

Spatial Priors for Part-Based Recognition using Statistical Models

D. Crandall
Cornell University

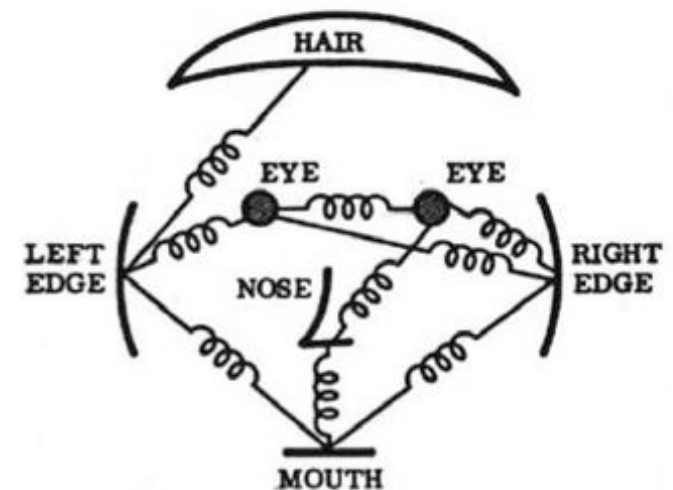
P. Felzenszwalb
University of Chicago

D. Huttenlocher
Cornell University

Part-based models

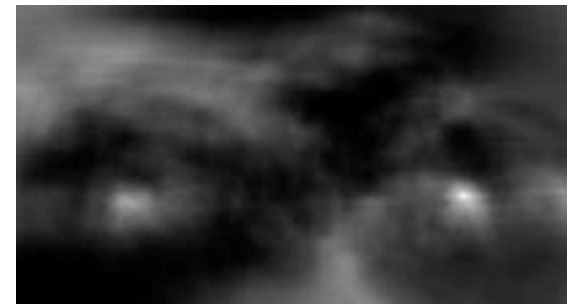
- Combination of appearance-based and geometrical models
 - Each part represents local visual properties
 - Spatial configuration captured by statistical model or spring-like connections
- Pictorial structures, Constellation of parts

History goes back to
Fischler and Elschlager, 1973



Main points

- Recognition without feature detection
 - Single overall inference problem
 - Parts have a match quality at each location



- Family of geometrical models
 - Represent using undirected graphical models
 - Choice has strong computational consequences

Statistical model

- Object with n parts labeled 1 through n
- Object configuration given by: $L = (l_1, \dots, l_n)$
 - Location of each part
- **Geometrical model:** $P(L)$
 - Undirected graphical model captures conditional independence of part locations
- **Appearance model:** $P(I|L) \propto \prod g_i(I, l_i)$
 - Factors into a term for each part

Localization problem

- Find the most probable configuration of the object,

$$L^* = \arg \max_L P(L|I)$$

- Posterior factors,

$$P(L|I) \propto P(I|L)P(L)$$

Appearance

Geometry

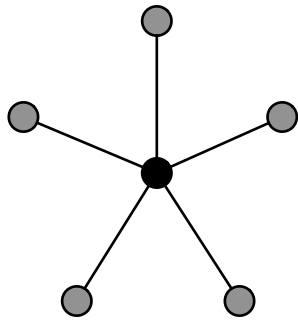
- Size of configuration space is exponential
 - n parts, m locations - $O(m^n)$ configurations
 - Use implicit search techniques

Choosing $P(L)$

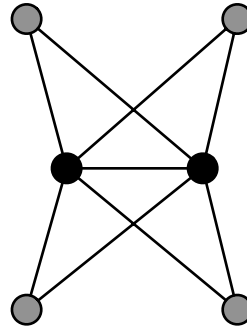
- Determines whether possible to do implicit search
- If no conditional independence assumptions are made the problem seems to be hard (joint Gaussian)
- If $P(L)$ is a tree model the problem can be solved in $O(nm)$ time
 - Use dynamic programming (DP) and distance transforms (DT) [FH00]
- What can we do between these two extremes?

k-fans graphical structure

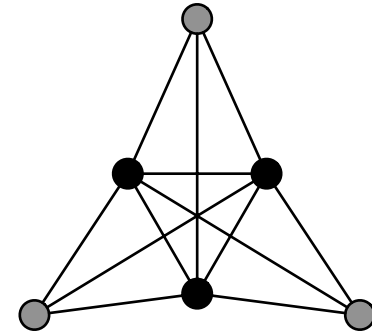
Family between a star graph and the complete graph



1-fan



2-fan



3-fan

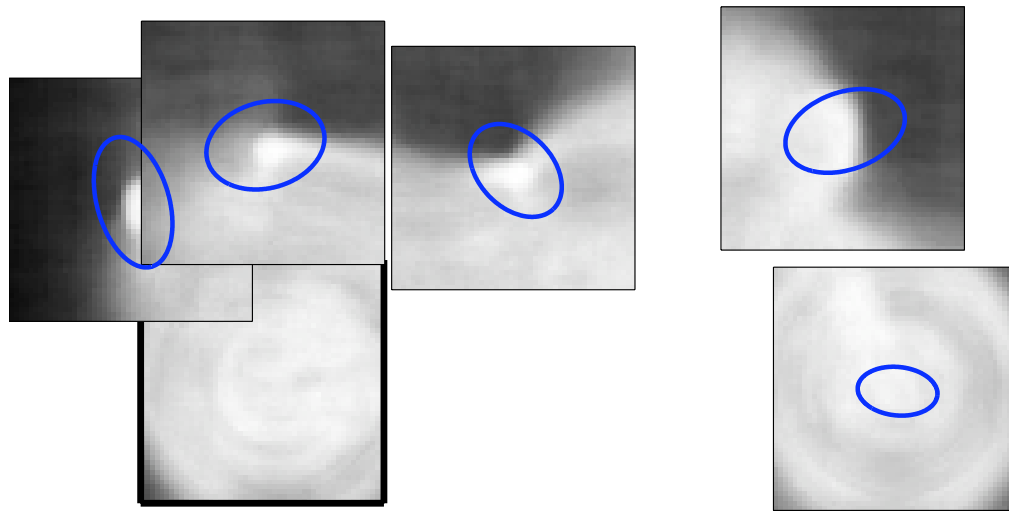
- k reference parts, n-k non-reference parts
- Non-reference parts are independent when reference locations are fixed
- Localization in $O(nm^k)$ time using DP and DT

Geometric interpretation

- Each maximal clique in a k -fan defines a soft constraint on the location of $k+1$ parts
- Can use location of reference parts to define a basis for other part locations
 - 1-fan allows for translation invariant models
 - 2-fan allows for models invariant to rigid motion and scale change
- Reference locations define a soft alignment

Motorbike model

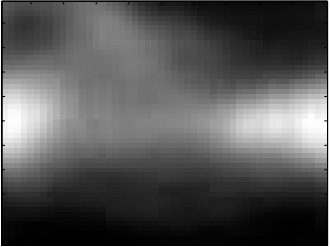
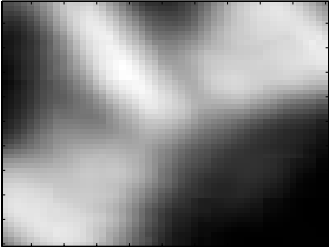
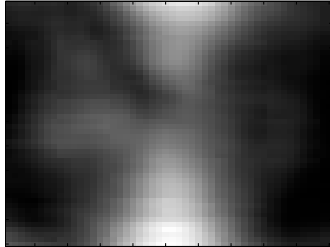
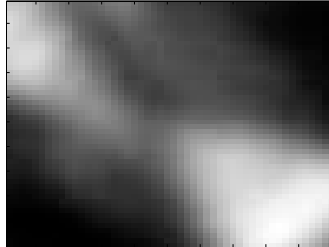
Part appearance defined by probability of an edge



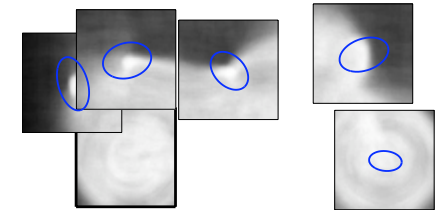
Ellipses show spatial uncertainty of non-reference parts

Reference

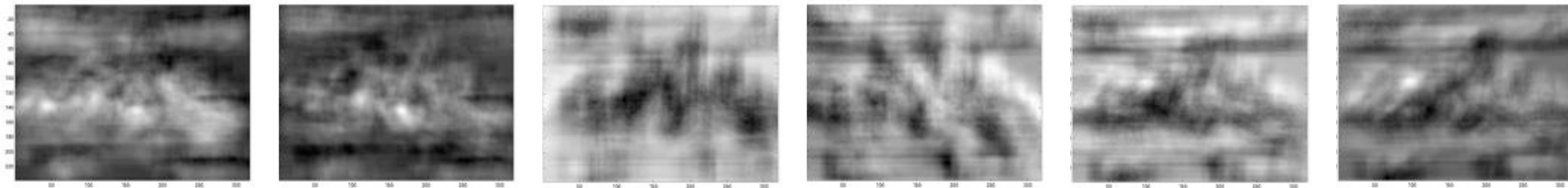
Front wheel oriented edge appearance model

	/	—	/
			

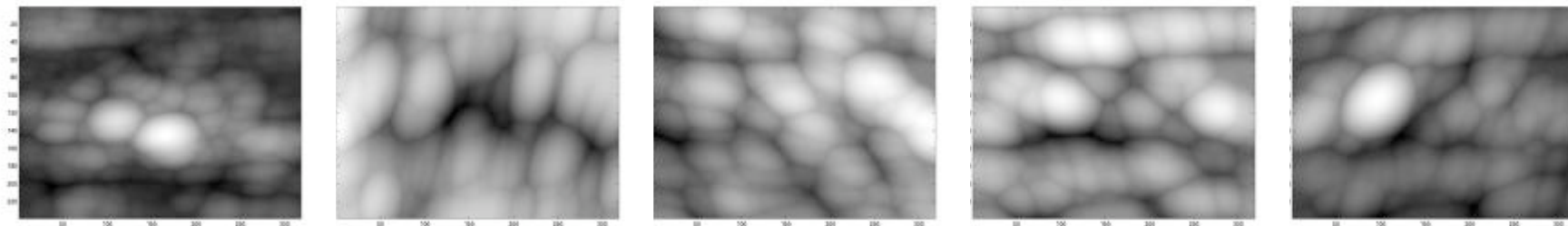
Processing steps



Match costs for each part - log probabilities

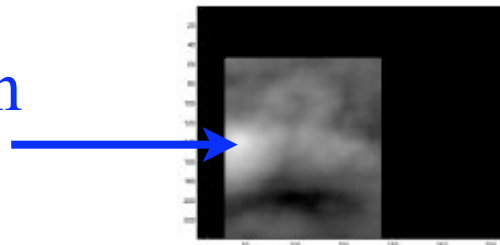


DT of non-reference match costs - correct form of blurring

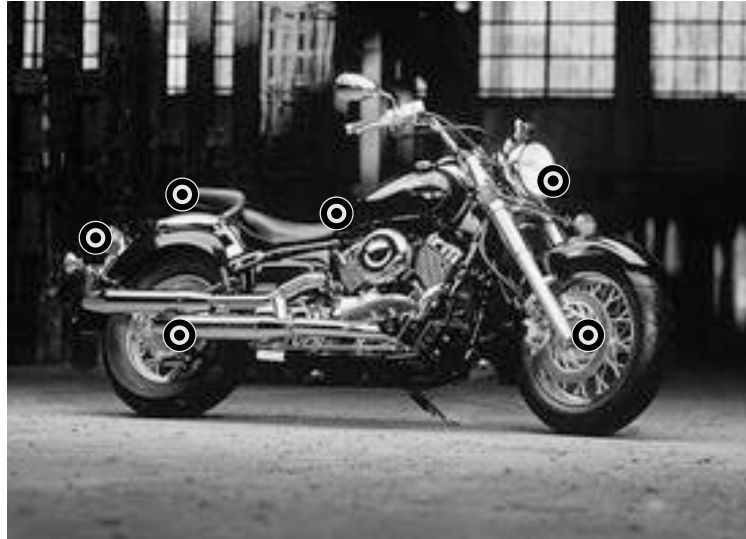


Sum shifted DTs and reference match cost

Find best location
for reference

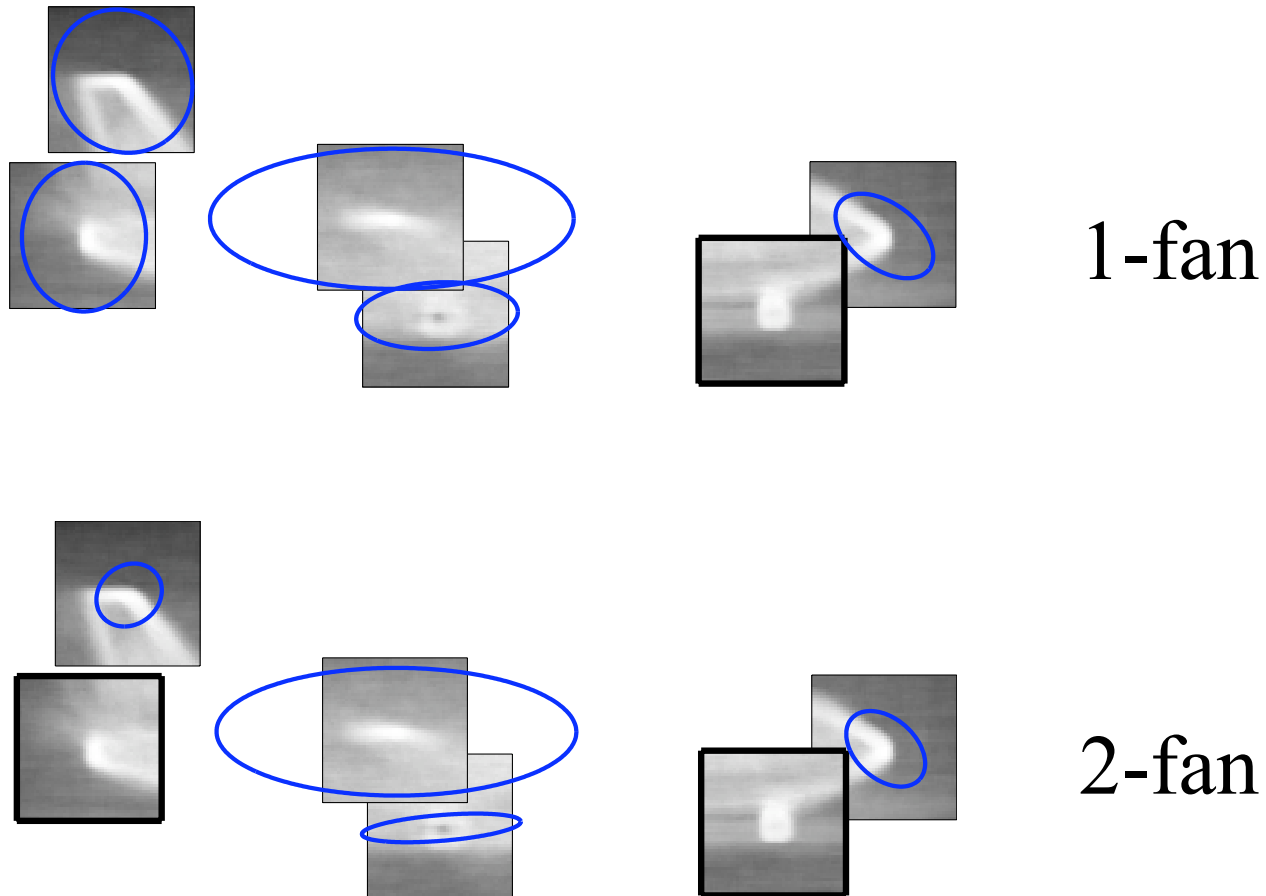


Localization with 1-fans

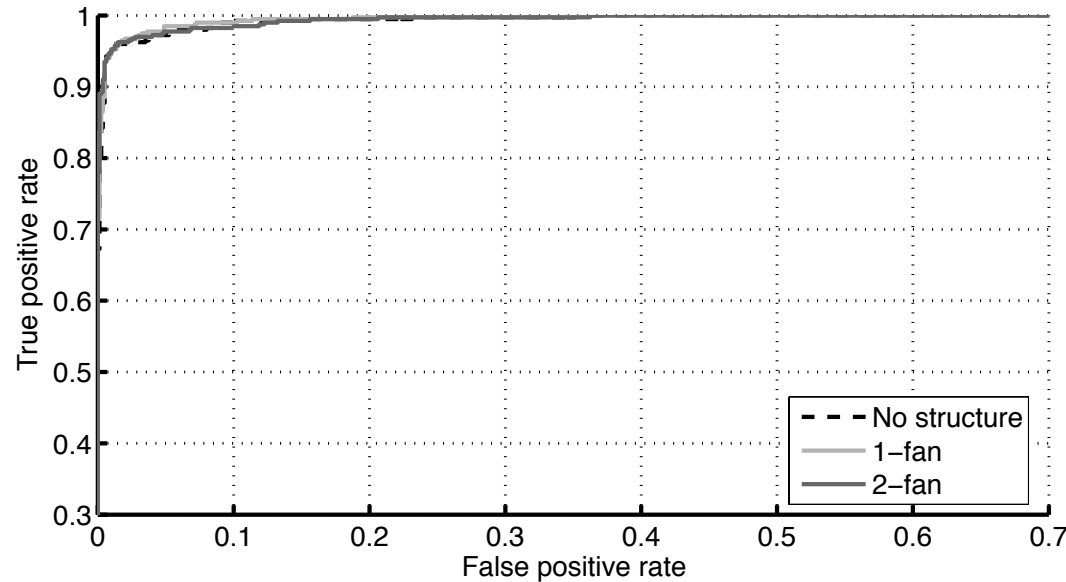


Airplane model

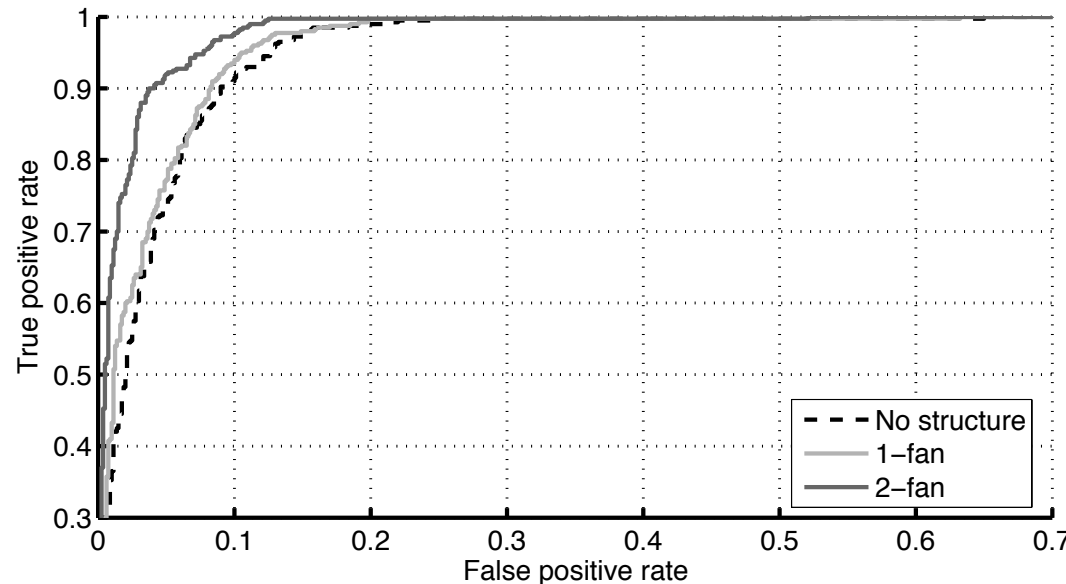
As k increases, the geometrical model becomes more precise



ROC curves



For motorbikes
1-fan is as good
as 2-fan



For airplanes
2-fan is better
than 1-fan

Comparable to other methods on the Caltech dataset

Discussion

- Detection times:
 - 0.1 sec for 1-fan vs. 3.3 sec for 2-fan
- Small amount of geometry can buy a lot
 - Appropriate amount depends on object class
 - Trade-off between model structure and computational complexity
- Recognition without feature detection combines bottom-up and top-down constraints
 - Each part is detected in the context of the others