BPT correspondence
(b) Matching algorithm

---
8. Partition Tree Matching: Distance

Proposal (ad-hoc): Weighted sum of part matches

\[ d = \sum_{i=0}^{N} w_i d_i \text{, where } w_i = \alpha_i \prod_{j \in \text{Ancestors}(Q_i)} \alpha_j = \alpha_i \cdot w_{\text{Parent}(Q_i)} \]

\[ \alpha_i = \frac{1}{\max(1, |C^i|)} \frac{\text{area}(i)}{\text{area}(P^i)} \]

Example:

\[ \alpha_0 = \frac{1}{2} = w_0 \]
\[ \alpha_1 = \frac{\text{area}(Q_1)}{\text{area}(Q_0)} \]
\[ \alpha_2 = \frac{\text{area}(Q_2)}{\text{area}(Q_0)} \]

Children of part \( i \)

Parents of part \( i \)
8. Partition Tree Matching: Algorithm

- Two proposals:
  - (a) Top-down Matching
  - (b) Bottom-Up Matching

Object Tree (OT)

- Q0
  - Q1
    - Q3
  - Q2
    - Q4

Target BPT
8. Partition Tree Matching: Algorithm

(a) Top-down Matching

Assumption: All OT nodes are contained in a single sub-BPT.

(b) Bottom-Up Matching

Assumption: OT nodes are spread through different sub-BPTs.
8. Partition Tree Matching: Algorithm

(b) Top-down QT Matching

Assumption: The best match for the OT is contained in a single sub-tree of the target BPT.

1. Set a match for Q0

2. Find matches for Q1 and Q2
8. Partition Tree Matching

What is the impact of decomposing the query into parts?

Image Retrieval (global scale)

MPEG-7 Region

Top-down match

Retrieval gain until the “correct” amount of parts (3).

X.Giró-i-Nieto, “Part-based Object Retrieval with Binary Partition Trees” 31/5/2012 @ UPC
8. Partition Tree Matching

What is the impact of decomposing the query into parts?

Search time

Exponential growth of search time with the amount of parts.
8. Partition Tree Matching: Algorithm

(a) Top-down Matching

Assumption: All OT nodes are contained in a single sub-BPT.

(b) Bottom-Up Matching

Assumption: OT nodes are spread through different sub-BPTs.
8. Partition Tree Matching: Algorithm

(c) Bottom-UP QT Matching

1. Set a match for Q1 & Q2

2. Assess the union of candidates for Q1 and Q2 to match Q0

Weaker assumption that in the top-down case:

→ OT leaves are among the nodes of the target BPT
8. Partition Tree Matching: Algorithm

(c) Bottom-UP QT Matching

How to rapidly describe the merged region generated during search time?

→ Approximate HBoR satisfies this requirement.
8. Partition Tree Matching

What is the impact of introducing the bottom-up match?

A clear gain, the object splits in the target BPTs are solved.
Outline

Part I
1. Introduction
2. Hierarchical Image Partitions
3. Interactive segmentation
4. Object Tree
5. Region Features

Part II
6. Codebooks
7. Fusion of Visual Modalities
8. Partition Tree Matching

Part III
9. **Part-based Object Model**
10. Conclusions
9. Part-based Object Model

**Reminder**: The Object Retrieval Problem (multiple instances)

Query objects

Target dataset

Ranked list
9. Part-based Object Model

Model-based retrieval

Query objects

Model

Training

Target image

Test

Retrieved object
9. Part-based Object Model

Parts definition

Inclusion classifier
Detector classifier

Fusion classifier

Inclusion classifier
Detector classifier

Fusion classifier
Contextual filtering

Training

Test
9. Part-based Object Model

Training

Parts definition

Inclusion classifier

Detector classifier

Fusion classifier

Contextual filtering

Test

Inclusion classifier

Detector classifier

Fusion classifier
9. Part-based Object Model

Parts definition

Unsupervised clustering with Quality Threshold [Heyer 1999].

Two parameters:
- Minimum amount of elements in cluster
- Maximum radius of a cluster
9. Part-based Object Model

- Parts definition
- Inclusion classifier
- Detector classifier
- Fusion classifier
- Contextual filtering

Training

Test

# Parts
9. Part-based Object Model

Positive: Parts and their ancestors
Negative: Parts' children and positive node's siblings.

Inclusion classifier (training)

HBoR
9. Part-based Object Model

Training

Parts definition

Inclusion classifier

Detector classifier

Fusion classifier

Test

Inclusion classifier

Detector classifier

Fusion classifier

Contextual filtering
9. Part-based Object Model

Efficient top-down exploration of the target BPT.

HBoR

Not explored (efficiency)
9. Part-based Object Model

Parts definition

Inclusion classifier

Detector classifier

Fusion classifier

Contextual filtering

Inclusion classifier

Detector classifier

Fusion classifier

Training

Test

# Parts

# Parts
9. Part-based Object Model

Positive: Parts.
Negative: Random sampling (balanced set).

Detector classifier (training)
9. Part-based Object Model

Inclusion classifier
Detector classifier
Fusion classifier
Contextual filtering

Training

Parts definition

Test

# Parts

# Parts
9. Part-based Object Model

Fusion classifier (training)

Positive training samples represent valid combinations of parts.
Negative training samples represent invalid combinations.

Parts definition

Type 1

Type 2

Type 3

Type 4

Parts of object i

Parts of object j

(1, 1, 0, 0)

(0, 1, 1, 1)
9. Part-based Object Model
9. Part-based Object Model

How important is a correct estimation of the types of parts?

The performance is dependent from the Quality Threshold parameters.
9. Part-based Object Model

What is the impact of codebook size in terms of time?

- **Training time**: Search time does not grow with codebook size (Efficient exploration).

- **Search time**: Does not grow with codebook size (Efficient exploration).
Outline

1. Introduction
2. Hierarchical Image Partitions
3. Interactive segmentation
4. Object Tree
5. Region Features
6. Codebooks
7. Fusion of Visual Modalities
8. Partition Tree Matching
9. Part-based Object Model
10. Conclusions
Conclusions

BPT presents several opportunities for local analysis:

→ Interactive segmentation
→ Semantic analysis

Three main contributions in this thesis:

Contribution #1
Hierarchy of Bags of Regions:
- Possibilistic quantization
- Max pooling through BPT

Contribution #2
Object Tree - BPT matching
- Top-down
- Bottom-up

Contribution #3
Part-based model for objects
- Automatic determination of parts
- Efficient BPT analysis.
Open work

- Test in public benchmark.
  → Instance Search task @ TRECVID 2012

- Richer image representations
  → Amount of BPT leaves adapted to content
  → Discard uninformative merges & allow multiple options

- Enrich region descriptors
  → Introduce interest points
  → Significance estimation

- Smarter codebooks
  → Unsupervised clustering
  → TF-IDF weights & Stop words

- Better model for combination of parts representing object

- Local segmentation of objects from global annotations
Outline

1. Introduction
2. Hierarchical Image Partitions
3. Interactive segmentation
4. Object Tree
5. Region Features
6. Codebooks
7. Fusion of Visual Modalities
8. Partition Tree Matching
9. Part-based Object Model
10. Conclusions

Part I

Part II

Part III
Thank you, moltes gràcies !

The Origin of the Theses