Computer Vision II

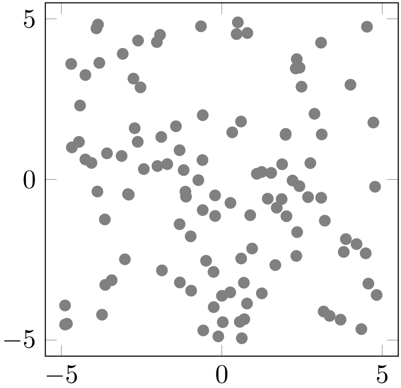
Bjoern Andres

Machine Learning for Computer Vision
TU Dresden

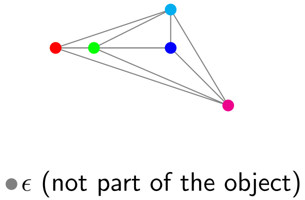
Object recognition

Object recognition is the task of finding any occurrences of an object in an image, given a **model** of the the geometry and appearance of the object.

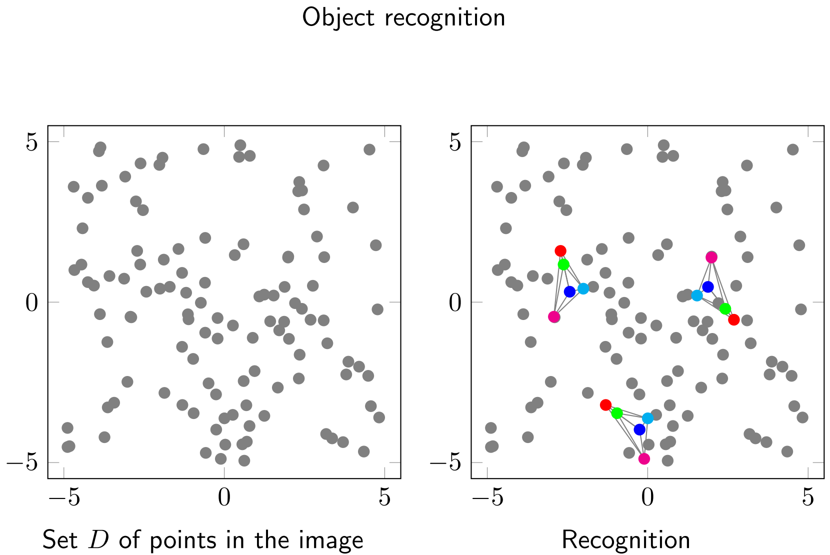
Object recognition



Set D of points in the image



Set V of object key points


Object recognition
5
0
−5
−5
0
5
Set D of points in the image
Recognition

Decisions at points

- For any point $d \in D$ in the image and any key point $v \in V$ of the object, let $y_{dv} \in \{0, 1\}$ indicate whether the point d is an occurrence of the key point v in the image
- We constrain each point in the image to be an occurrence of precisely one key point, possibly ϵ. Hence, we consider the feasible set

$$Y_{DV} = \left\{ y \colon D \times V \to \{0,1\} \;\middle|\; \forall d \in D \colon \; \sum_{v \in V} y_{dv} = 1 \right\} \; .$$

Costs at points

- For any point $d \in D$ and any key point $v \in V$, let $c_{dv} \in \mathbb{R}$ a cost associated with the decision $y_{dv} = 1$
- This cost typically depends on the contents of the image at the point d.

Decisions for pairs of points

- For any pair $\{d, d'\} \in \binom{D}{2}$ of points, let $x_{\{d,d'\}} \in \{0,1\}$ indicate whether d and d' belong to the same occurrence of an object in the image
- We require these decisions to be transitive, i.e.

$$\forall d \in D \; \forall d' \in D \setminus \{d\} \; \forall d'' \in D \setminus \{d, d'\} \colon \quad x_{\{d,d'\}} + x_{\{d',d''\}} - 1 \leq x_{\{d,d''\}} \tag{1}$$

Hence, we consider the feasible set

$$X_D = \left\{ x \colon \tbinom{D}{2} \to \{0,1\} \;\middle|\; (1) \right\}$$

Costs for pairs of points

- For any pair $(d, d') \in D^2$ of points such that $d \neq d'$ and any pair $(v, w) \in V^2$ of key points, let
  - $c'_{dd'vw} \in \mathbb{R}$ a cost associated with the decision $y_{dv}\, y_{d'w} x_{\{d,d'\}} = 1$
  - $c''_{dd'vw} \in \mathbb{R}$ a cost associated with the decision $y_{dv}\, y_{d'w}(1 - x_{\{d,d'\}}) = 1$
- These costs can depend, e.g., on the distance between d and d' in the image plane.

Optimization problem

- The task of object recognition can now be stated as the optimization problem

$$\min_{(x,y)\in X_D\times Y_{DV}} \sum_{d\in D}\sum_{v\in V} c_{dv}\, y_{dv} + \sum_{d\in D}\sum_{d'\in D\setminus\{d\}}\sum_{(v,w)\in V^2} c'_{dd'vw}\, y_{dv}\, y_{d'w}\, x_{\{d,d'\}} + \sum_{d\in D}\sum_{d'\in D\setminus\{d\}}\sum_{(v,w)\in V^2} c''_{dd'vw}\, y_{dv}\, y_{d'w}(1-x_{\{d,d'\}})$$

- This is a joint graph decomposition and node labeling problem
- The local search algorithm we have considered before (for the task of joint image decomposition and pixel labeling) can be applied!