A Critique of Local Invariant Features for Object Recognition

Daniel Huttenlocher
IMA Workshop
May, 2006

The Return of Local Features

- Long history of using sparse local features and geometric constraints for recognition
  - Roberts in 1960's
  - LFF, Interpretation Tree , Alignment, … in 1980's
- 1990's saw more global approaches
  - Appearance methods such as Turk&Pentland, Murase&Nayar, …
  - Geometric invariants, Rothwell et al, …
  - Hausdorff matching
- 2000's have seen a return to local features
  - Now often using invariant descriptors

Historical Context

- Roberts' work based on corners and edges
  - Motivated by ease of humans recognizing line drawings, extracting less variable information
- In 1980's Marr generalized such feature-based approaches
  - Primal sketch, 2½D sketch similarly based on detecting intermediate structures
  - Neurophysiologic and psychophysical evidence
    - But not clearly support for detection over filtering
- A feature dominated world ever since …

Good and Bad of Local Features

- Good
  - Less sensitive to clutter and occlusion than global measures
  - Can measure both appearance and geometric information
- Bad
  - Requires error-prone detection decisions
  - Often very sparse description
  - Difficult to combine into global model
    - Combinatorial explosion for correspondence
    - Bag models can over-count evidence

Good and Bad of Invariant Features

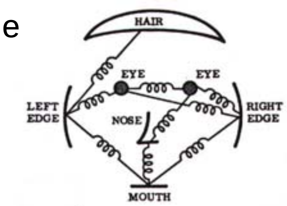
- Good
  - More reliably detectable under wider range of image conditions
- Bad
  - Feature geometry should be consistent
    - E.g., orientation and scale should match across features of a given object
  - Larger areas of support for more transformation parameters (e.g., affine)
    - Possibly more sensitive to clutter and occlusion, sparser, or overlapping

Are Features Actually Helping?

- Filtering (operators or transforms)
  - Map from images to "images"
    - Not necessarily in same coordinate system
- Feature detection
  - Map from images to sets of discrete locations
    - Again not necessarily in same coordinates
- Filtering enhances what is important without making decisions
  - E.g., likelihood function or cost map rather than set of locations

Pictorial Structures

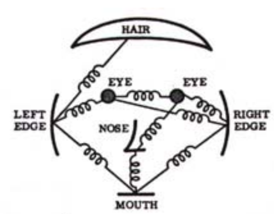
- Fischler and Elschlager took fundamentally different view in 1970's; became a sideline
  - Feature operators or cost maps rather than detection
    - Related Chamfer matching school of thought, e.g. Barrow&Tenenbaum, 1977
  - Combine feature maps in single overall optimization problem
    - Not computationally tractable at the time
    - Still challenging, but so is feature matching



All positions of all parts

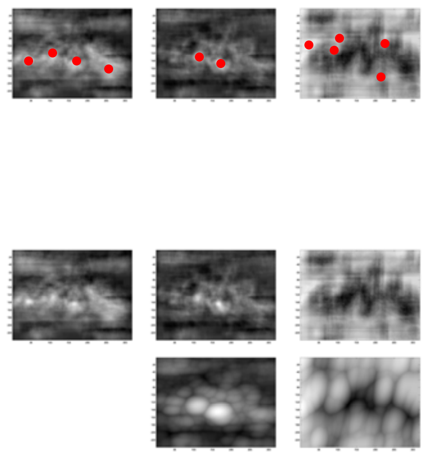
Single Optimization Problem

- Degree to which each "location" is like a given part or feature – no detection
  - Can express in Bayesian framework as likelihood of image given parts at particular locations
- Degree to which particular part locations fit the spatial configuration
  - Prior spatial model
- No error-prone local decisions about features



Contrasting the Approaches

- Feature based
  - Local feature detection
  - Explicitly handle missing data and outliers



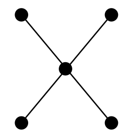
- Single optimization
  - Determine feature responses (likelihood)
  - Dynamic programming (e.g., distance transform) techniques to combine with spatial model (prior)

Computational Issues

- Feature detection analogous to computing likelihood function and then thresholding
- In principle feature detection can focus attention and reduce computation
  - In practice combinatorial problem
    - Often exponential, subsets for missing features
    - Limitation to objects with small number of parts
- While exhaustive nature of single optimization appears prohibitive
  - Dynamic programming and branch-and-bound, huge literature on optimization

Better Living Without Features

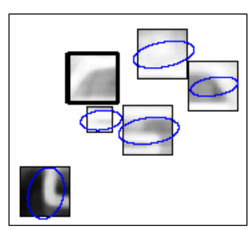
- Single optimization can be more accurate and faster than feature detection
  - Optimization approach for star model vs. feature detection for full joint Gaussian [CFH05,FPZ05]
  - 6 parts under translation, Caltech-4
  - Single class, fixed scale, equal ROC error



	Airplane	Motorbike	Faces	Cars
Features [FPZ03]	90.2%	92.5%	96.4%	90.3%
Optim. [FPZ05]	93.6%	97.3%	90.3%	87.7%
Optim. [CFH05]	93.3%	97.0%	98.2%	92.2%

Learning Without Features

- Weakly supervised learning – just specifying positive vs. negative exemplars
  - [FPZ05] used feature detector for weakly supervised learning
  - [CHF05] required extensive supervision, specifying location of each "part"
- At ECCV, weakly supervised learning without feature detection [CH06]
  - E.g., 6 part car model
  - Edge strengths



Detection Results

- Weak supervision often beats strong, no features beats features
- More parts/feature operators ("denser model") is better
  - Still not as good as bag of feature models

	Airplane	Motorbike	Faces	Cars
Features, 6 [FPZ05]	93.6%	97.3%	90.3%	87.7%
Strong, 6 [CFH05]	93.3%	97.0%	98.2%	92.2%
Weak, 6 [CH06]	93.6%	98.1%	95.6%	92.6%
Weak, 25 [CH06]	95.2%	99.0%	98.0%	95.0%

Discussion

- Doing away with features
  - Higher dimensional transformation spaces still pose challenge to feature map approaches
    - Better branch and bound search, as in Hausdorff matching under affine transformation
- Combining multiple parts/features
  - Bag models do better than anything else
  - Why?
    - Mixtures
    - Simple datasets where spatial relations not important (single feature gets 75-85% correct)