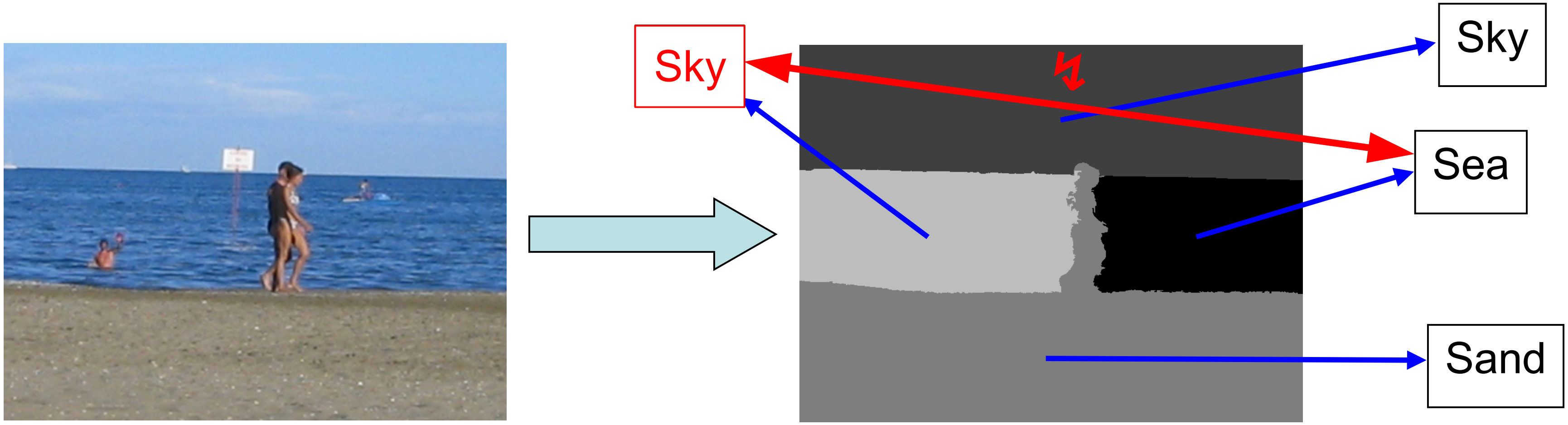
Labelling Image Regions Using Wavelet Features and Spatial Prototypes

Carsten Saathoff, Marcin Grzegorzek and Steffen Staab
SAMT 2008, Koblenz, Germany

Motivation



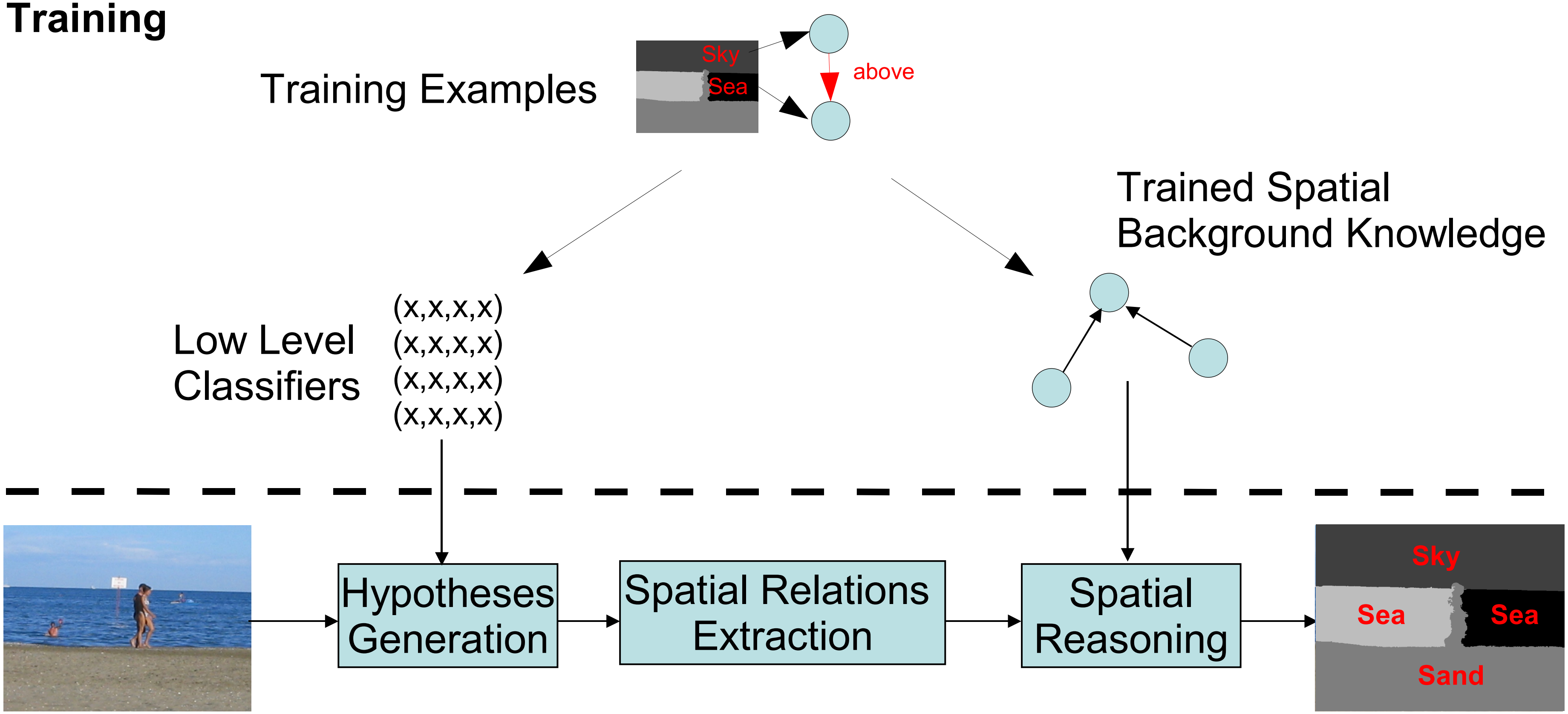
- Local features often not sufficient for classification
- Exploit explicitly defined spatial knowledge to improve labelling
  - e.g. Sky not allowed left or right of Sea
- Allow for efficient training of classifiers and spatial knowledge
  - Good labelling performance with few training examples

Agenda

- Analysis Framework
- Exploiting Spatial Context, 1st try
  - Fuzzy Constraint Satisfaction (WIAMIS08)
- Contribution
- Low-Level Region Classification
  - Training of statistical models
  - Classification
- Exploiting Spatial Context, revisited
  - Binary Integer Programming
- Evaluation
- Conclusions

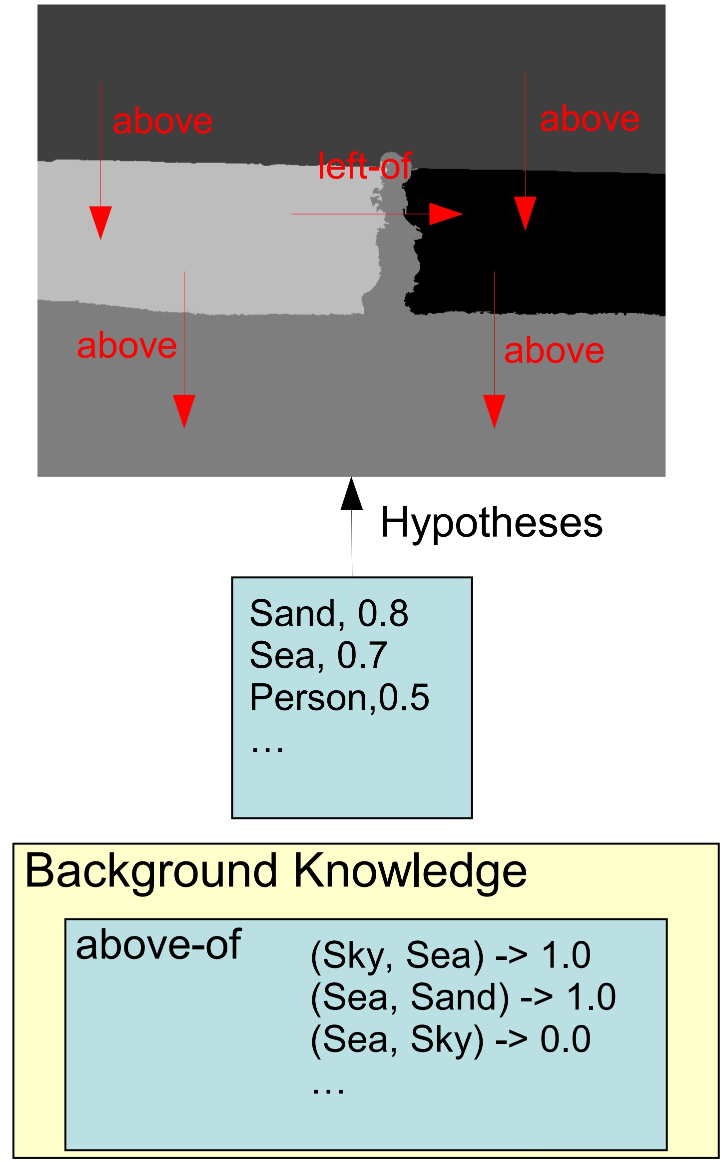
Analysis Framework

Training

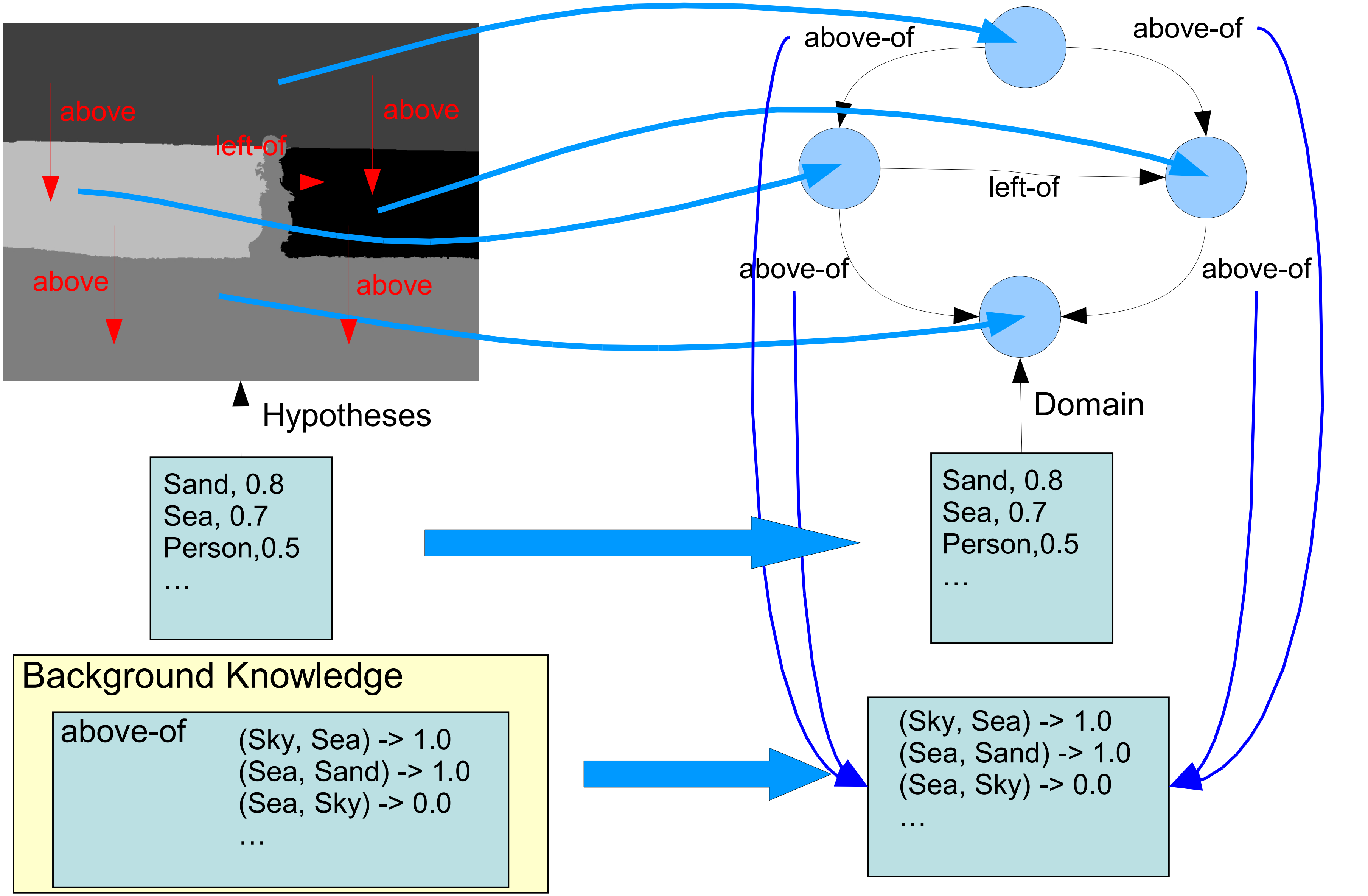


Analysis

- Create appropriate optimization problem from
  - Regions
  - Spatial Relations
  - Hypotheses sets
  - Spatial background knowledge
- Approaches
  - Fuzzy Constraint Satisfaction
    - WIAMIS08
  - Binary Integer Programming
    - later...

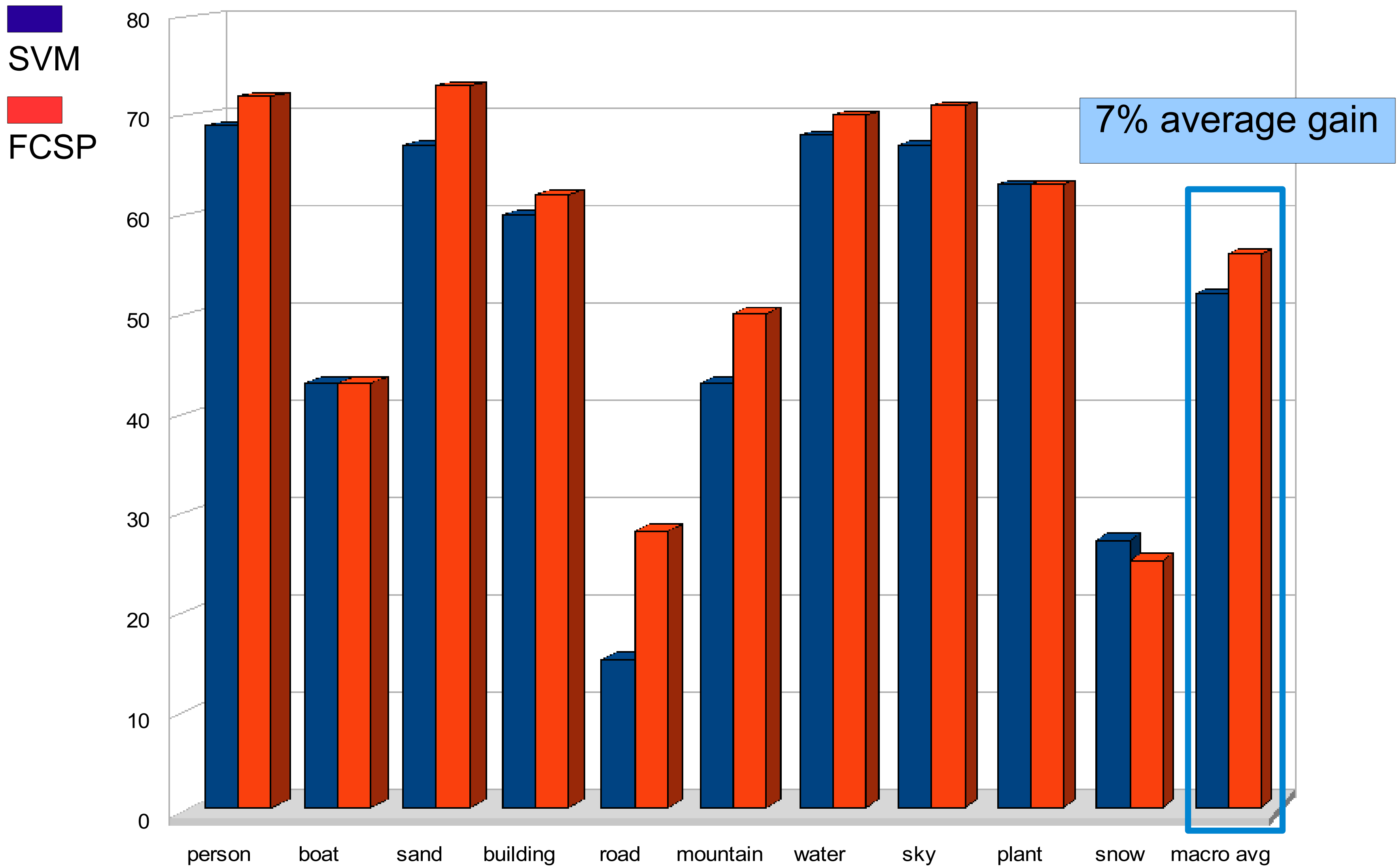


Region Labelling as a FCSP



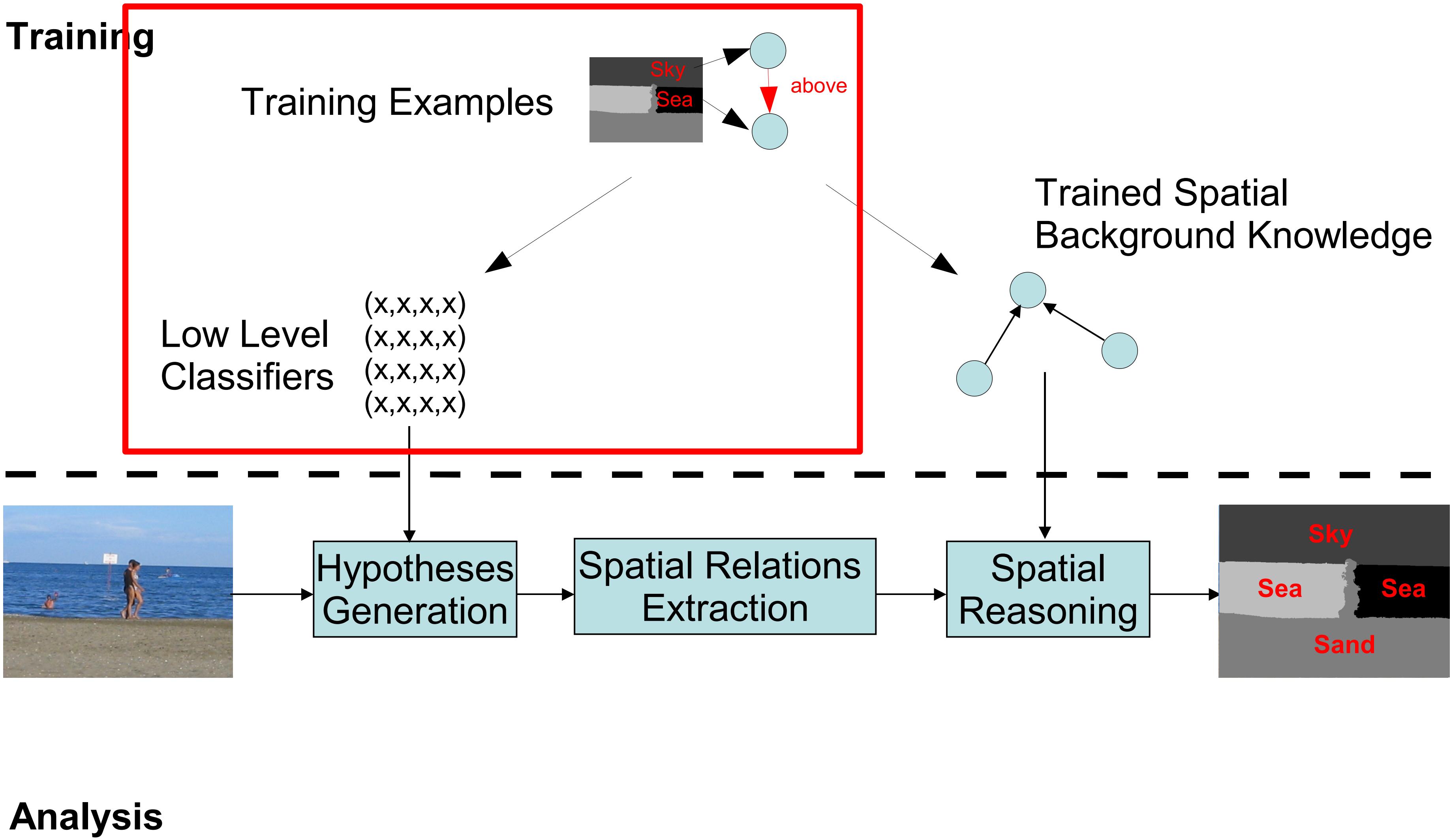
F-Measure (WIAMIS08)

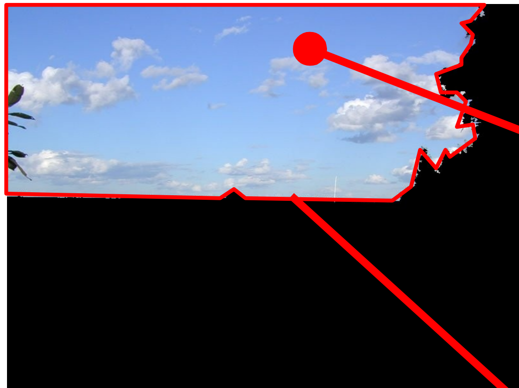
F-Measure for all concepts average F-Measure.

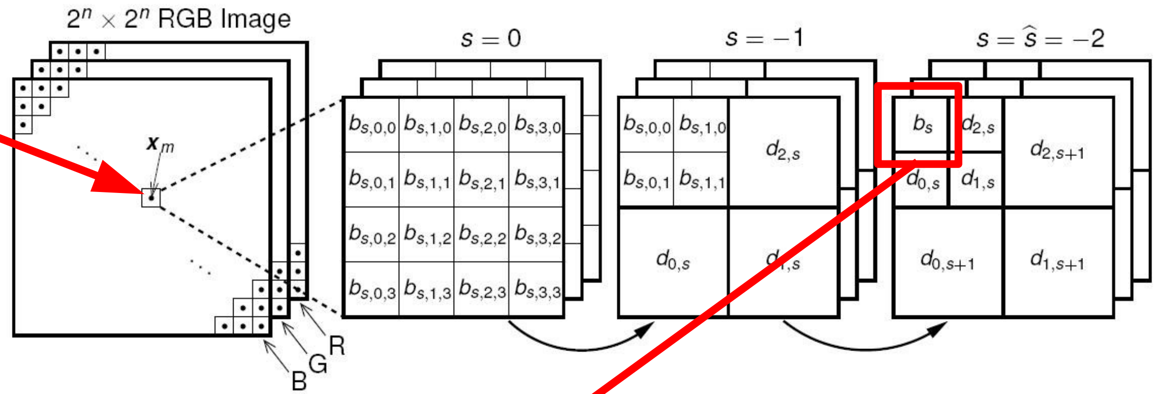


Contribution

- Region-Level classification based on
  - Wavelet Features
  - Statistical Classification
- Spatial Reasoning
  - Training of Spatial Knowledge
  - Formalization as Binary Integer Programming
- Comparison with alternative approach over different training set sizes


Training
Training Examples
Sky
Sea
above
Trained Spatial
Background Knowledge
Low Level
Classifiers
(x,x,x,x)
(x,x,x,x)
(x,x,x,x)
(x,x,x,x)
Hypotheses Generation
Spatial Relations Extraction
Spatial Reasoning
Sky
Sea
Sea
Sand
Analysis





Mean pixel value

Mean value over wavelet coefficients of all neighborhoods

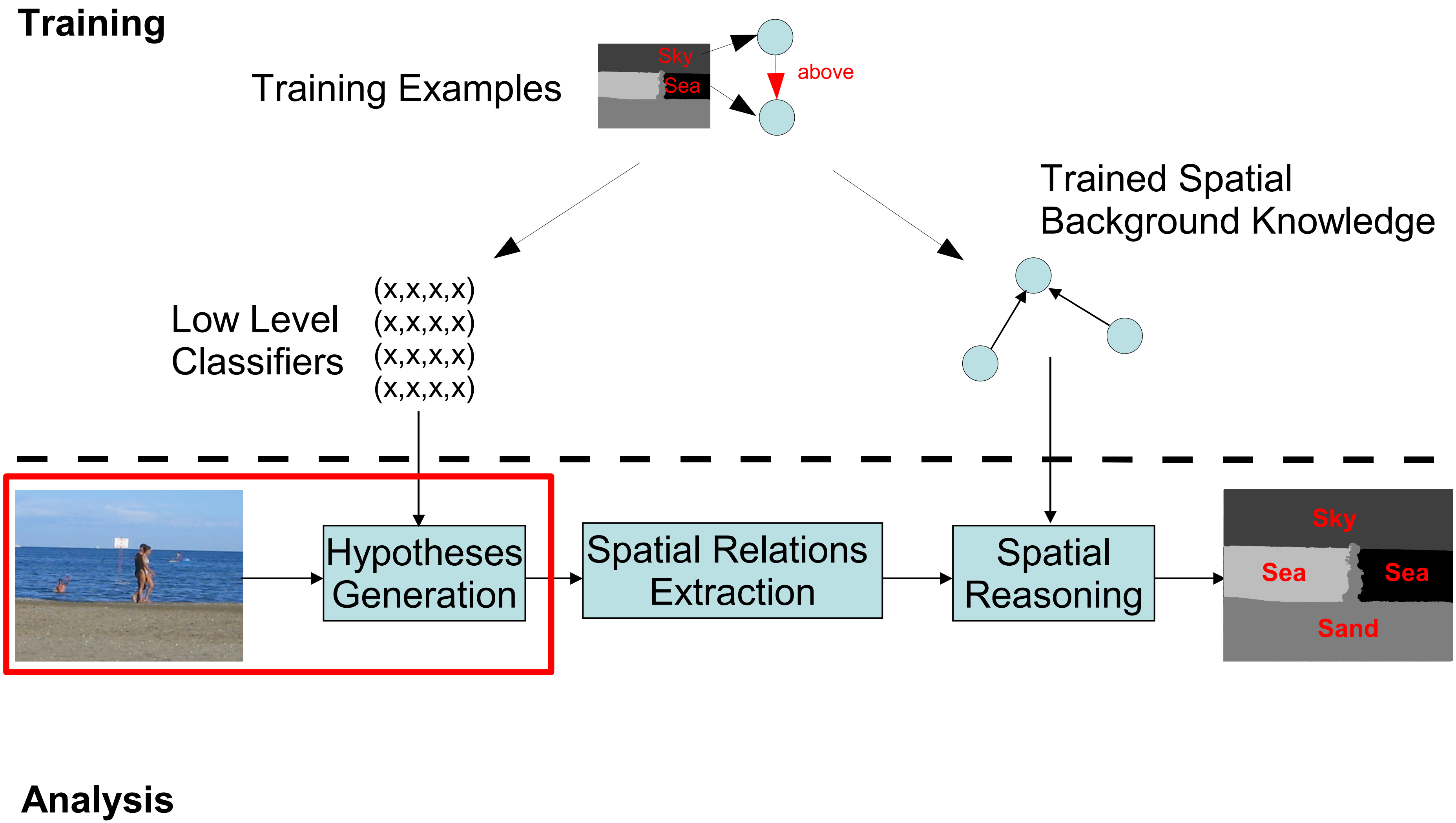
Feature Vector: (m, b_R, b_G, b_B)

Compute mean and standard deviation

Statistical model for concept $\boldsymbol{K}$

$$\boldsymbol{\mu}_K = (\mu_m, \mu_R, \mu_G, \mu_B)$$
$$\boldsymbol{\sigma}_K = (\sigma_m, \sigma_R, \sigma_G, \sigma_B)$$

Analysis Framework



<isweb>

Feature extraction: $c_r = (m, b_R, b_G, b_B)$

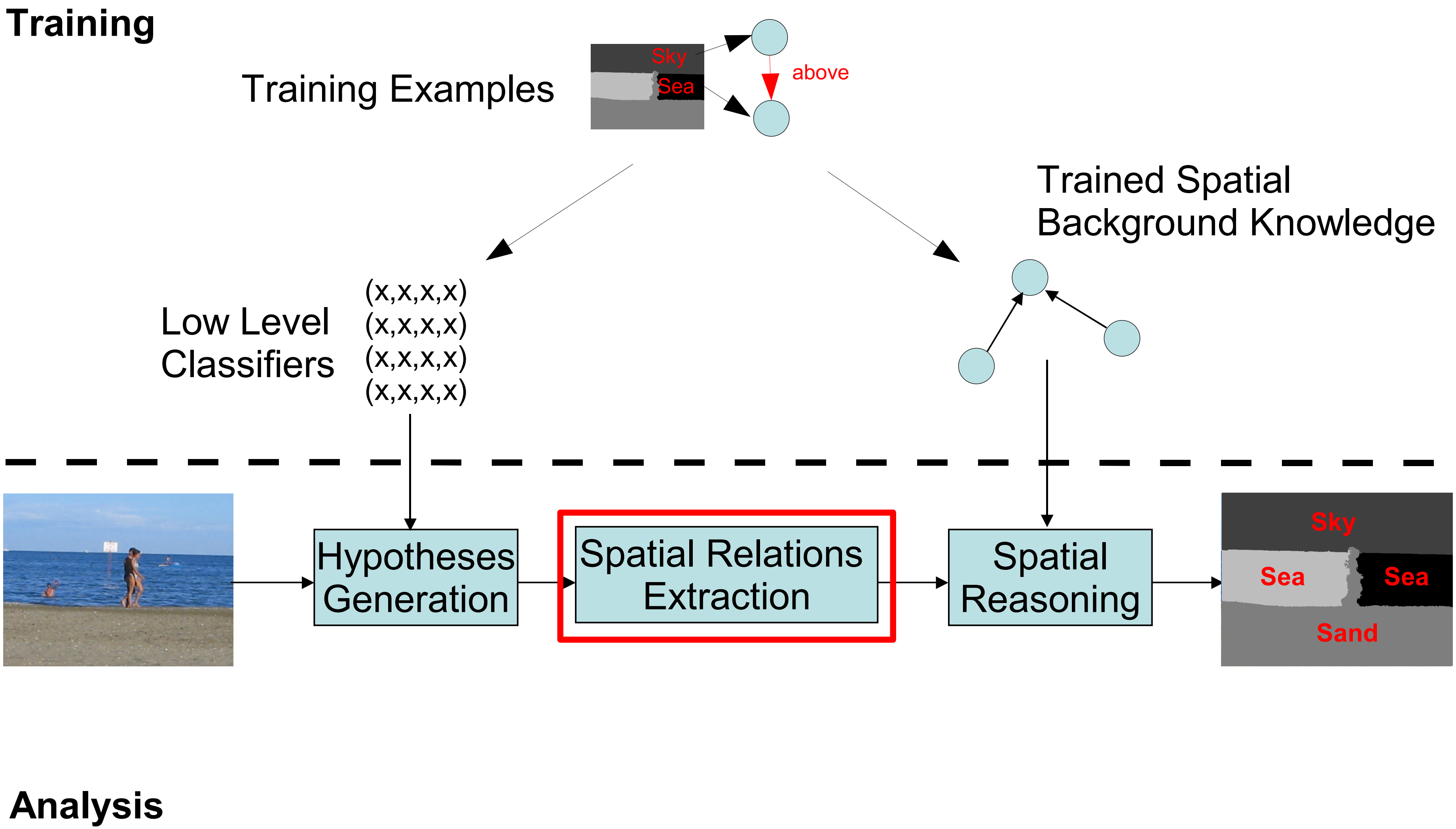
Image Region

$f_r \rightarrow c_r$

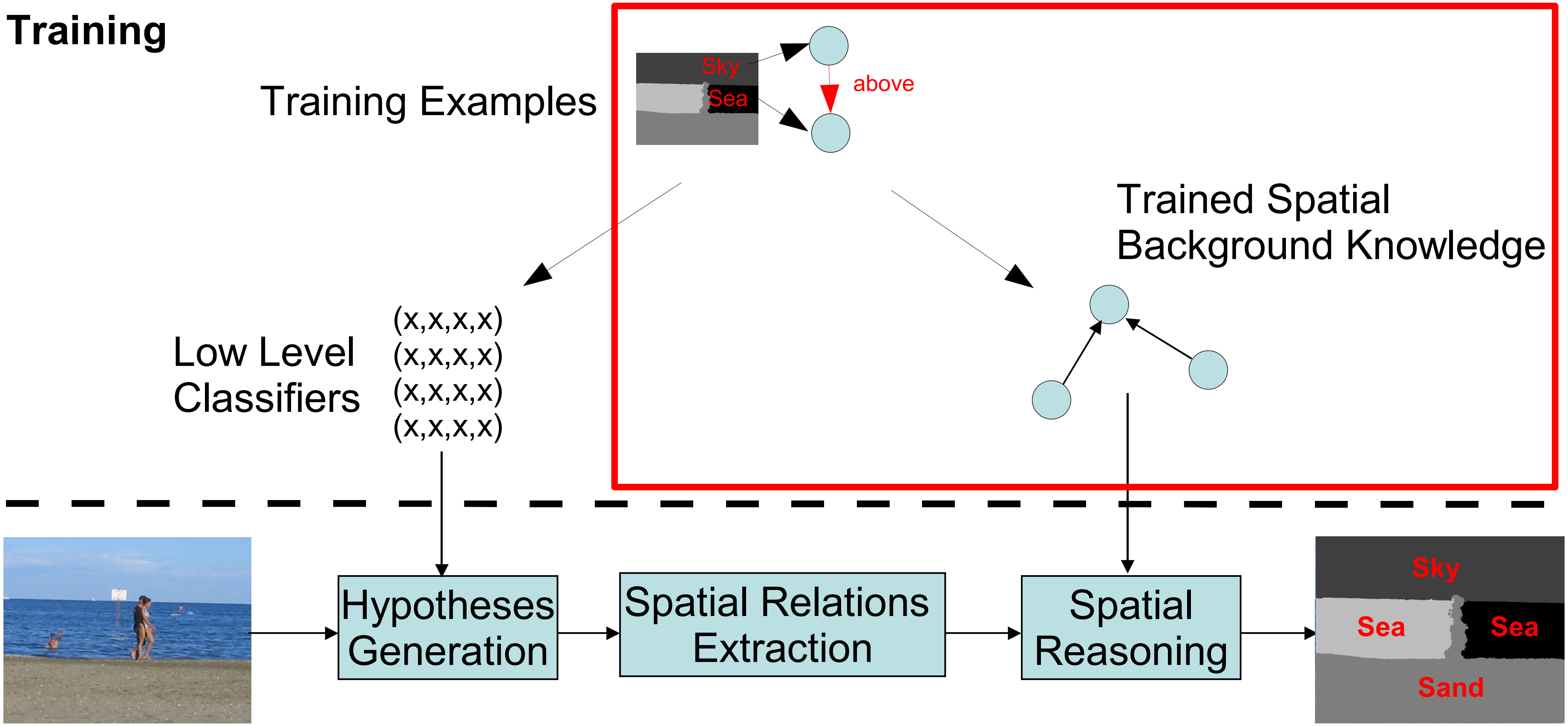
Sky, 0.8
Sea,0.76
Sand, 0.68
Person, 0.67
Building, 0.54
…

Input for spatial reasoning

Analysis Framework



Analysis Framework



Analysis

Constraint Acquisition

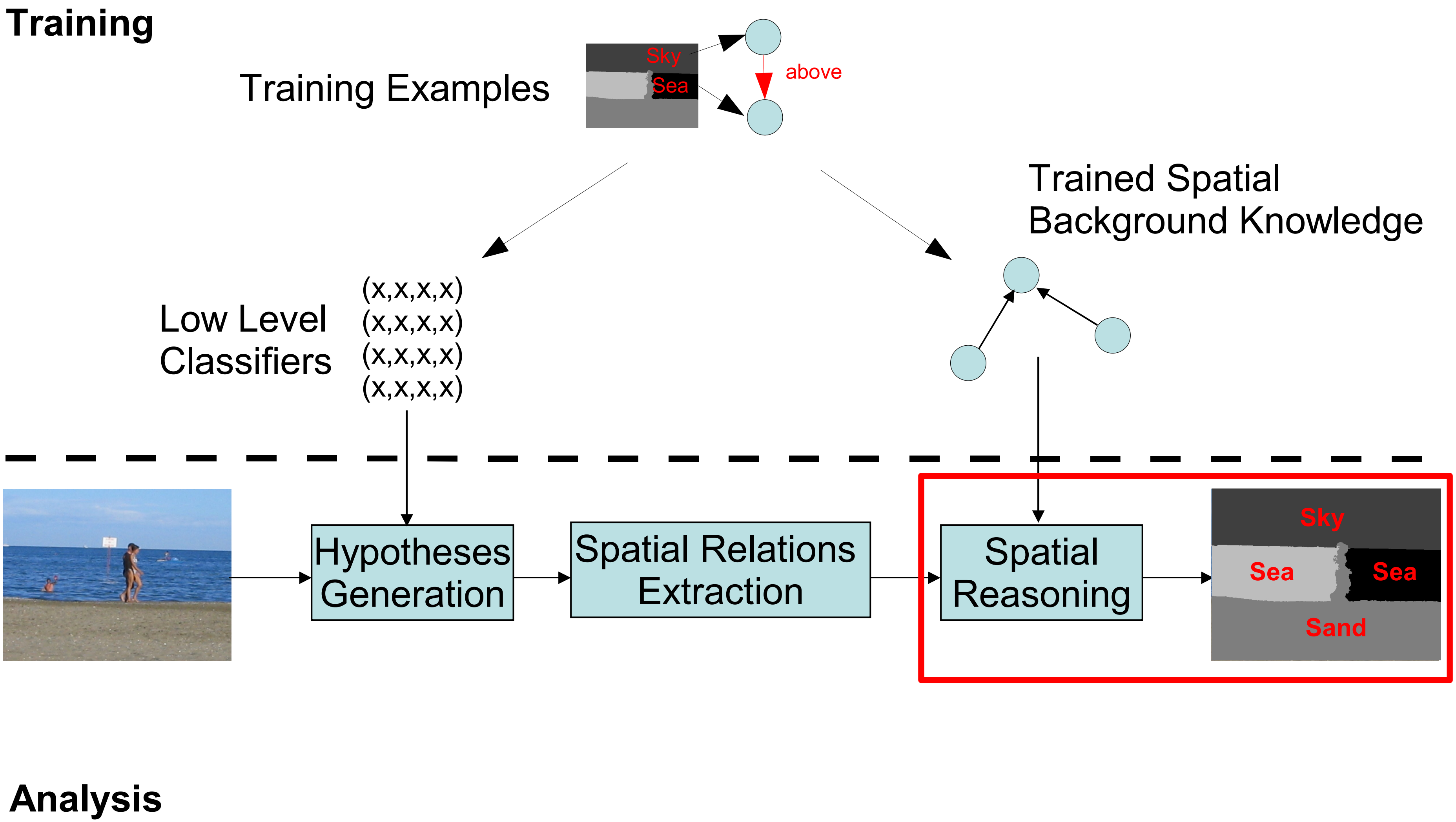
- Background Knowledge consists of *Spatial Constraint Templates*
  - degree of satisfaction for spatial arrangements of concepts
  - Simplest version: crisp degrees
    - e.g. above = (Sky, Sea):**1.0**, (Sea, Sky):**0.0**
- Acquired by mining from a set of labelled examples (spatial prototypes)
  - mining based on confidence (and support)

Background Knowledge

above-of
(Sky, Sea) -> 1.0
(Sea, Sand) -> 1.0
(Sea, Sky) -> 0.0
…

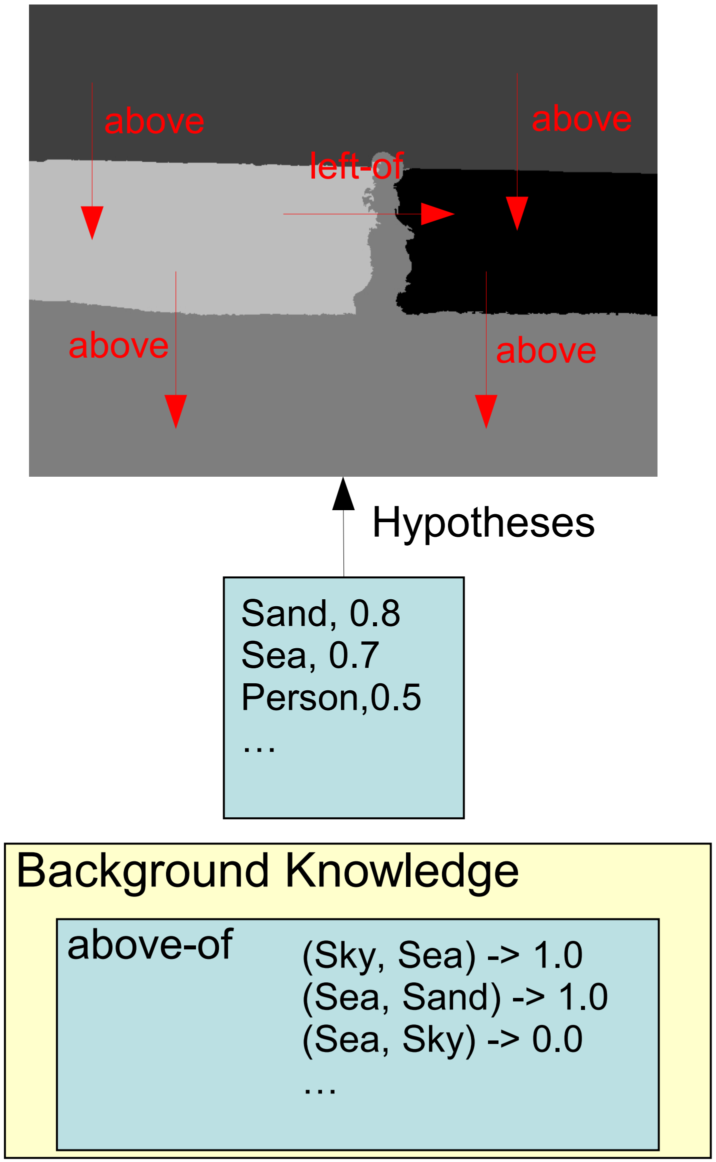
left-of
(Sky, Sea) -> 0.0
(Sea, Sand) -> 1.0
(Sea, Sea) -> 1.0
…

Analysis Framework



Exploiting Spatial Context

- Create appropriate optimization problem from
  - Regions
  - Spatial Relations
  - Hypotheses sets
  - Spatial background knowledge
- Approaches
  - Fuzzy Constraint Satisfaction
    - WIAMIS08
  - **Binary Integer Programming**



Problem of the form:

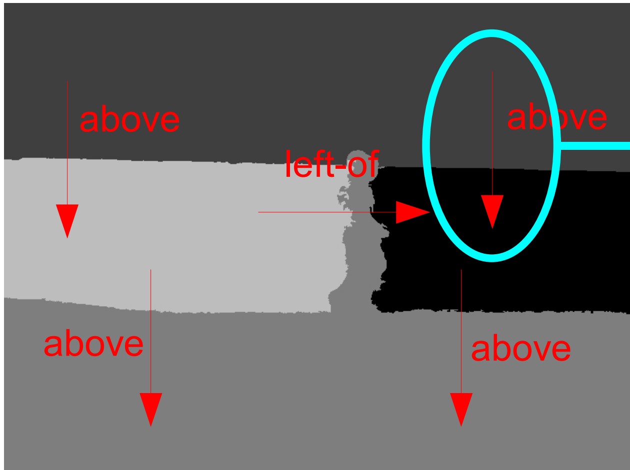
$$\text{minimise } Z = \boldsymbol{c}^T \boldsymbol{x}$$
$$st.$$
$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$$
$$\boldsymbol{x} \in \{0,1\}$$

Example: Assignment Problem

- $i=1...n$ workers
- $j=1..n$ machines
- c_{ij} cost of assigning worker i to machine j
- $x_{ij}=1$ assigns worker i to machine j

$$\text{minimise}$$
$$\sum_{j=1}^{n} \sum_{i=1}^{n} c_{ij} x_{ij}$$
$$st.$$
$$\sum_{j=1}^{n} x_{ij} = 1 \text{ for } i = 1, \ldots, n$$
$$\sum_{i=1}^{n} x_{ij} = 1 \text{ for } j = 1, \ldots, n$$
$$x_{ij} \in \{0,1\}$$

Spatial Reasoning as a BIP



Hypotheses

Sand, 0.8
Sea, 0.7
Person,0.5
…

Background Knowledge

above-of

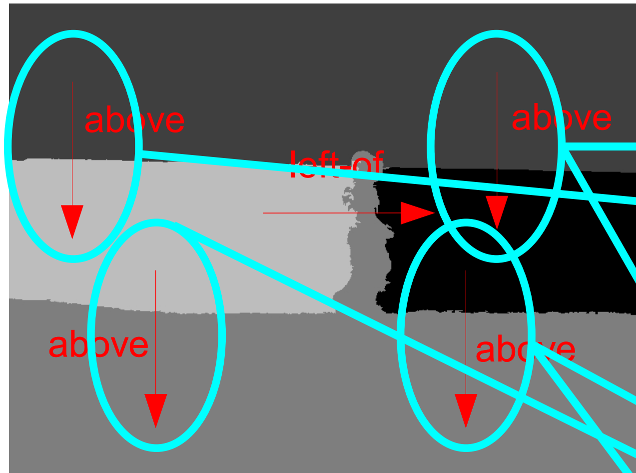
(Sky, Sea) -> 1.0
(Sea, Sand) -> 1.0
(Sea, Sky) -> 0.0
…

Set of variables: c_{itj}^{ko}

$$c_{itj}^{ko} = 1 \Rightarrow s_i = l_k, s_j = l_o$$

Only one assignment per relation

$$\forall r_t = (s_i, s_j) : \sum_{l_k} \sum_{l_o} c_{itj}^{ko} = 1$$



Hypotheses

Sand, 0.8
Sea, 0.7
Person,0.5
…

Background Knowledge

above-of	(Sky, Sea) -> 1.0 (Sea, Sand) -> 1.0 (Sea, Sky) -> 0.0 …

Link outgoing relations:

Base relation: r_{t_O}

Link remaining outgoing relations to base relation:

$$\forall l_k : \sum_{l_o} c^{ko}_{it_O j} - \sum_{l_{o'}} c^{ko'}_{it j'} = \cdot$$

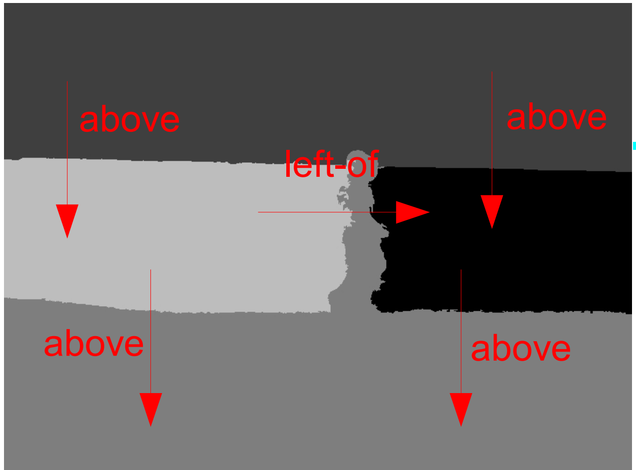
Accordingly for incoming relations:

$$\forall l_k : \sum_{l_o} c^{ok}_{j t_E i} - \sum_{l_{o'}} c^{o'k}_{j' ti} = 0$$

Link outgoing and incoming relations

$$\forall l_k : \sum_{l_o} c^{ko}_{it_O j} - \sum_{l_{o'}} c^{o'k}_{j' t_E i} = 0$$

Objective Function



Hypotheses

Sand, 0.8
Sea, 0.7
Person,0.5
…

Background Knowledge

above-of

(Sky, Sea) -> 1.0
(Sea, Sand) -> 1.0
(Sea, Sky) -> 0.0
…

Set of variables: c_{itj}^{ko}

Linking constraints

Hypotheses $\Theta_i(l_k)$

Templates $T_t(l_k, l_o)$

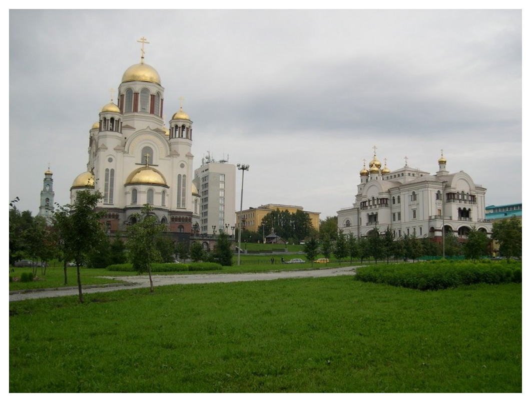
Objective Function:

$$\sum_{r_t}\sum_{l_k}\sum_{l_o} min(\Theta_i(l_k), \Theta_j(l_o)) * T_t(l_k, l_o) * c_{itj}^{ko}$$

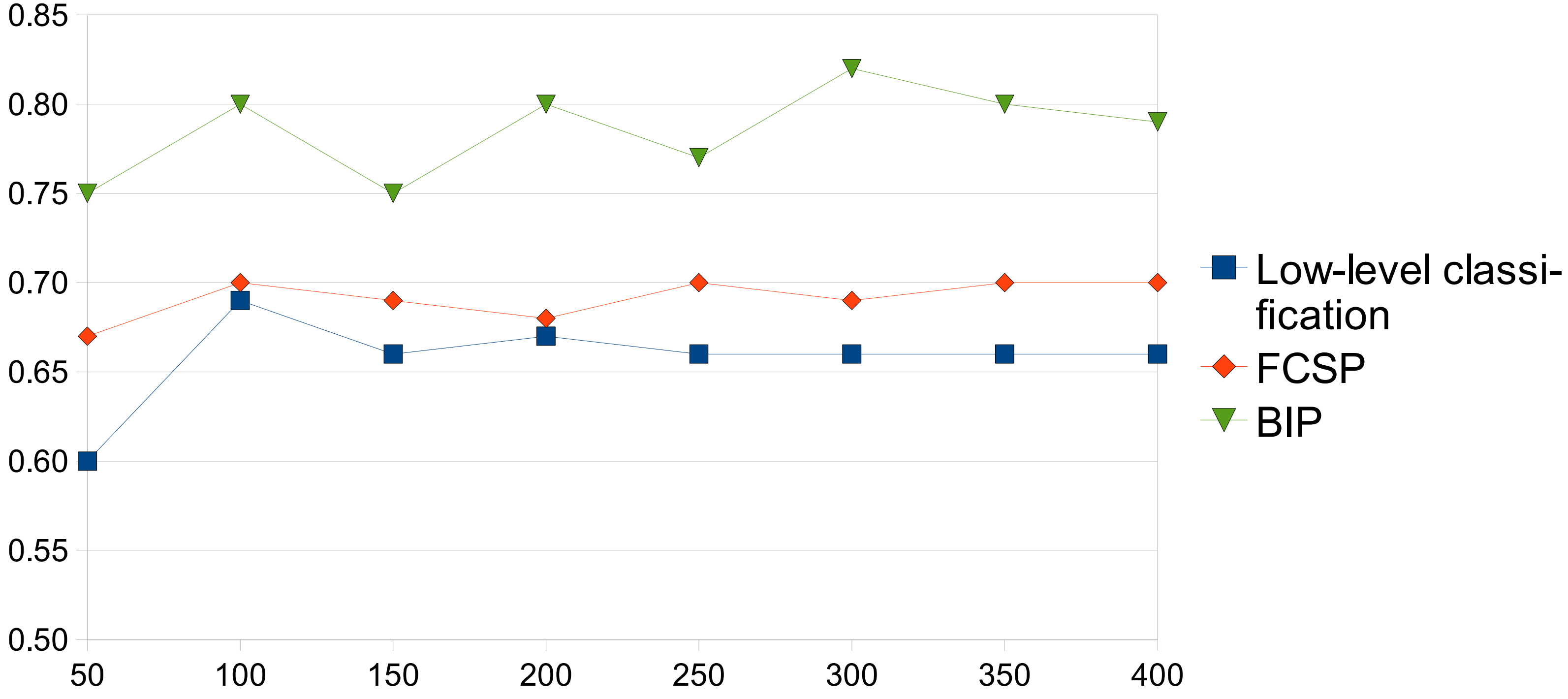
Experimental Setup

- Measure improvement achieved with
  - FCSP
  - Linear Programming
- over low-level classification with different training set sizes
- Data set with 923 natural and urban images
- Set of 10 concepts
  - building,foliage,mountain,person,road,sailing-boat,sand,sea,sky,snow
  - 5690 labelled regions
  - 568 labelled „unknown“ -> ignored
- Ground truth defined on automatic segmentation
  - Regions labelled with dominant concept
  - Ground truth created for this work

Overview of dataset

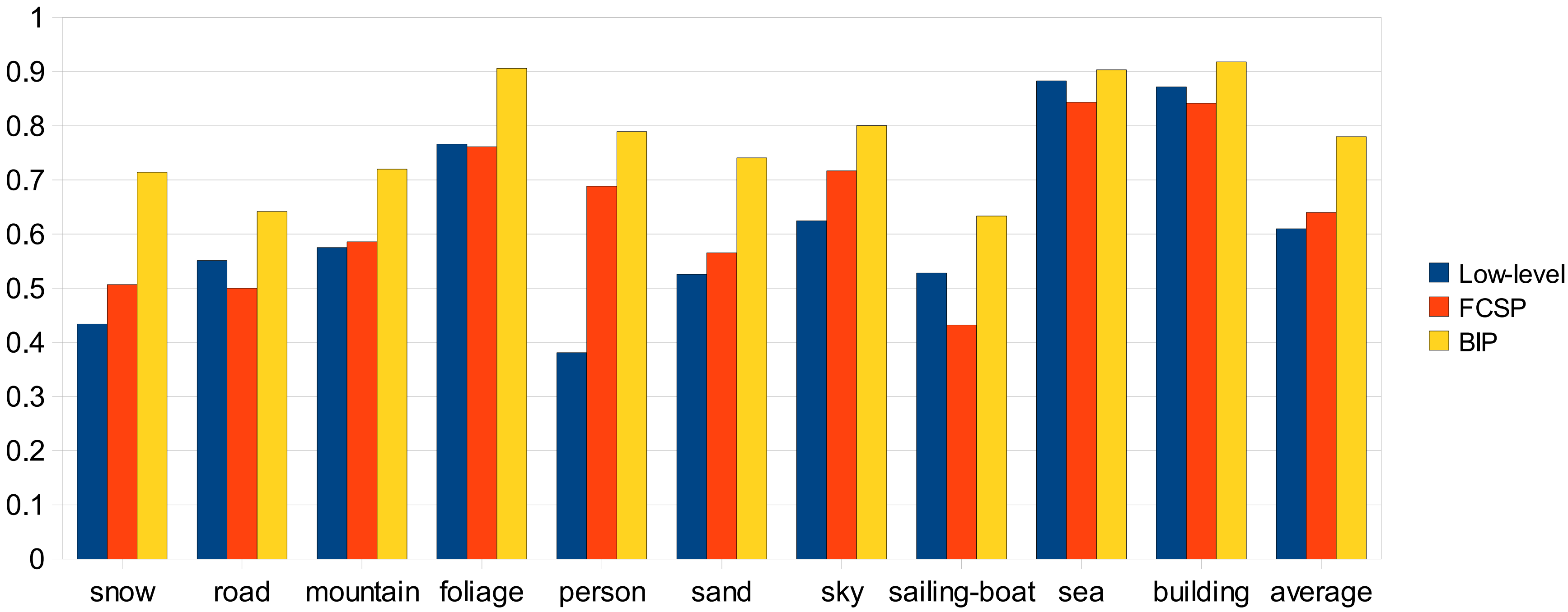


Classification Rate over Training Set Size



F-Measure per Concept

Training Set with 300 images (best performing one for BIP)



Conclusions

- Combination of
  - Wavelets based region classification
  - Spatial context with explicit knowledge
- BIP outperforms FCSP
  - Classification rate
  - Efficiency (BIP avg ~1 sec, FCSP ~40secs)
- Validation of earlier observation
  - Explicit knowledge provides good model for spatial context exploitation
  - Low amount of training data required to acquire good performing models

Thanks!