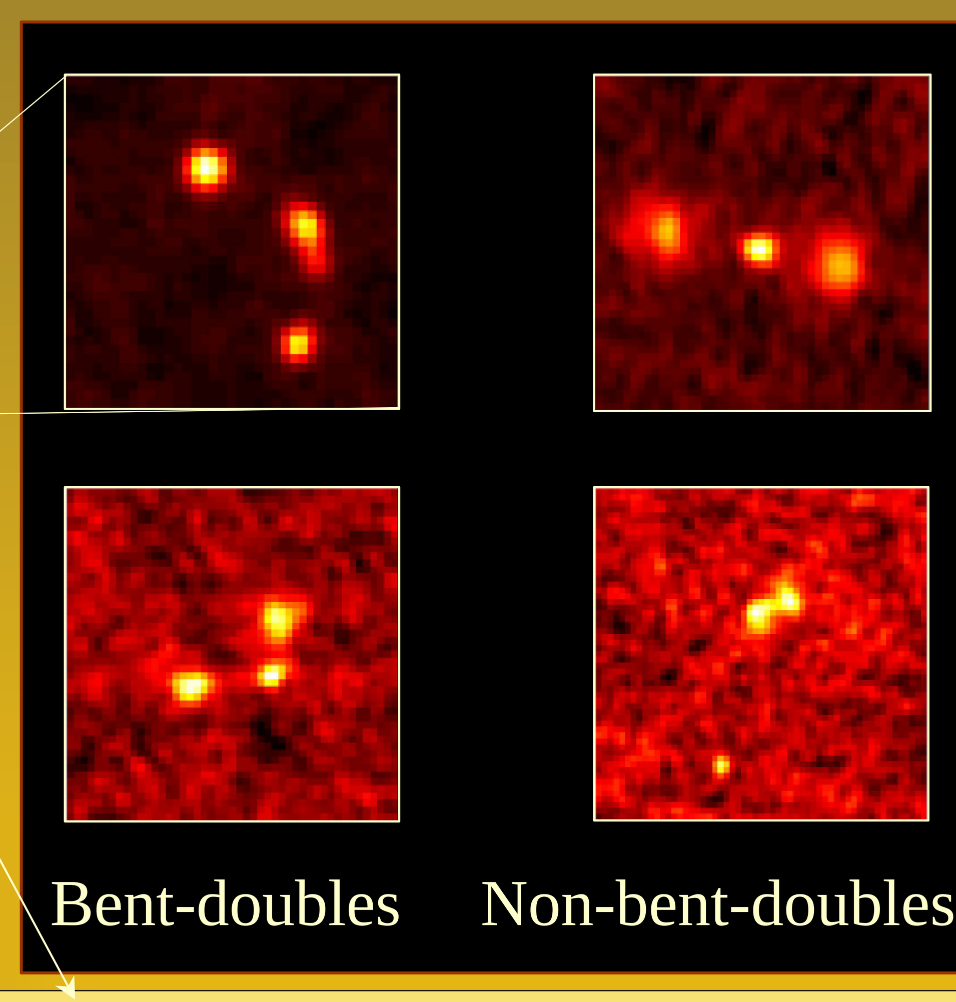
Probabilistic Model-Based Detection of Bent-Double Radio Galaxies

Sergey Kirshner[1], Igor V. Cadez[1], Padhraic Smyth[1], Chandrika Kamath[2], Erick Cantú-Paz[2]

[1]Department of Information and Computer Science, University of California, Irvine, CA 92697-3425, U.S.A.

[2]Center for Applied Scientific Computing, Lawrence Livermore National Laboratory, Livermore, CA 94551, U.S.A

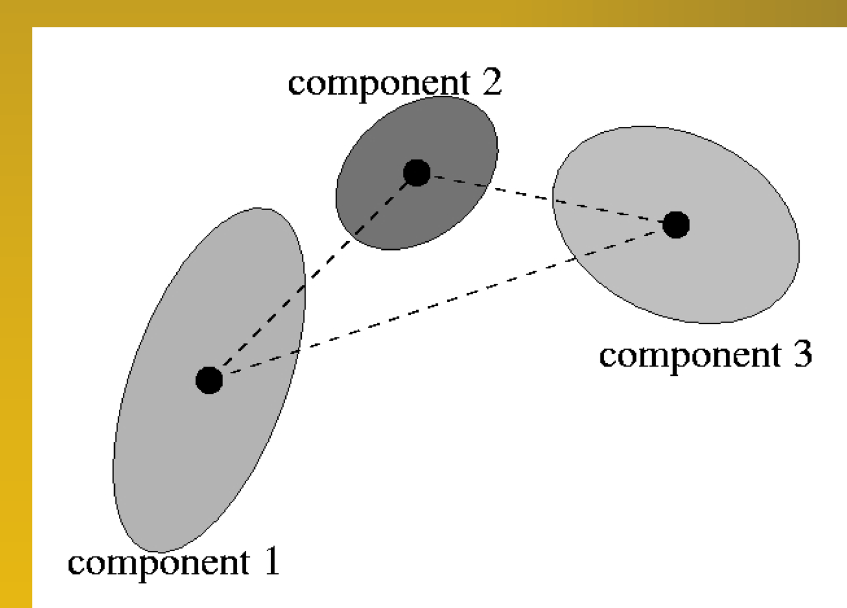
We describe an application of probabilistic modeling to the problem of recognizing radio galaxies with a bent-double morphology. The data is available in both raw images and a catalog of fitted elliptical components. Due to the size of the catalog (~770K sources), it is infeasible to manually label all of bent-double configurations.

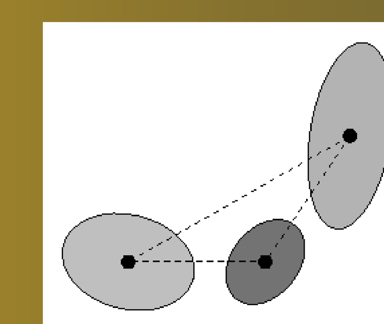


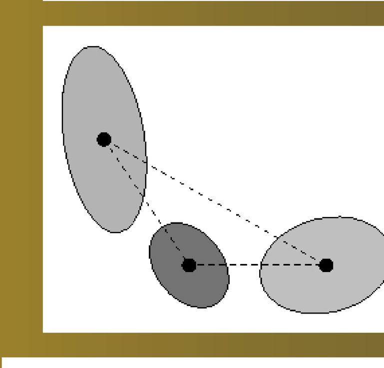
The VLA FIRST survey (http://sundog.stsci.edu) covered about 8500 squared degrees and identified about 7.7*10^5 sources.

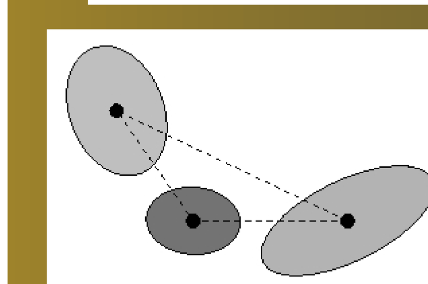
RA	Dec	Peak Flux (mJy/bm))	Major Axis (arcsec)	Minor Axis (arcsec)	Position Angle (degrees)
16 39 31.273	+28 54 31.83	6.35	11.03	6.00	20.5
16 39 31.399	+28 54 9.36	6.76	6.25	5.48	153.3
16 39 32.800	+28 54 44.58	8.97	6.83	6.72	0.1

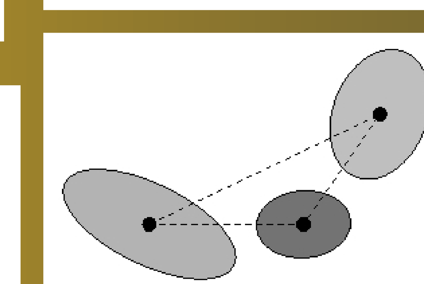
Simplified Representation of Bent-Double Configurations and Possible Orientations

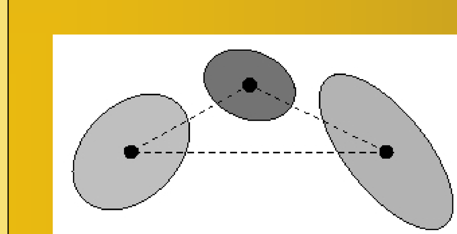


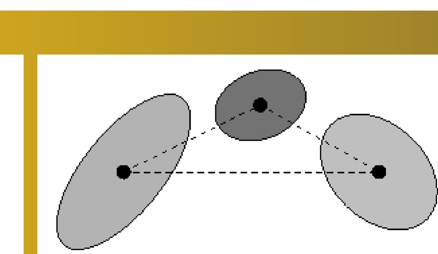












How to Build a Probabilistic Model?

- Build a probabilistic model only for bent-doubles.
- Bent-doubles exhibit reflection symmetry.
  - Call symmetrical elliptical components "lobes"
  - Call the center components "cores"
- Given orientation, can calculate features and use them.
  - Center angles
  - Side ratios
  - Peak intensity ratios
  - 3 per configuration per orientation
- Once features calculated, can calculate the probability of the configuration being a bent-double under a particular orientation.
- Assuming conditional independence of features: $P((\mathbf{c}_a,\mathbf{c}_b,\mathbf{c}_c)|\theta)=P(\alpha_{a,\theta},\alpha_{b,\theta},\alpha_{c,\theta})P(sr_{a,\theta},sr_{b,\theta},sr_{c,\theta})P(ir_{a,\theta},ir_{b,\theta},ir_{c,\theta})$.
- Knowing cores for all configurations in the set, we can further express probabilities of individual features.
  - Kernel density estimators (KDE)
  - Normal distributions on transformed features

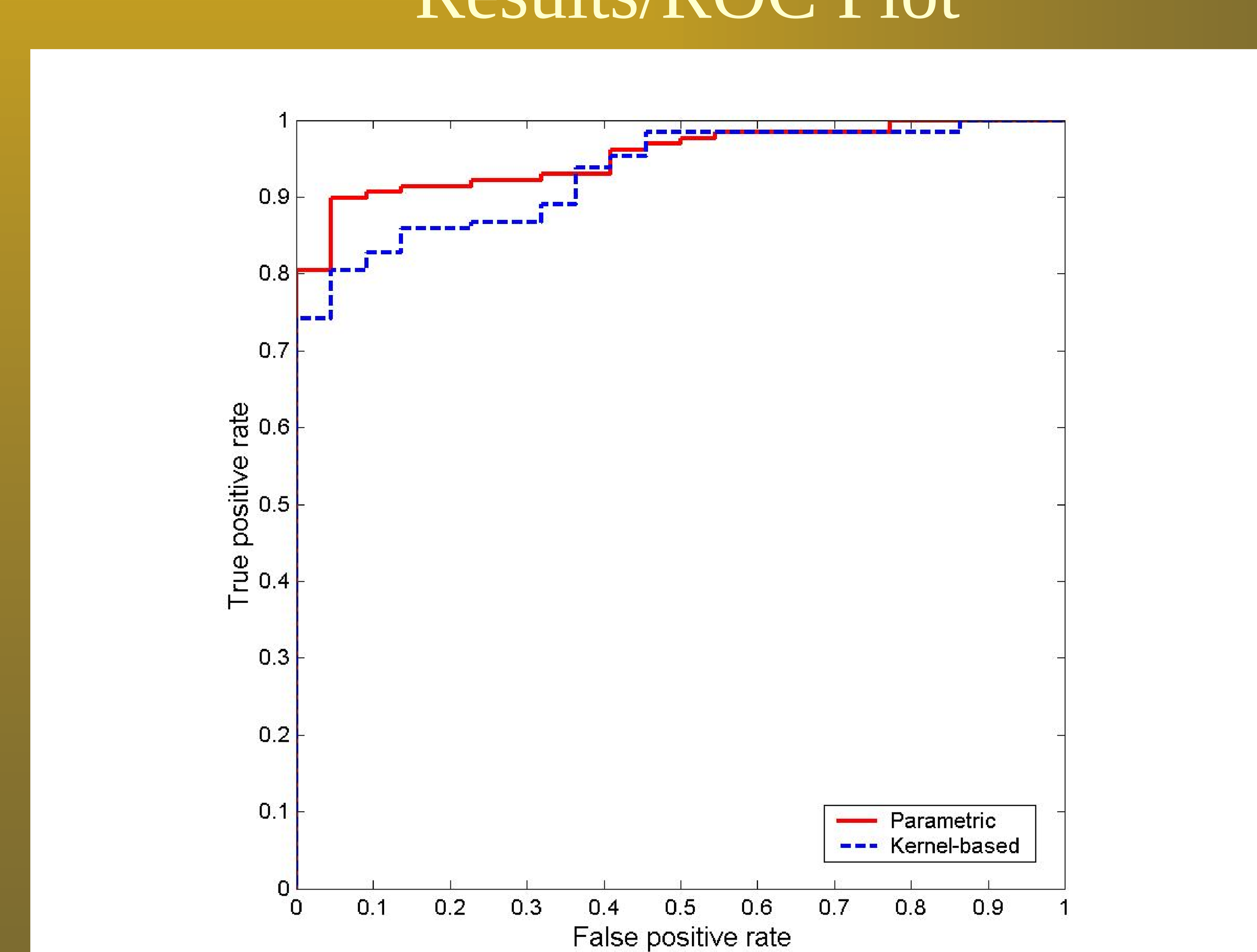
How to Find Proper Orientations?

- What is left is to find the cores for the training set.
- We can ask the scientists to identify the cores.
- Alternatively, we can use an iterative algorithm to find cores automatically.
  - Initially, core components for each examples in the training set are chosen at random.
  - At each iteration, a core for each configuration is re-selected to maximize the probability of the features for the configuration.
  - Iterations are stopped once there are no changes from one iteration to the next or after a fixed number of iterations has elapsed.
- The procedure does not guarantee convergence to the set of cores which maximize the probability of the training set, but works well in practice.

Results and Future Directions

- We are using receiver operating characteristics (ROC) curve plot for evaluation of the method.
- For each of the configurations, we calculate its probability score given the rest of positive examples from the set.
- Scores are then arranged in increasing order, and ROC plot is built.
- Algorithm can be converted into classifier by setting a threshold for positives.
- Features other than described can be used as well.
- In the future, we can use EM algorithm to learn orientations with parametric method.
- We can also mixtures of classes with EM. (Early results for the last two bullets has been submitted to NIPS '02).

Results/ROC Plot



For parametric method, the area under the ROC curve is 0.0469. For KDE-based method, the area under the ROC curve is 0.0696. Random score assignment would yield the area of 0.5 while perfect assignment would yield 1.