# Composite Object Detection in Video Sequences:
## Application to Controlled Environments

Xavier Giró and Ferran Marqués

Image Processing Group
Signal & Communications Department
Technical University of Catalonia (UPC)

# Outline

1. **Introduction ←**
2. Generic algorithm
   - Perceptual Analysis
   - Semantic Analysis
3. Controlled environment
   - Perceptual
   - Semantic
4. Results
5. Conclusions

# Introduction

- Our goal: Use context to improve object detection performance

There are four laptops in the room located in (X,Y) and with orientation Dº

Semantic description

ANALYSIS

Perceptual data (visual)

# Introduction

- Our approach: Adapt a generic indexing technique to a specific real-time detection problem in a controlled environment.

| Indexing | Detection |
|---|---|
| →CBIR from database | →Real-time applications |
| →Generic environment | →Controlled environment |
| →Universal analysis | →Class-specific analysis |
| →Bottom-up | →Top-down |
| →Loose time restrictions | →Tight time restrictions |

ANALYSIS

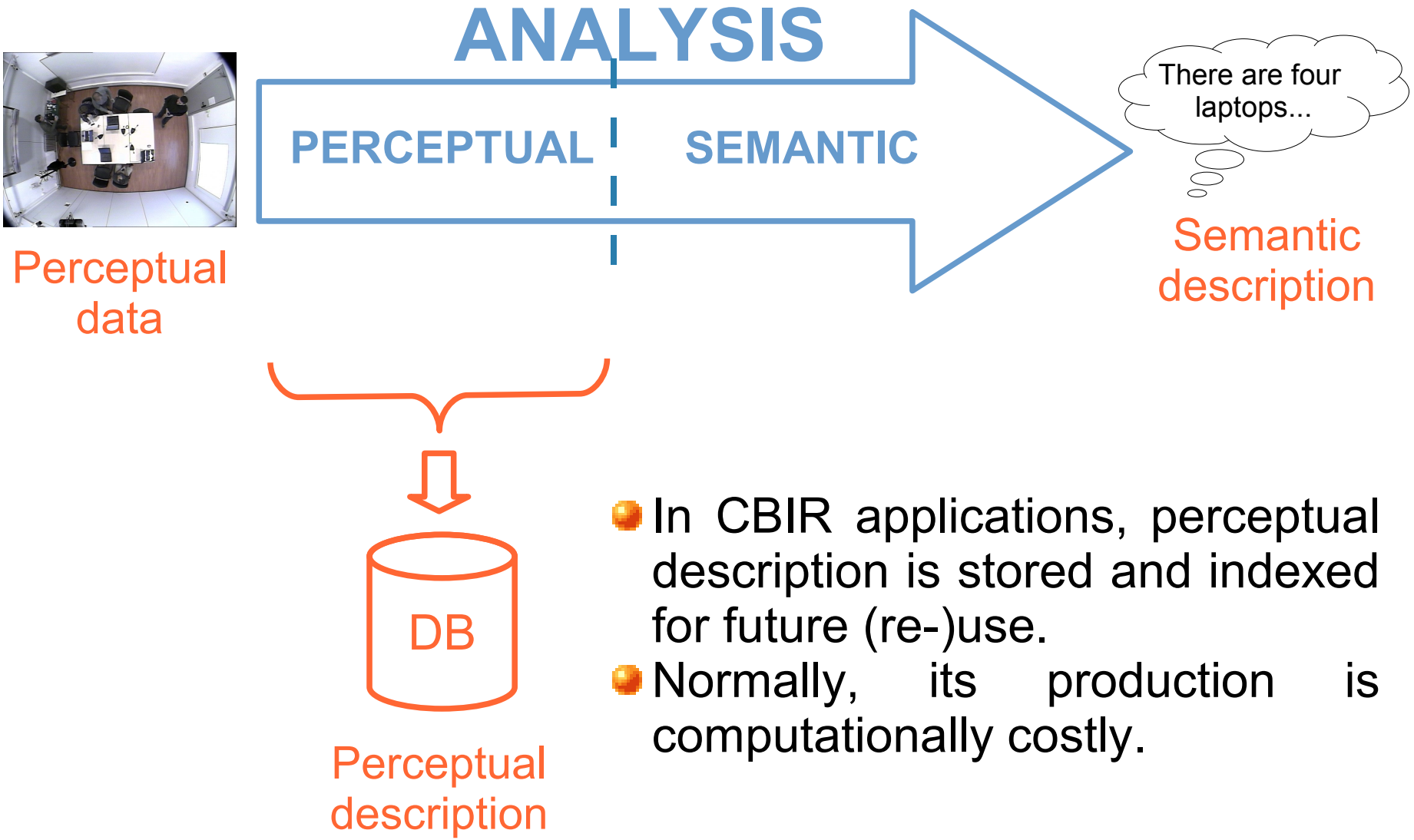
There are four laptops...

# Outline

1. Introduction
2. **Generic algorithm ←**
   - Perceptual Analysis
   - Semantic Analysis
3. Controlled environment
   - Perceptual
   - Semantic
4. Results
5. Conclusions

# Generic algorithm: Perceptual

- The proposed generic analysis can be divided in two parts: a perceptual and a semantic one.




- In CBIR applications, perceptual description is stored and indexed for future (re-)use.
- Normally, its production is computationally costly.

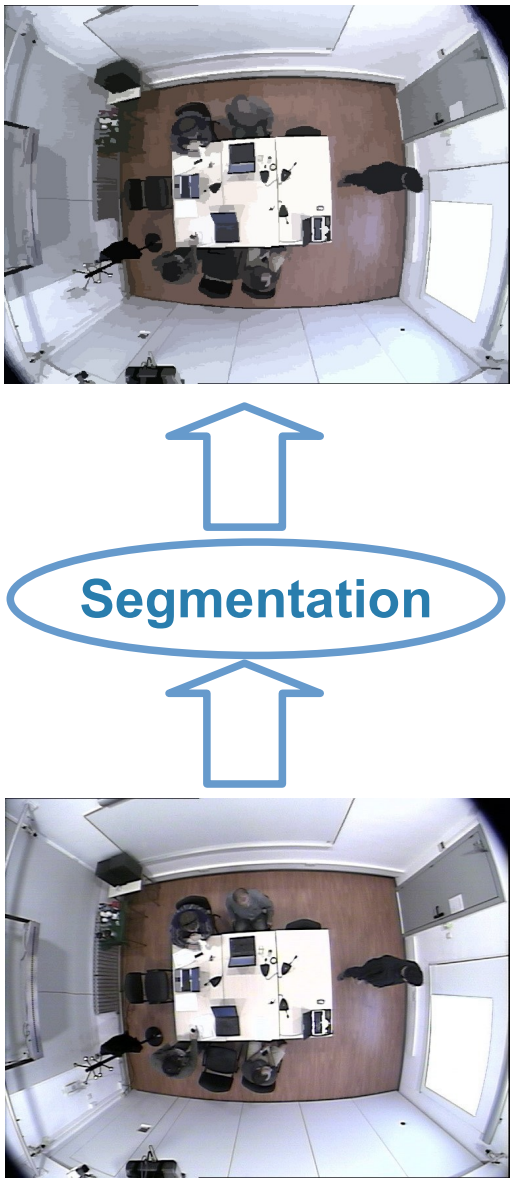
# Generic algorithm: Perceptual

Signal Theory & Communications Department
Technical University of Catalonia (UPC)

## Region-based [1]

- Advantage
  - Objects being union of regions
  - Solves some over-segmentations
  - Simplification: # regions << # pixels

- Limitation
  - Still a large number of possible unions to consider
  - Initial partition may present over- and sub-segmentation problems



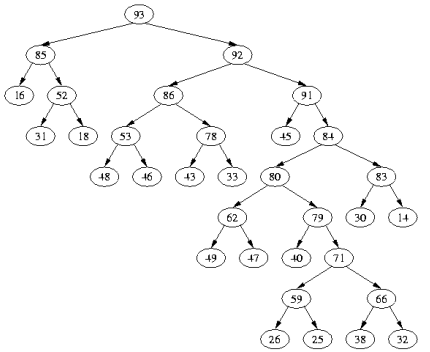

[1] P.Salembier and F.Marqués, “Region-based representations of image and video: segmentation tools for multimedia services”, IEEE Trans. Circuits and Systems for Video Technology (1999).

# Generic algorithm: Perceptual



## Hierarchy-based [2]

Binary Partition Tree (BPT)




- Advantage
  - Reduced universe of region unions
  - Still multi-resolution

BPT creation

- Limitation
  - Merging criterion must match the semantic interpretation.

Regions

[2] P.Salembier and L.Garrido, “Binary Partition Tree as an efficient representation for image processing, segmentation and information retrieval”, IEEE Trans. on Image Processing (2000)
[2+] Related approaches: Quad-trees [Pietikainen], component trees [Monasse], min-max trees.

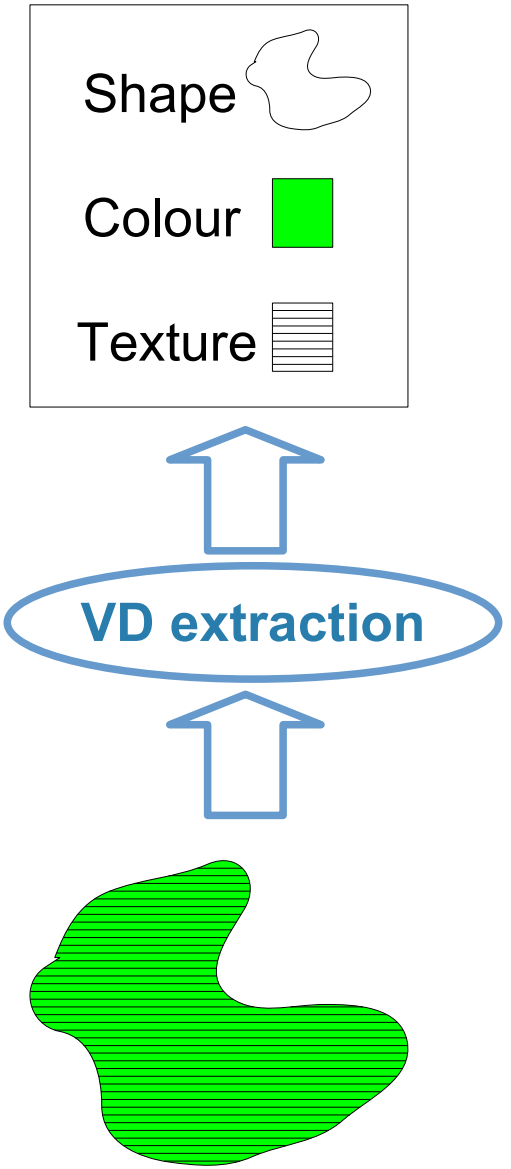
# Generic algorithm: Perceptual



## Visual descriptors [3]

- Advantage
  - → Quantization of compact perceptual features relevant for semantic analysis
- Limitation
  - → Extracted descriptors must efficiently characterize the object (or its parts)
  - → Compromise:

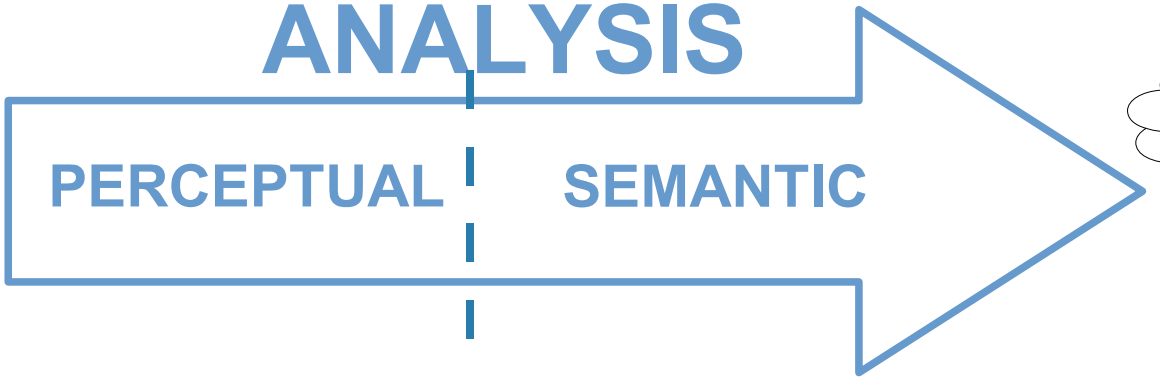
↑↑ descriptors ⇒ ↑↑ computation




[3] Manjunath, Salembier and Sikora, "Introduction to MPEG-7", Wiley (2002)

# Generic algorithm: Semantic

- The proposed generic analysis can be divided in two parts: a perceptual and a semantic one.

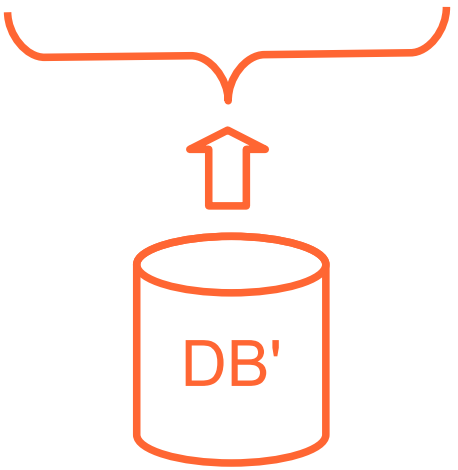



Perceptual data

There are four laptops...

Semantic description

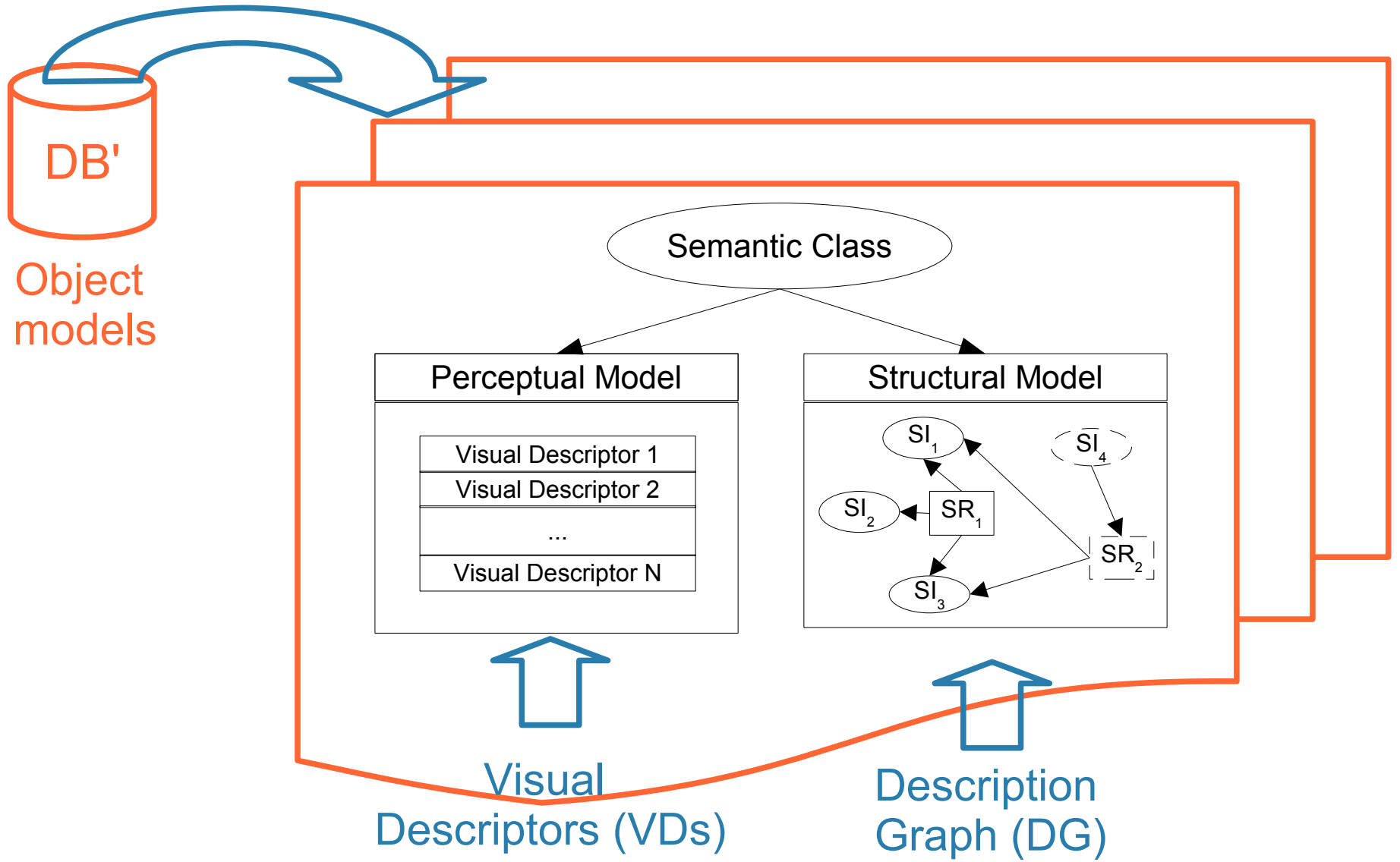
- Object models are built during a learning process.
- User feedback may update/improve model accuracy.




Object models

# Generic algorithm: Semantic

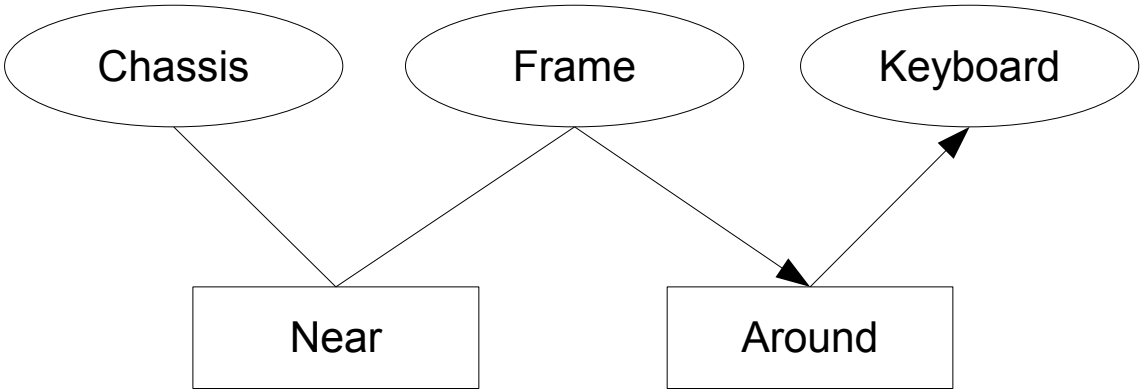
- Dual models: perceptual (simple) and structural (composite).



# Generic algorithm: Semantic

- Structural models with Description Graph [4].
- A Description Graph (DG) models a semantic class by assigning semantic instances and their Relations to its vertices.

Example: DG of the semantic class "Laptop"



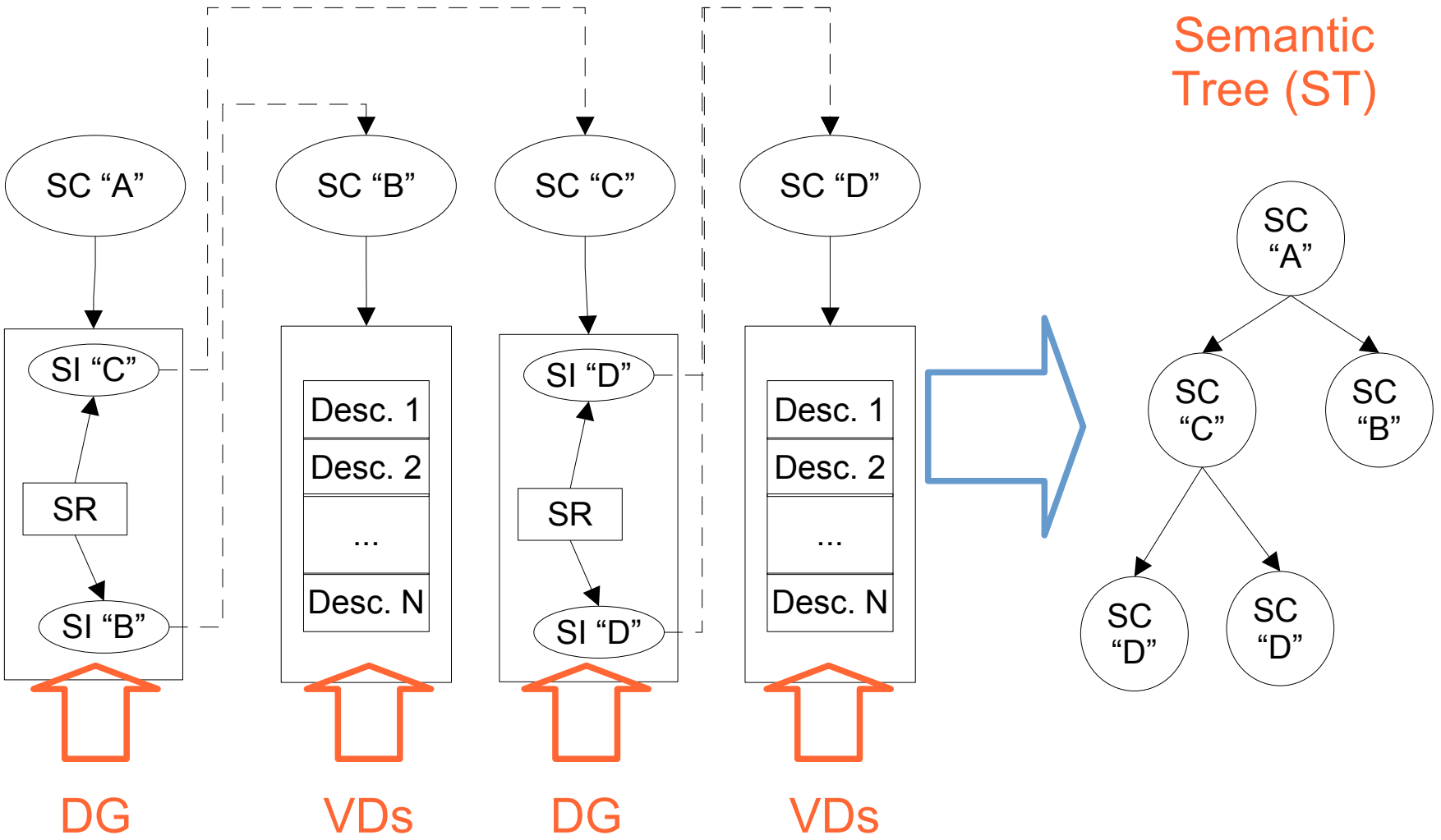

[4] X.Giró and F.Marqués, "Detection of semantic objects using Description Graphs", ICIP, Genoa, Italy (2005).
[4+] Related approaches: Constellations [Burl et al.], Factor Graphs [Kozintsev et al], Attributed Relational Graphs [Jaimes et al, Lee et al].



Signal Theory & Communications Department
Technical University of Catalonia (UPC)

# Generic algorithm: Semantic

- Hierarchical model decomposition in Semantic Trees (STs)



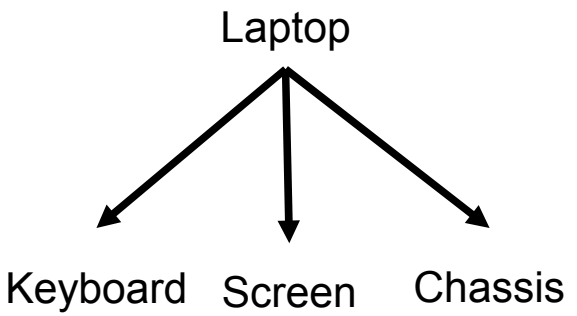
# Generic algorithm: Semantic



## Semantic Trees creation [5]

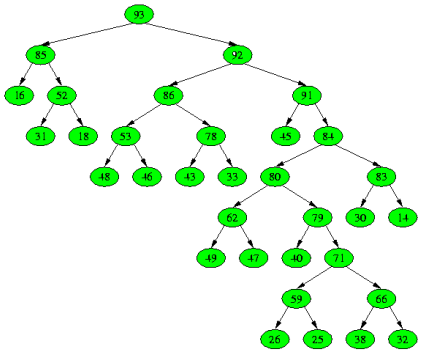
- Advantage
  - → Unique detection algorithm for simple and composite objects.
  - → Supports non-connected objects.
- Limitation
  - → Semantic models and classifier must effectively deal with perceptual variability.

Semantic Tree (ST)




ST creation

BPT + VDs



[5] X.Giró and F.Marqués, "From Partition Trees to Semantic Trees", MRCS. Istanbul 2006.

# Outline

1. Introduction
2. Generic algorithm
   - Perceptual Analysis
   - Semantic Analysis
3. **Context-based analysis ←**
   - Perceptual
   - Semantic
4. Results
5. Conclusions

# Context-based analysis

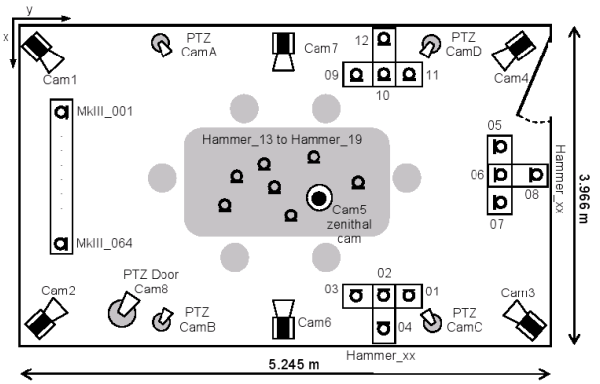
- Problem: The generic algorithm developed for indexing is not efficient in object detection in controlled environment.

- Improvement: Use the context to reduce the search space

## Spatial

- Fixed zenithal camera
- Fixed table
- Static and known background

## Smart room

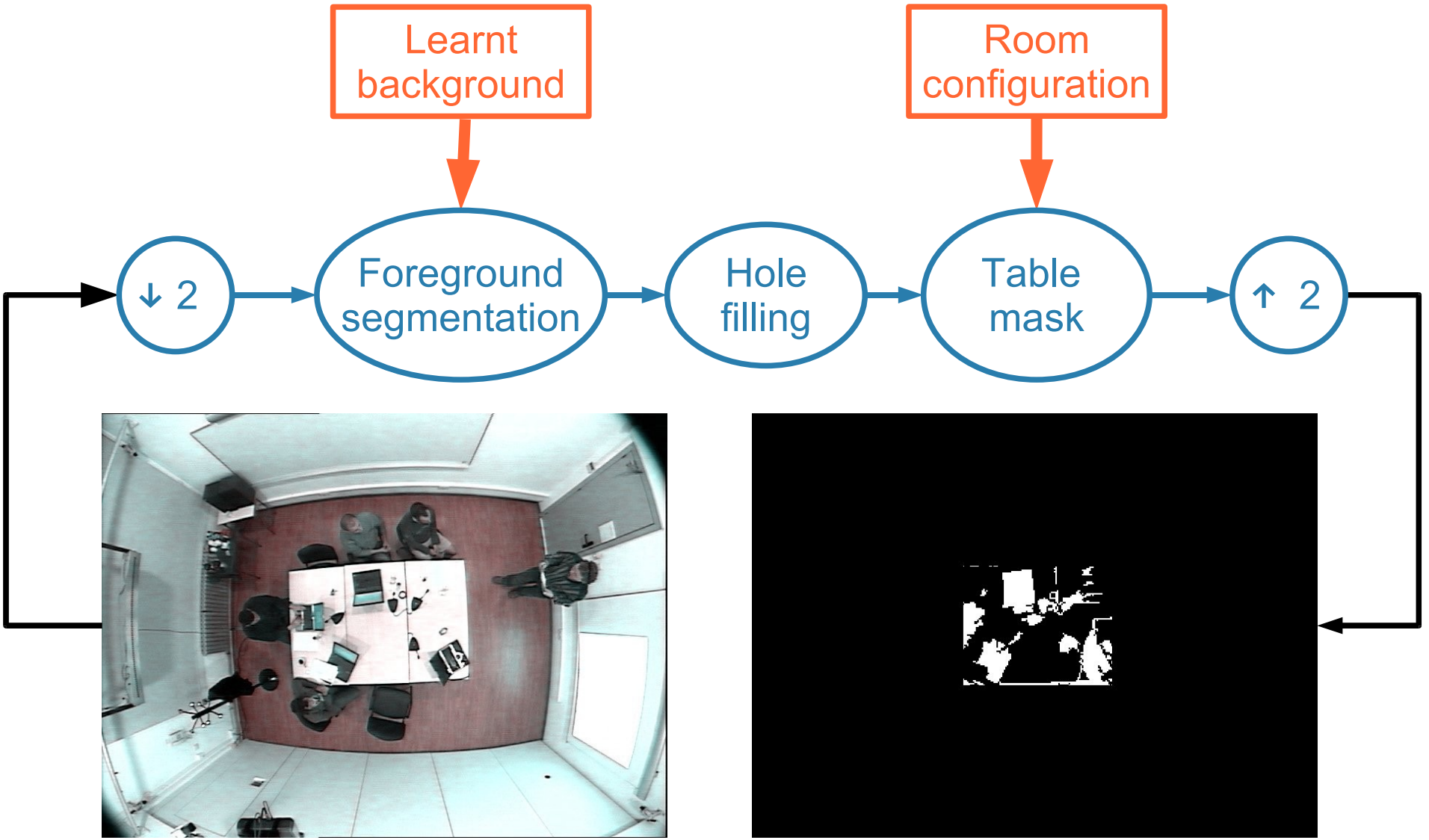



## Application

- Class of interest are laptops.
- Limited universe of laptops.
- Only interested in laptops on the table
- Video sequence

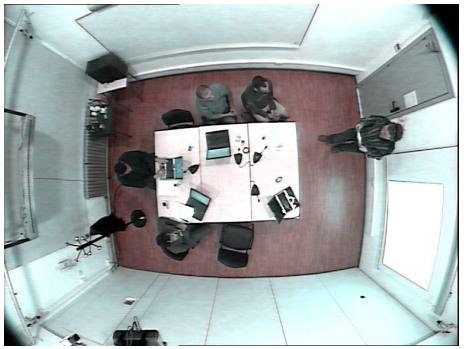
# Perceptual analysis: spatial context

- Dynamic generation of foreground masks on table

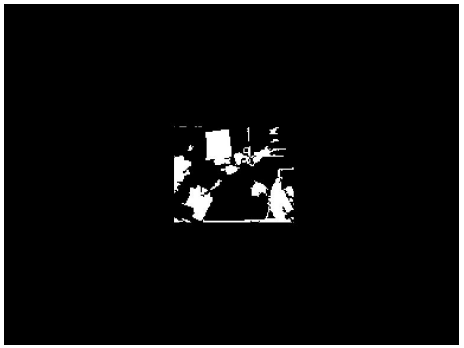


# Perceptual analysis: spatial context

- Disconnected segments in mask generate BPT forests.

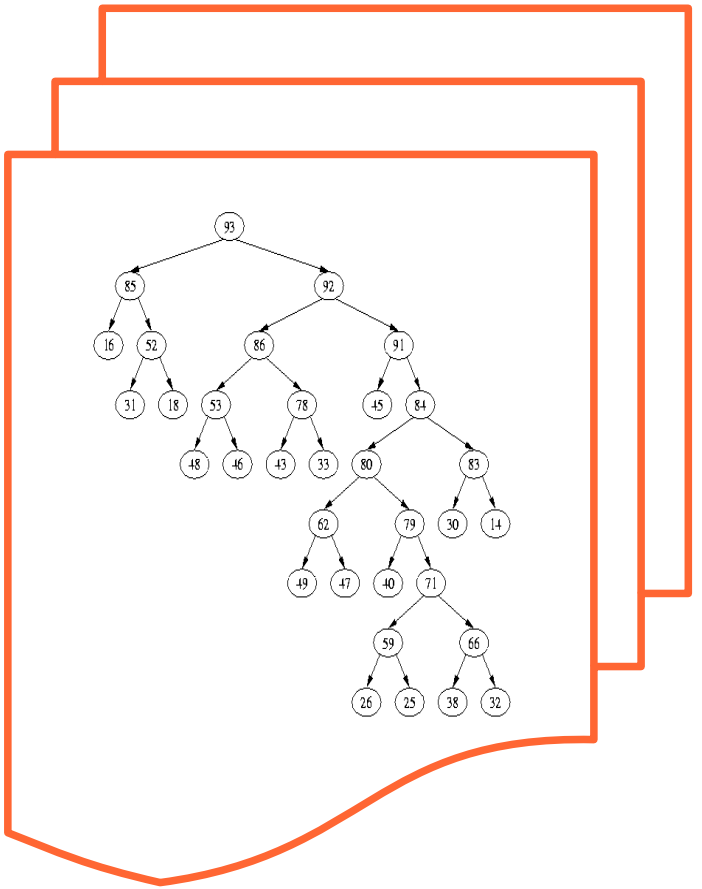


+



N connected segments

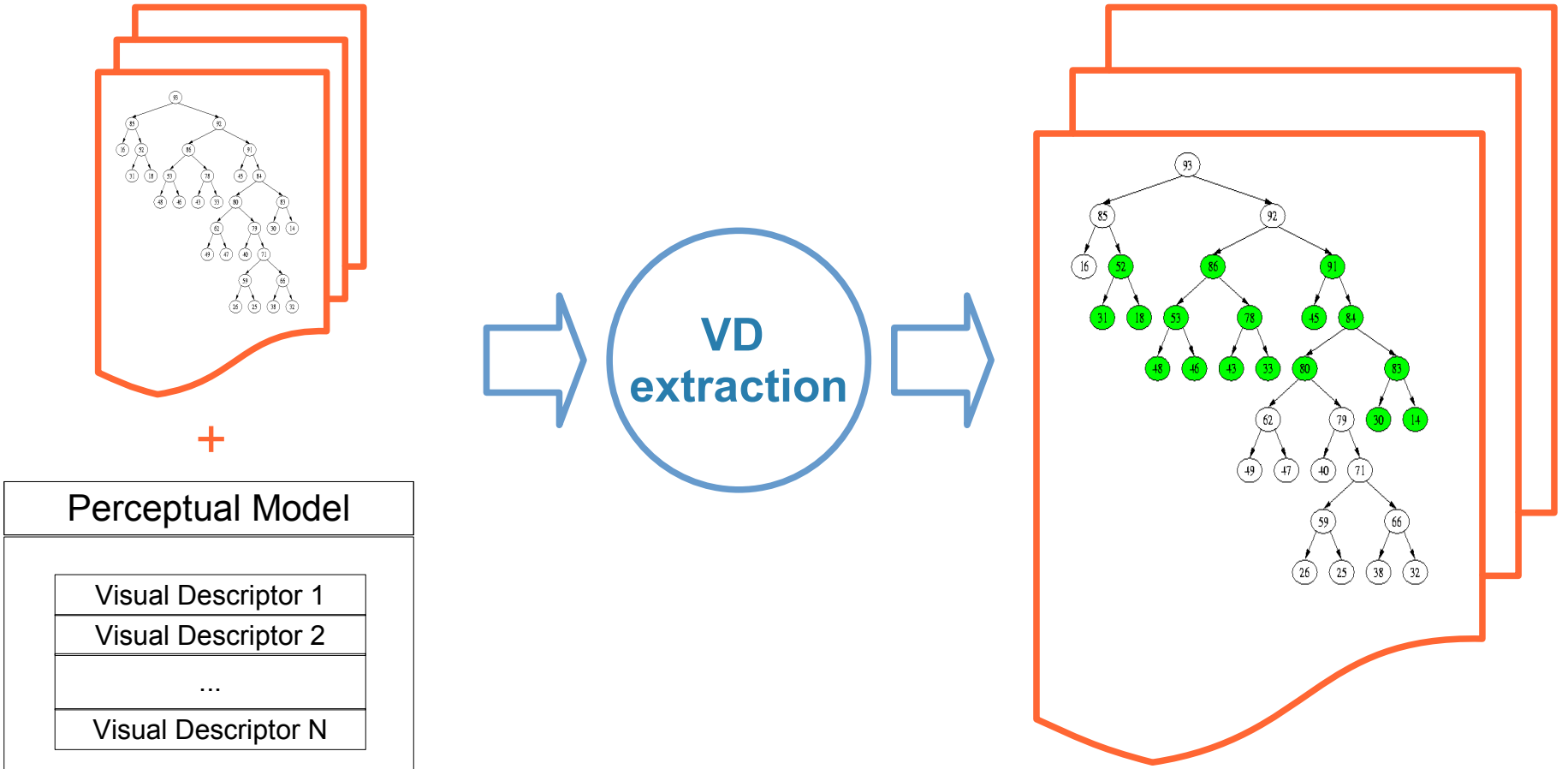
BPT creation




N BPT's (BPT forest)

# Perceptual analysis: application context

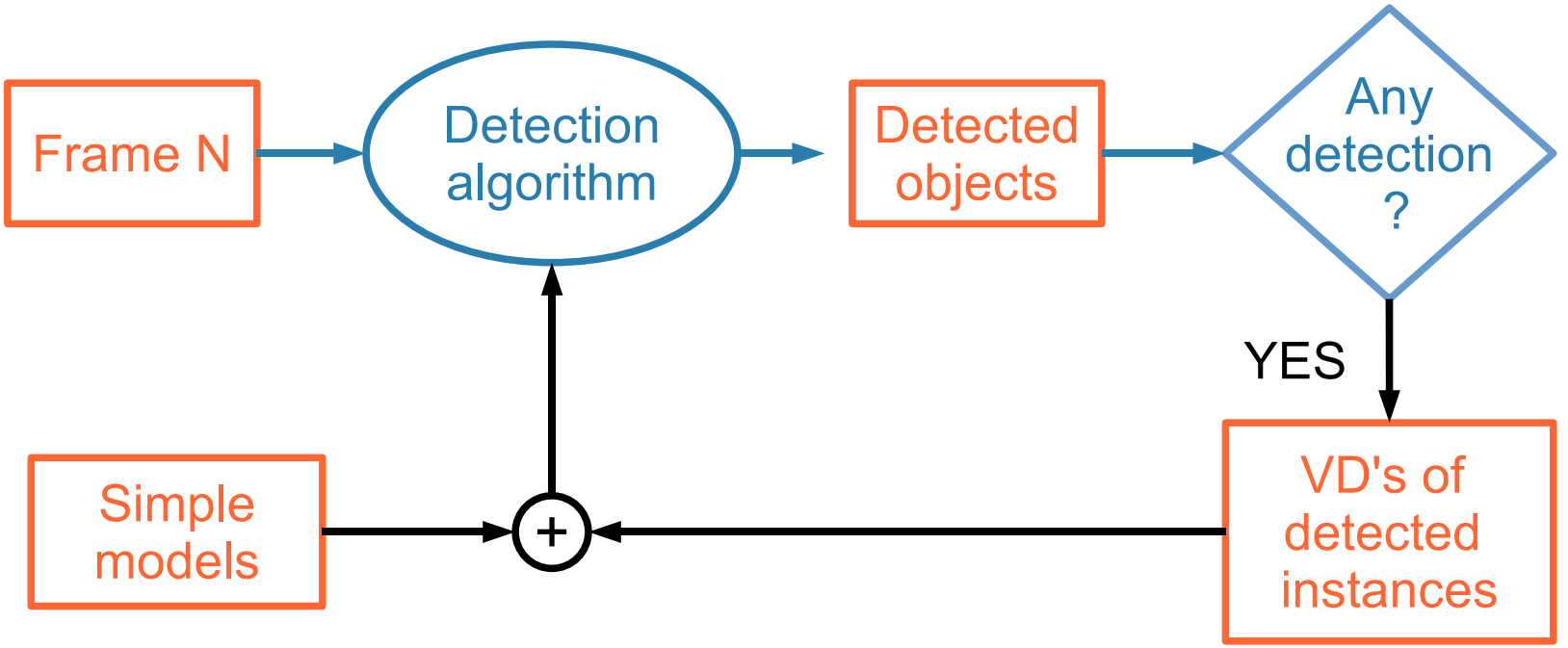
- Simple descriptors in models may discard several BPT nodes for VD extraction.



# Semantic analysis: application context

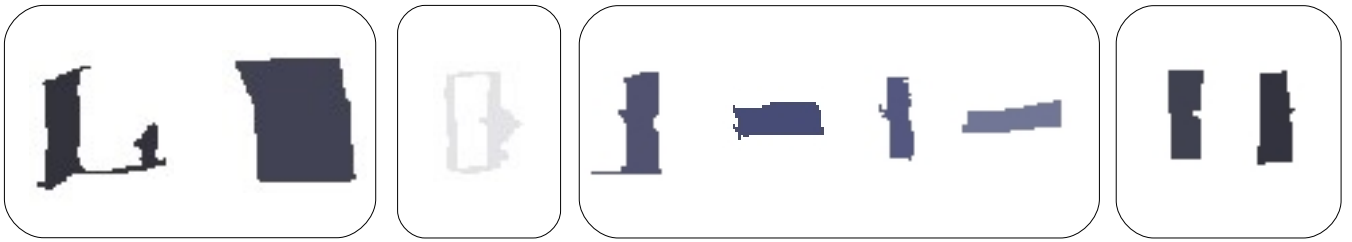
- Problem: Partial occlusions gradually in time.

- Improvement: Simple models are updated after each object detection to cope with gradual changes.

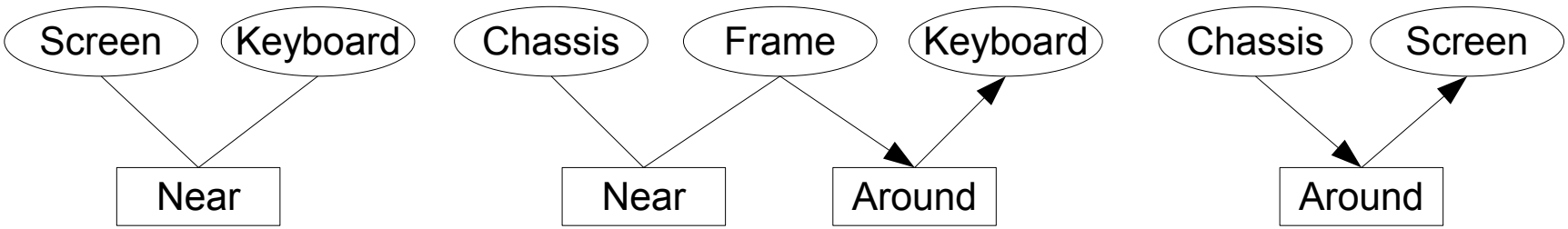


# Semantic analysis: application context

- Problem: A single semantic class can be represented by multiple simple and/or composite models.
- Improvement: Support multiple models for single class.
  - simple: multiple sets of descriptors/classifiers




  - composite: multiple structural models (eg. DG's)



# Outline

1. Introduction
2. Generic algorithm
   - Perceptual Analysis
   - Semantic Analysis
3. Context-based analysis
   - Perceptual
   - Semantic
4. **Results ←**
5. Conclusions

# Results

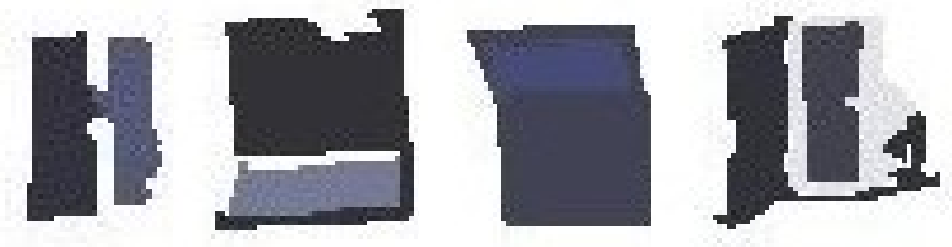
- Perceptual analysis in context

  Dynamic Masks → 90-99% pixels discarded

  Discriminative descriptor : (area) → 75-80% regions discarded

- Semantic analysis in context

  Detection of four laptops with perceptual variability



1 2 3 4

| Laptop | Precision |
|---|---|
| 1 | 0,8 |
| 2 | 0,86 |
| 3 | 0,8 |
| 4 | 0,97 |

# Outline

1. Introduction
2. Generic algorithm
   - Perceptual Analysis
   - Semantic Analysis
3. Context-based analysis
   - Perceptual
   - Semantic
4. Results
5. **Conclusions ←**

# Conclusions

- Methodology: CBIR → Real time object detection.
- Context provides optimization.
  - → Spatial: less pixels to analyse.
  - → Semantic: less descriptors to extract.
  - → Temporal: models update

- Future goals:
  - → Study of descriptors according to their BPT expansion, discrimination and computation cost.

# Outline

Further details: http://gps-tsc.upc.es/imatge/_Xgiro/start.html

1. Introduction
2. Generic algorithm
   - Perceptual Analysis
   - Semantic Analysis
3. Context-based analysis
   - Perceptual Analysis
   - Semantic Analysis
4. Results
5. Conclusions

Thank your for your attention