Preprint
UCRL-136286

Use of Morphological Operators and Pattern Recognition Techniques for sorting the First Data

D. Proctor

This article was submitted to
Center for Advanced Signal and Image Sciences, 1999 Signal and Imaging Sciences Workshop, Livermore, CA, November 11-12, 1999

U.S. Department of Energy

Lawrence
Livermore
National
Laboratory

November 4, 1999

Approved for public release; further dissemination unlimited

DISCLAIMER

This document was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor the University of California nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or the University of California, and shall not be used for advertising or product endorsement purposes.

This is a preprint of a paper intended for publication in a journal or proceedings. Since changes may be made before publication, this preprint is made available with the understanding that it will not be cited or reproduced without the permission of the author.

This report has been reproduced
directly from the best available copy.

Available to DOE and DOE contractors from the
Office of Scientific and Technical Information
P.O. Box 62, Oak Ridge, TN 37831
Prices available from (423) 576-8401
http://apollo.osti.gov/bridge/

Available to the public from the
National Technical Information Service
U.S. Department of Commerce
5285 Port Royal Rd.,
Springfield, VA 22161
http://www.ntis.gov/

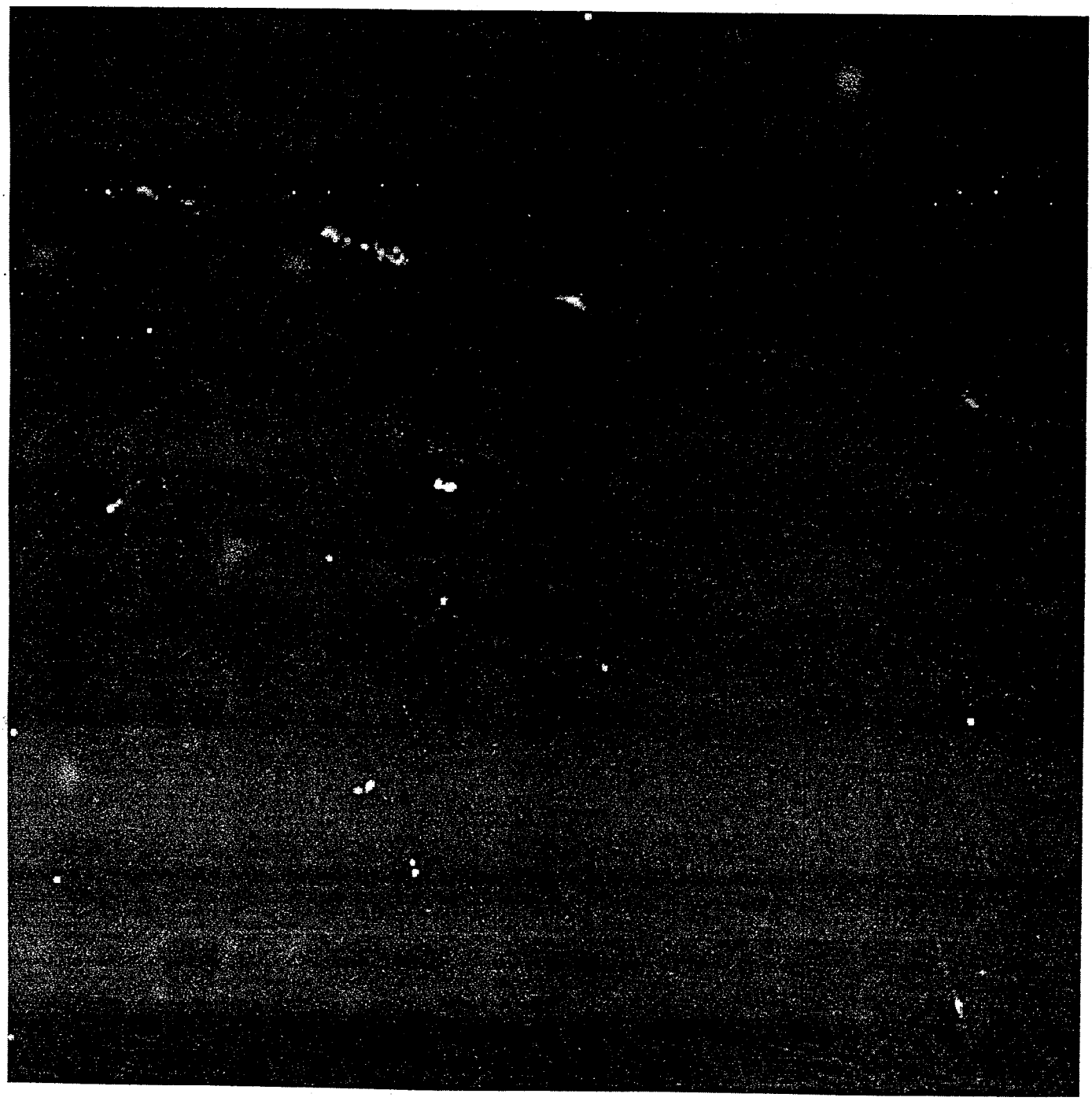
OR

Lawrence Livermore National Laboratory
Technical Information Department's Digital Library
http://www.llnl.gov/tid/Library.html

USE OF MORPHOLOGICAL OPERATORS AND PATTERN RECOGNITION TECHNIQUES FOR SORTING THE FIRST DATA

Deanne Proctor
Scientific Computing Applications, Computations
Lawrence Livermore National Laboratory

November, 1999.



Use of Morphological Operators and Pattern Recognition Techniques for Sorting the FIRST Data.

Deanne Proctor
(925) 423-4039
ddproctor@llnl.gov
L-413

The application of morphological and pattern recognition techniques to the sorting of FIRST data is presented. FIRST (Faint Images of the Radio Sky at Twenty Centimeters) is designed to produce the radio equivalent of the Palomar Observatory Sky Survey, using the National Radio Astronomical Observatory Very Large Array. A catalog of source positions and Gaussian-fit parameters for the sources is available.

The goal of this part of the project is the automatic classification of types of sources, since it is expected that there will be on the order of one million sources in the final catalog.

The data consists of radio maps, each approximately 1500x1100 pixels and containing on average about 30 sources per map. The resolution of these maps is 5" (approximately 3 pixels). The sources range in size from a few pixels to 70 or 80 pixels maximum diameter and have various morphologies. Of particular interest is the extraction of 'bent doubles', core-jets, and gravitational lens candidates. These interesting types constitute on the order of one percent of the sources.

The emphasis in this presentation is on image-based pattern recognition. We report on the use of morphological operators for source extraction and compare use of decision trees and neural networks in the classification of source morphologies and accounting for the finite resolution.

Keywords: pattern recognition, morphology

OUTLINE

I. INTRODUCTION

II. SOURCE EXTRACTION
- A. Catalog
- B. Morphological Operators
- C. Groupings

III. PATTERN RECOGNITION
- A. Invariance
- B. Training and Test Sets
- C. Feature Selection
- D. Decision Trees and Neural Networks

IV. RESULTS

RESOURCES

FIRST maps, (FIRST survey, Robert Becker)
IDL Interactive Data Language
(Research Systems, Inc.)
IDL UIT database software (Landsman)
IDL FIRST catalog in database form (Michael Gregg)
OC1 decision tree software
(Murthy, Salzberk, Kasif Johns Hopkins Univ.)
SNNS Stuttgart Neural Network Simulator
(Institute for Parallel and Distributed High Performance Systems, Univ. of Stuttgart)

Thanks to IGPP, Charles Alcock, Kem Cook for office and computing facilities.

SOURCE EXTRACTION

CATALOG

Up to 4 Gaussians fit per source island.
Software to group catalog sources by proximity (~1').
Group grows until no new source within distance.

Singles	71%
Doubles	18%
Triples	6%
Four and more	5%

MORPHOLOGICAL

Erosions, Dilations and Label Region

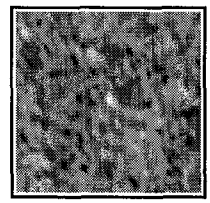
TABLE 1

SAMPLE PAGE FROM THE FIRST CATALOG

RA (2000) (1)	Dec (2000) (2)	W[a] (3)	S_p[b] (4)	S_i[c] (5)	RMS[d] (6)	Maj[e] (7)	Min[e] (8)	PA[f] (9)	Field[g] (10)
08 00 00.099	+32 06 09.83		1.11	1.07	0.138	2.8	−2.8	173.7	08000+32118B
08 00 00.250	+38 14 21.10		8.99	12.35	0.143	4.1	2.4	23.8	08000+38071B
08 00 00.453	+34 52 18.24	W	1.13	2.17	0.157	11.9	−3.3	91.9	08000+35071B
08 00 00.591	+39 50 49.00		2.03	1.31	0.148	−2.3	−3.8	163.1	08000+39393B
08 00 00.653	+37 33 50.90		22.01	21.91	0.266	0.0	−1.0	34.4	08000+37367B
08 00 00.752	+32 46 50.88		4.24	3.95	0.106	−1.0	−1.8	66.2	08000+32407B
08 00 02.306	+31 17 20.94		2.19	2.75	0.095	3.3	2.0	170.2	08000+31143B
08 00 02.515	+35 49 27.15		1.11	0.85	0.153	1.5	−3.6	133.8	08000+36065B
08 00 03.247	+36 29 42.13		46.53	58.70	0.138	3.7	1.5	21.6	08000+36364B
08 00 03.683	+33 58 24.35		1.50	1.08	0.144	−2.8	−2.8	0.3	08000+34082B
08 00 04.079	+32 02 09.58		40.37	54.07	0.155	4.3	1.8	133.9	08000+32118B
08 00 04.348	+31 23 15.24		4.16	4.98	0.115	3.6	0.0	126.8	08000+31143B
08 00 05.466	+32 11 23.16		2.95	3.17	0.133	2.6	−1.5	26.1	08000+32118B
08 00 05.859	+38 19 21.50	W	1.04	9.24	0.155	19.3	11.8	102.0	08000+38071B
08 00 06.096	+34 04 34.60		1.30	1.62	0.132	3.6	1.5	79.5	08000+34082B
08 00 06.436	+34 58 02.24		1.11	1.69	0.146	5.6	2.0	107.4	08000+35071B
08 00 06.555	+32 01 48.05		1.11	0.75	0.157	−2.0	−3.6	149.9	08000+32118B
08 00 06.574	+34 04 36.10		1.39	1.45	0.132	4.5	−3.1	98.9	08000+34082B
08 00 06.591	+39 38 05.90		1.32	2.02	0.132	6.7	−1.5	32.7	08000+39393B
08 00 07.046	+29 36 56.93		8.97	13.42	0.151	4.2	3.4	145.2	08000+29486A
08 00 07.118	+34 15 07.26		1.17	0.89	0.137	0.0	−3.5	18.1	08000+34082B
08 00 07.145	+33 40 05.92		7.69	9.06	0.129	2.7	1.8	139.7	08000+33389B
08 00 07.371	+38 07 34.51		7.20	8.05	0.138	2.3	1.3	176.0	08000+38071B
08 00 07.745	+34 06 14.68		1.28	1.07	0.133	2.1	−3.4	133.0	08000+34082B
08 00 07.872	+41 13 06.70		3.76	3.71	0.126	0.0	−1.0	0.0	08000+41132B
08 00 07.954	+36 04 08.82		4.83	5.08	0.136	1.5	0.9	62.0	08000+36065B
08 00 08.817	+38 46 23.73		1.15	0.96	0.142	1.5	−3.3	17.2	08000+38377B
08 00 09.145	+36 04 07.45		2.00	4.64	0.137	7.1	5.3	56.4	08000+36065B
08 00 09.623	+38 06 49.86		1.08	1.30	0.138	2.6	2.1	180.0	08000+38071B
08 00 09.689	+33 59 56.95		6.11	8.32	0.140	4.1	2.3	53.3	08000+34082B
08 00 10.125	+30 31 51.56		1.31	0.88	0.156	−1.5	−3.8	151.5	08000+30456B
08 00 10.262	+41 20 36.81		1.36	0.80	0.133	−3.4	−3.6	94.1	08000+41132B
08 00 10.335	+35 57 41.31		1.09	1.54	0.145	5.8	−1.5	79.8	08000+36065B
08 00 10.567	+34 02 20.85		1.59	3.45	0.137	8.9	2.8	178.8	08000+34082B
08 00 10.613	+28 51 23.53		1.88	4.94	0.095	11.8	2.4	159.4	08000+28521B
08 00 10.794	+30 36 06.29		9.51	10.03	0.148	1.5	0.9	66.0	08000+30456B
08 00 11.197	+30 31 48.01		25.66	25.99	0.156	0.8	0.2	117.0	08000+30456B
08 00 11.282	+34 00 07.72		3.23	3.77	0.140	2.7	1.6	89.7	08000+34082B
08 00 11.479	+30 14 34.11		3.13	2.72	0.136	1.0	−2.8	141.0	08000+30170A
08 00 11.495	+35 50 17.56		1.10	0.80	0.154	−2.0	−3.4	176.0	08000+36065B
08 00 12.075	+30 48 48.79		3.02	2.74	0.138	−1.5	−2.0	0.9	08000+30456B
08 00 12.089	+30 05 16.34		9.89	10.43	0.151	1.4	1.1	90.0	08000+30170A
08 00 12.833	+29 23 46.99		1.09	1.35	0.138	5.2	−2.5	67.2	08000+29203A
08 00 14.079	+41 46 12.10		3.09	3.05	0.133	1.5	−1.5	17.7	08000+41450B
08 00 14.111	+41 30 47.54		1.71	2.48	0.146	5.0	2.0	97.9	08000+41450B
08 00 15.829	+35 44 29.46		3.79	4.41	0.141	2.8	1.3	135.9	08000+35367B
08 00 16.065	+40 29 56.01		189.37	199.72	0.145	1.3	1.3	90.0	08000+40417B
08 00 16.990	+32 14 50.34		3.47	3.42	0.135	1.8	−2.0	163.3	08000+32118B
08 00 17.075	+35 42 42.29		3.88	3.93	0.139	1.0	0.0	156.5	08000+35367B
08 00 17.567	+37 02 59.81		33.55	34.94	0.138	1.2	0.9	180.0	08000+37065B

[a] Warning flag 'W' indicates source is possibly a sidelobe of a nearby, bright source.
[b] Peak flux density (mJy/beam) at 1.4 GHz for Gaussian fit to source.
[c] Integrated flux density (mJy) for Gaussian fit to source.
[d] Estimated RMS noise (mJy/beam) in coadded image at source position.
[e] Major and minor axes FWHM (arcsec) for source after 5″.4 synthesized beam size has been deconvolved. Negative values indicate that source was smaller than beam and allow fitted size to be recovered (see text.)
[f] Position angle of major axis (degrees east of north.)
[g] Name of coadded image including this source.

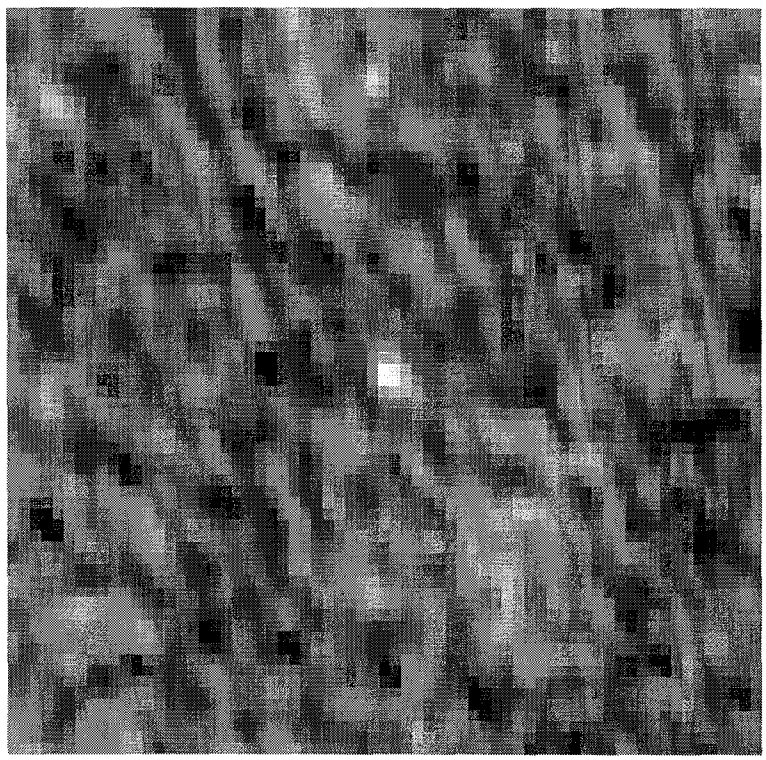
file: s98_000128.fits
dir: .



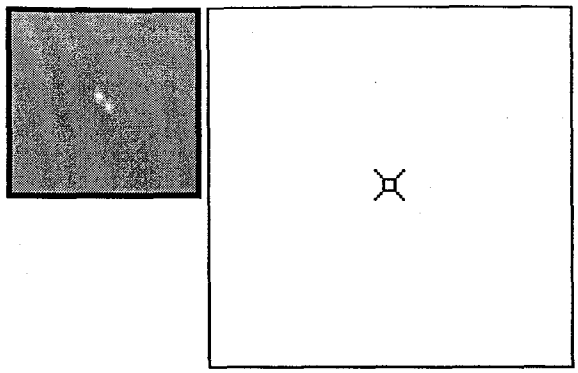
0:00:51.910 -10:16:25.55 J2000

70.8 94.8 x

single

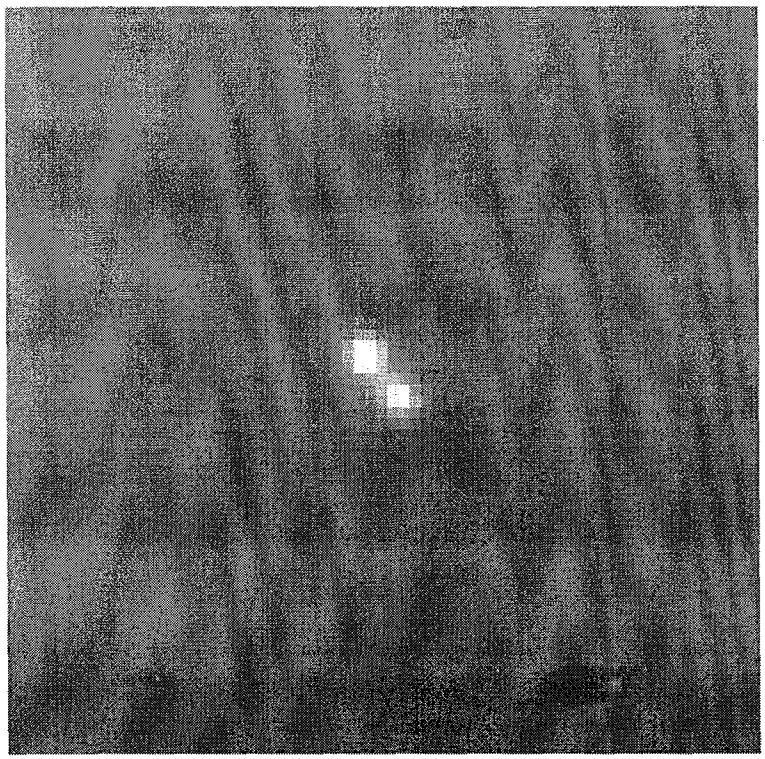


file: north98_074343.fits
dir: RAND_N

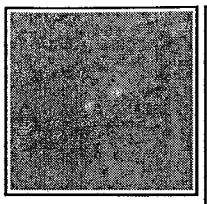


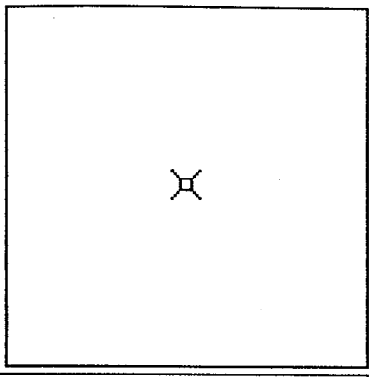
89.2 95.0 x

Double



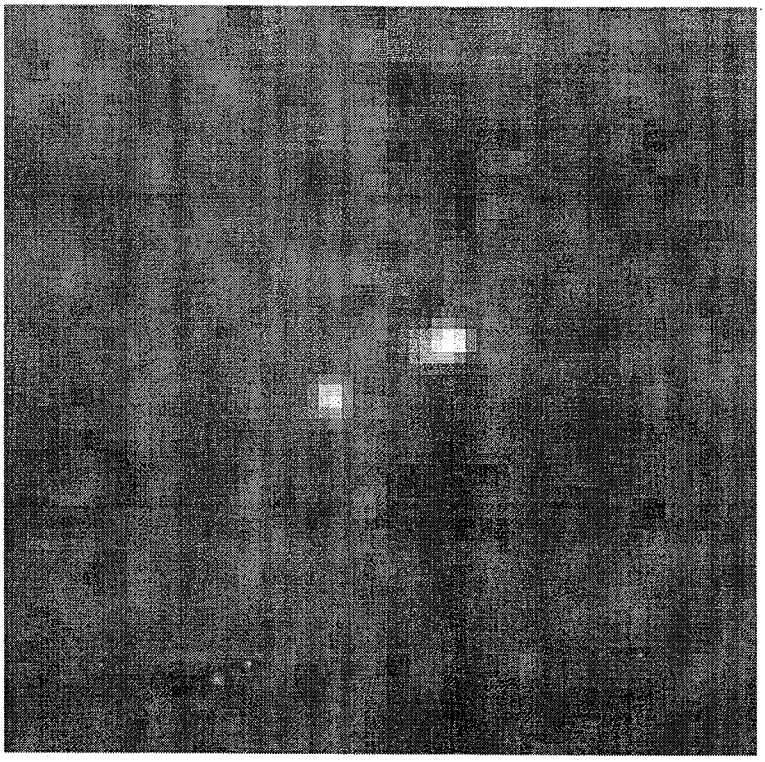
file: north98_076235.fits
dir: RAND_N



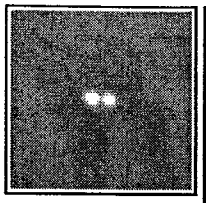


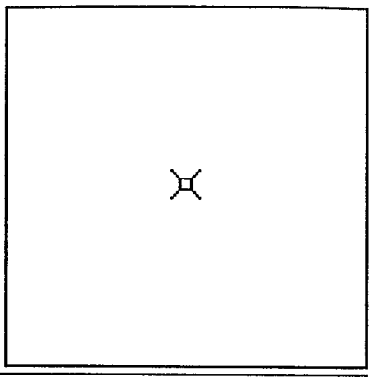
44.5 96.2 x

Double



file: north98_103126.fits
dir: RAND_N

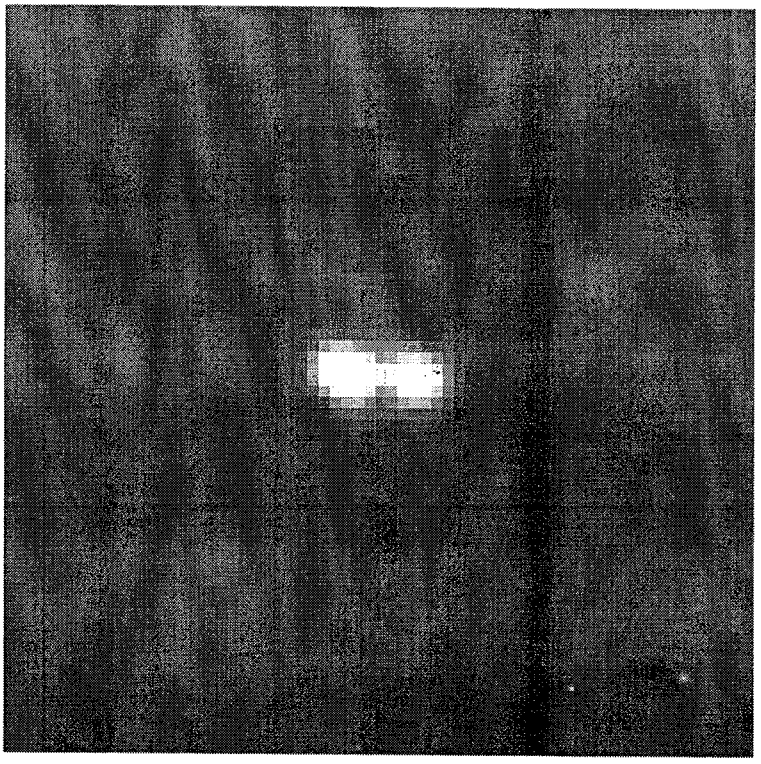




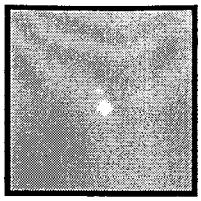
9:57:08.001 +36:18:34.04 J2000

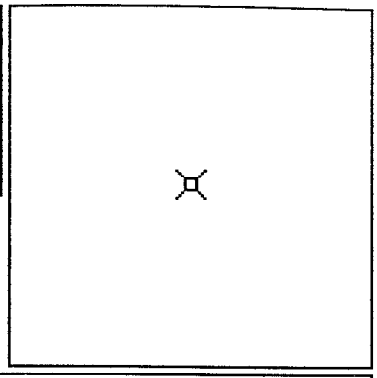
27.2 92.8 x

Double



file: north98_108485.fits
dir: RAND_N

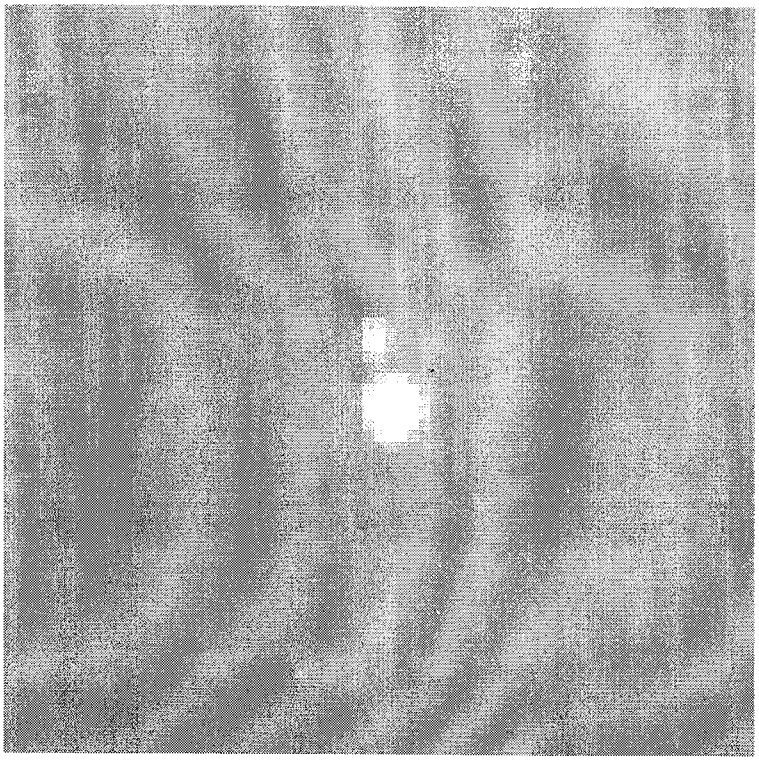




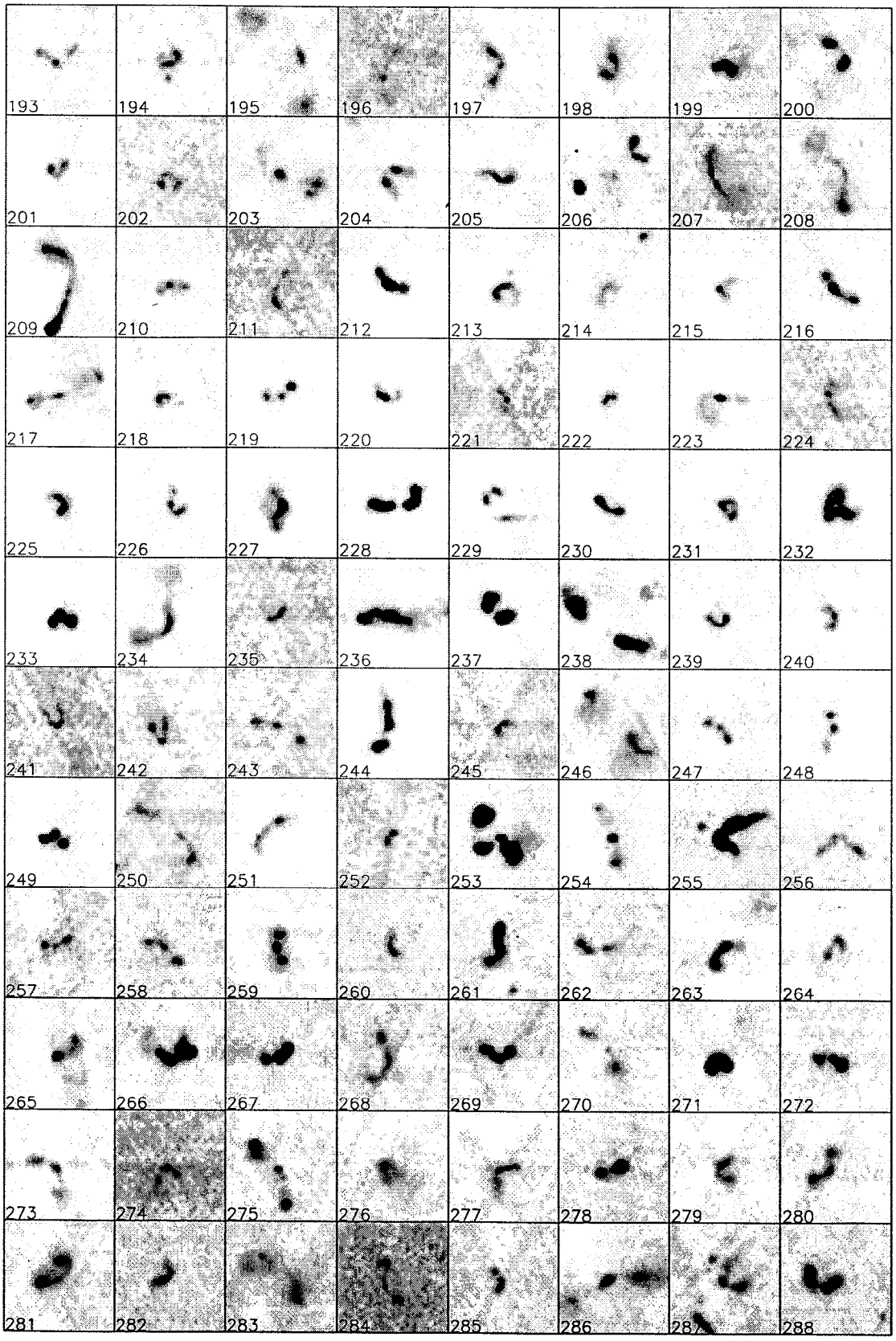
10:05:45.922 +34:36:49.63 J2000

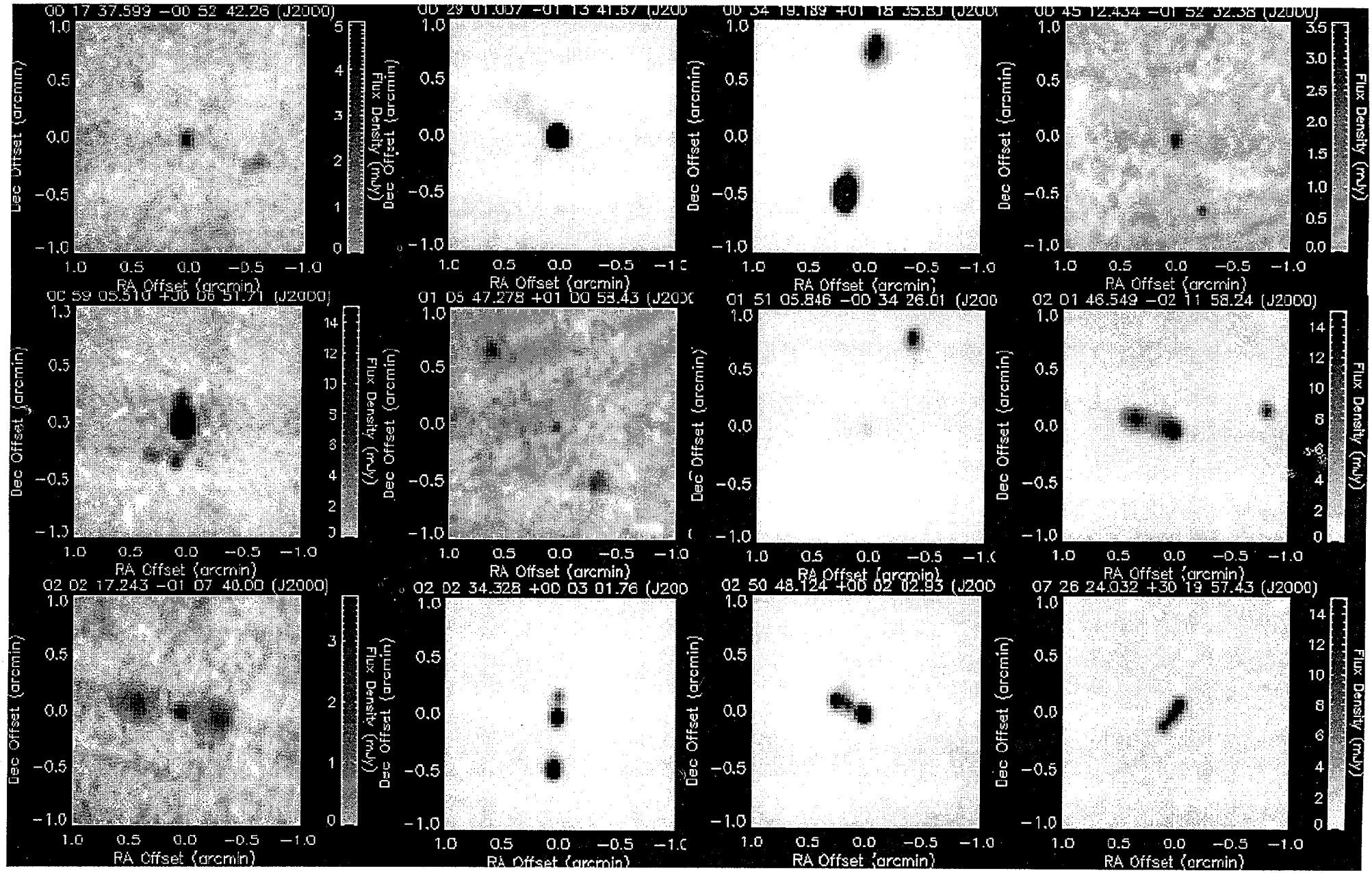
70.2 96.2 x

Double



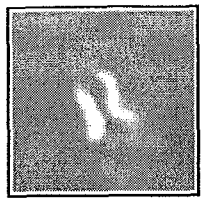
Example Bents NAT, WAT





Example Non Dent

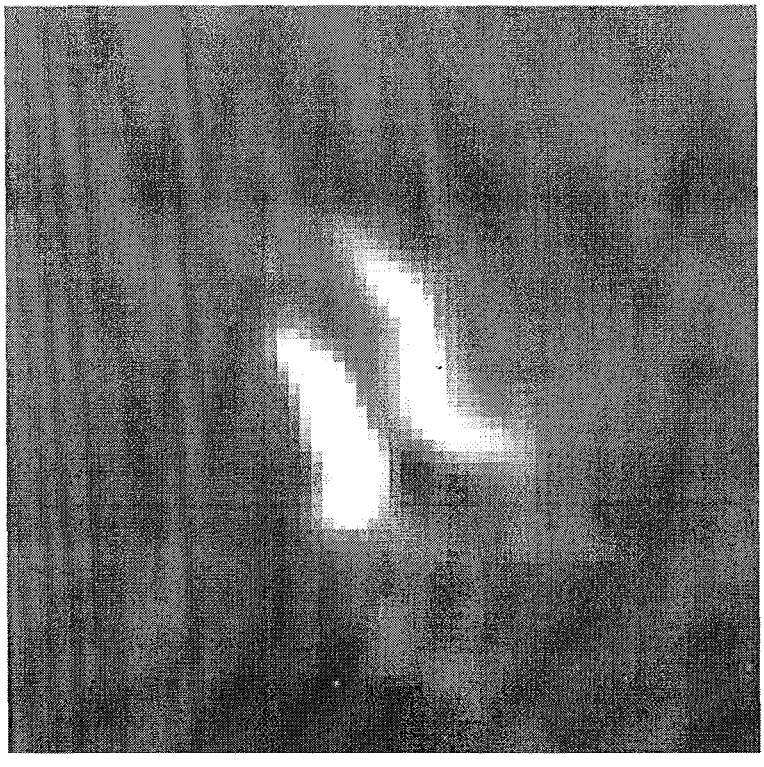
file: south98_008442_008447.fits
dir: .



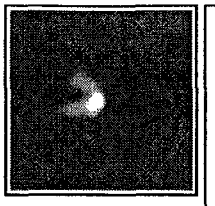
23:59:56.160 +00:01:57.45 J2000

55.5 97.5 x

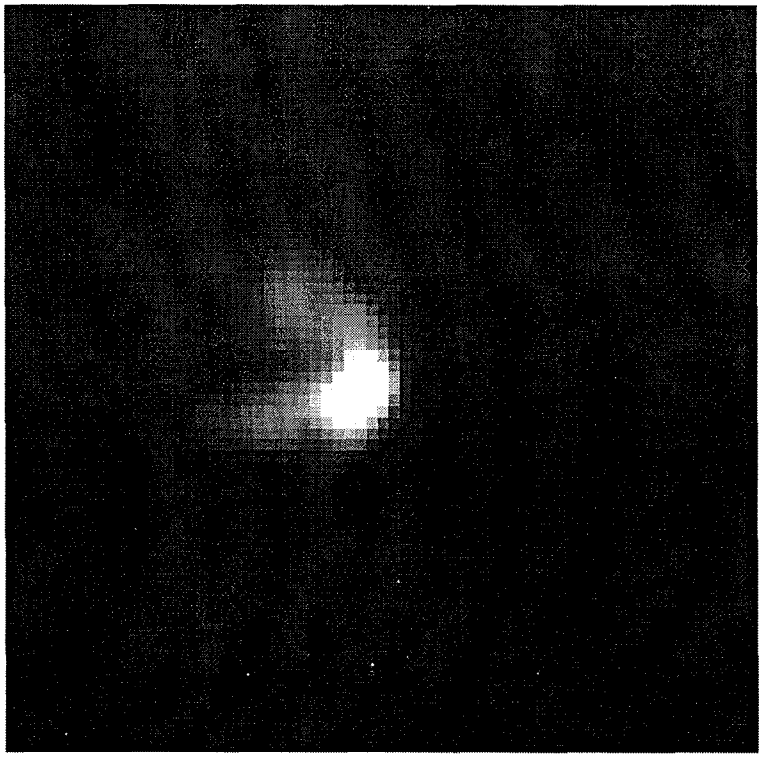
NAT



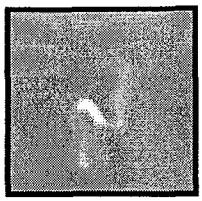
file: s98.00210.first.fits
dir: .



83.8 97.0 x

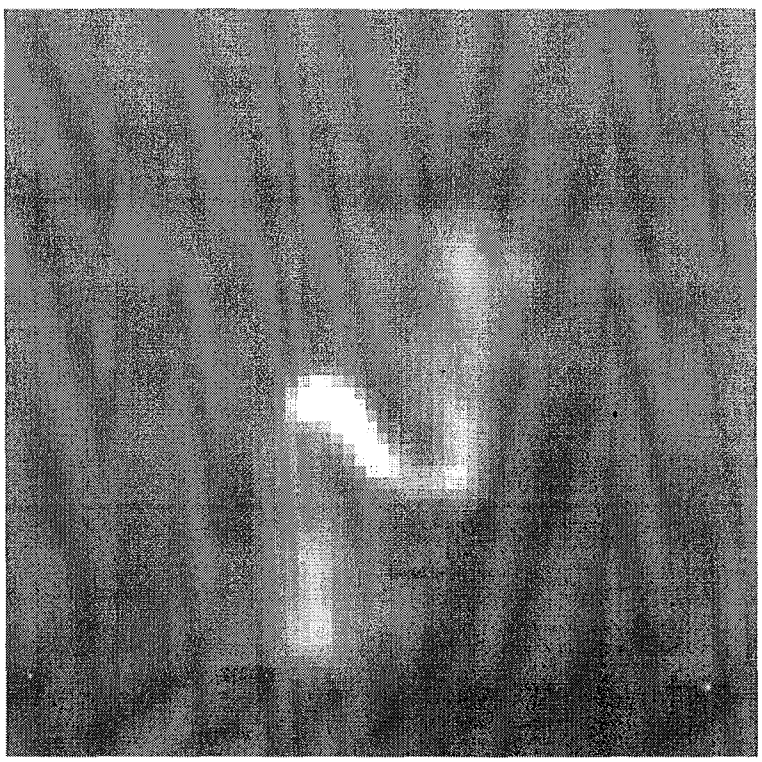


file: north98_069940_069971.fits
dir: GROUPS_4+Sources



23:59:58.680 +00:01:57.45 J2000

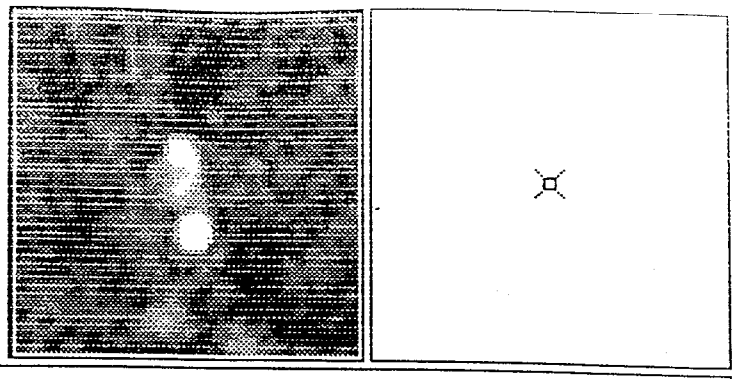
53.2 97.2 x



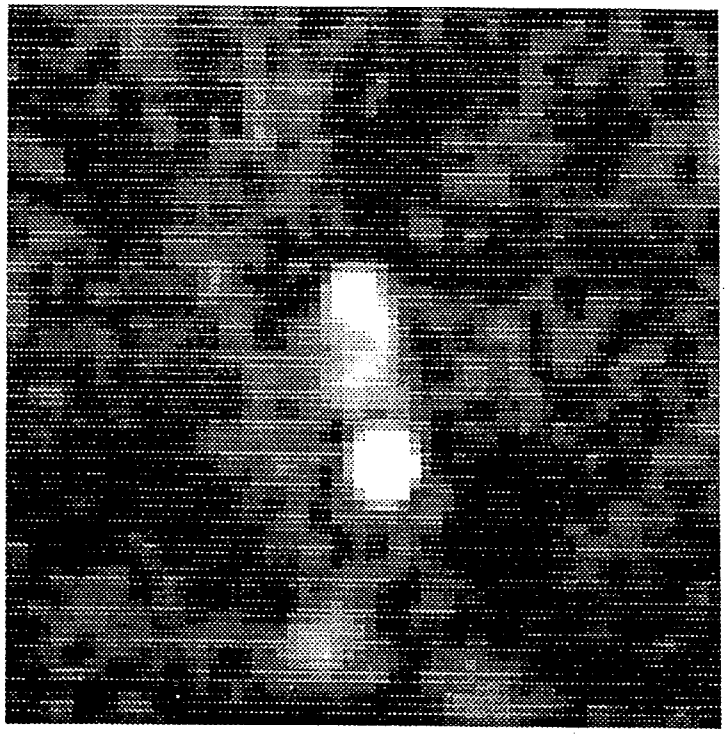
file: image.fits
dir: /tmp

RA 10 7 54.800
DEC 30 20 13.39

63.2 94.0 x



doub
10 7 54.8 30 20 13.39



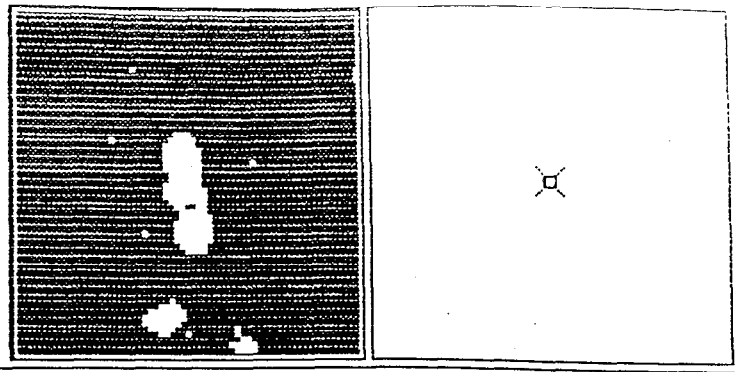
II.1

/tmp/image.fits

SAOimage proctor@tulsir5 Wed Aug 9 14:45:32 1995

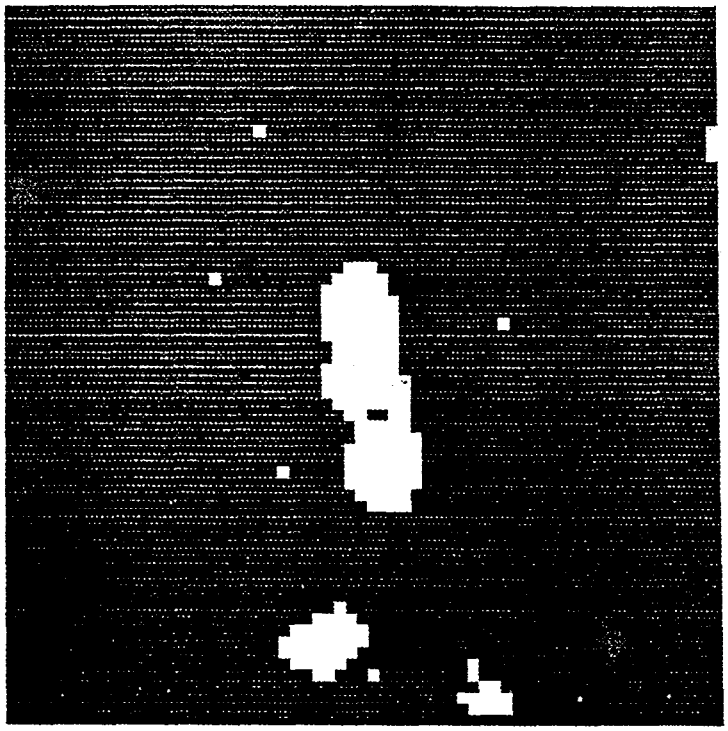
file: image.fits
dir: /tmp

127.0 268.0 x



b = doub gc .00014943

b :



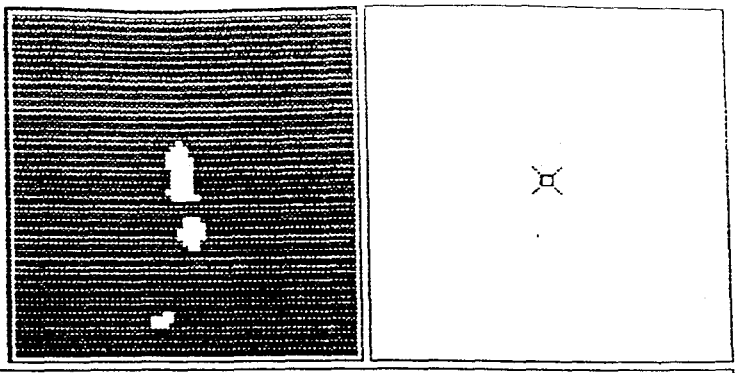
II.2

/tmp/image.fits

SAOimage proctor@taisir5 Wed Aug 9 14:51:08 1995

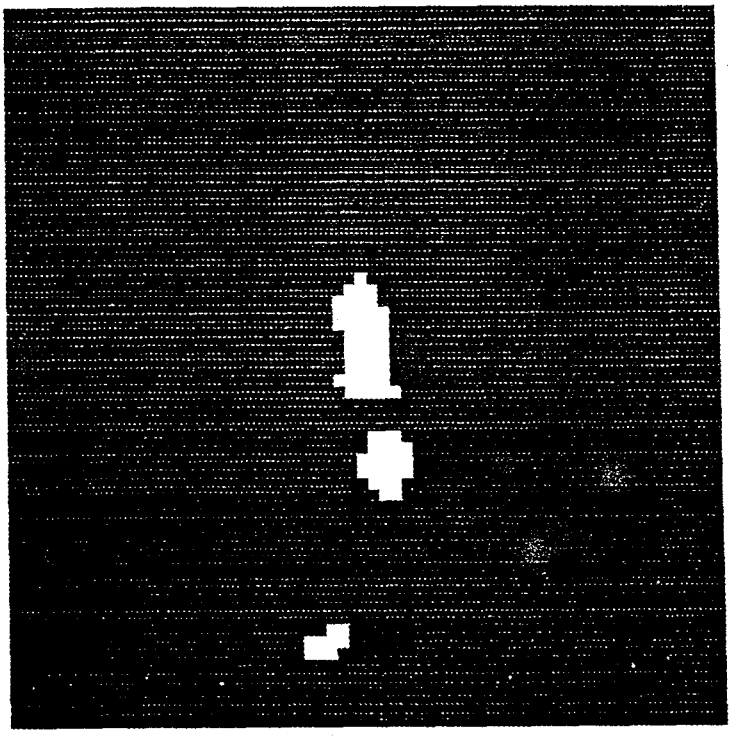
file: image.fits
dir: /tmp

88.8 88.8 x



$B \ominus S$

erode(b,s)



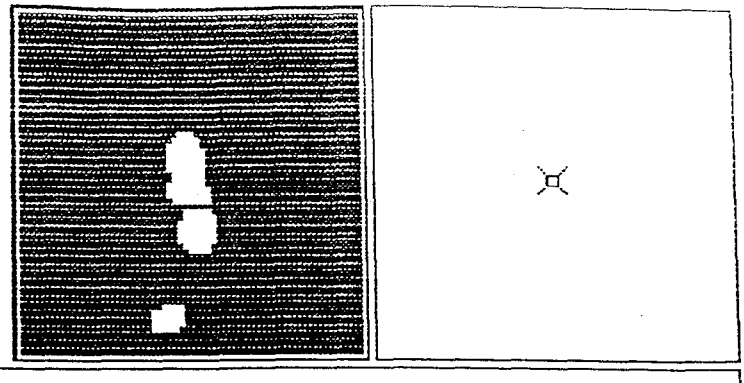
II.3

/tmp/image.fits

SAOimage proctor@taisir5 Wed Aug 9 15:05:48 1995

file: image.fits
dir: /tmp

80.8 97.0 x



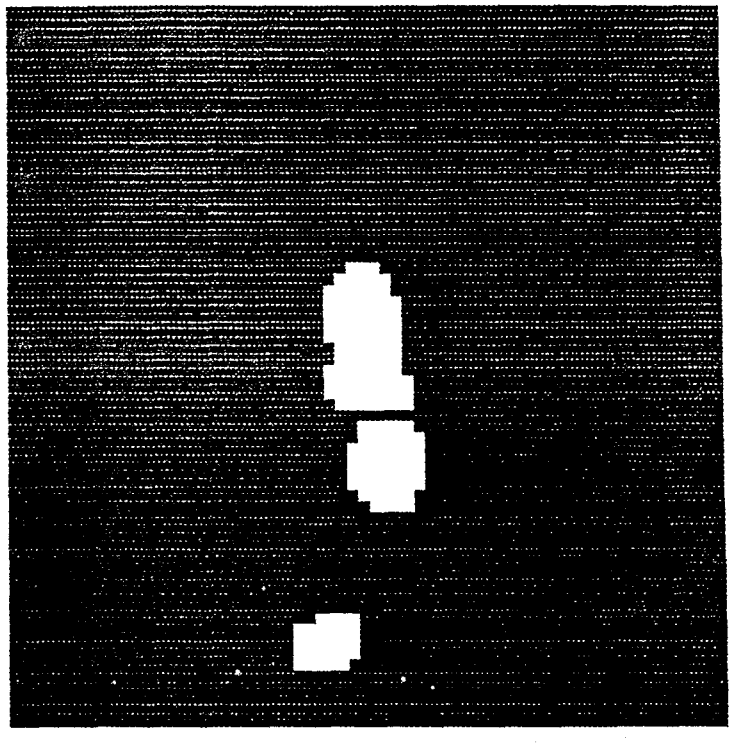
b = doubge .000149×3

S = 1 1 1
1 1 1
1 1 1

C = dilate(erode,S),S)

C =

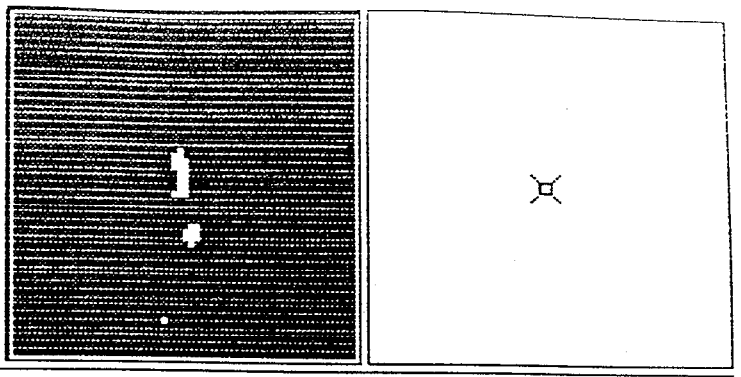
$(B \ominus S) \oplus S$



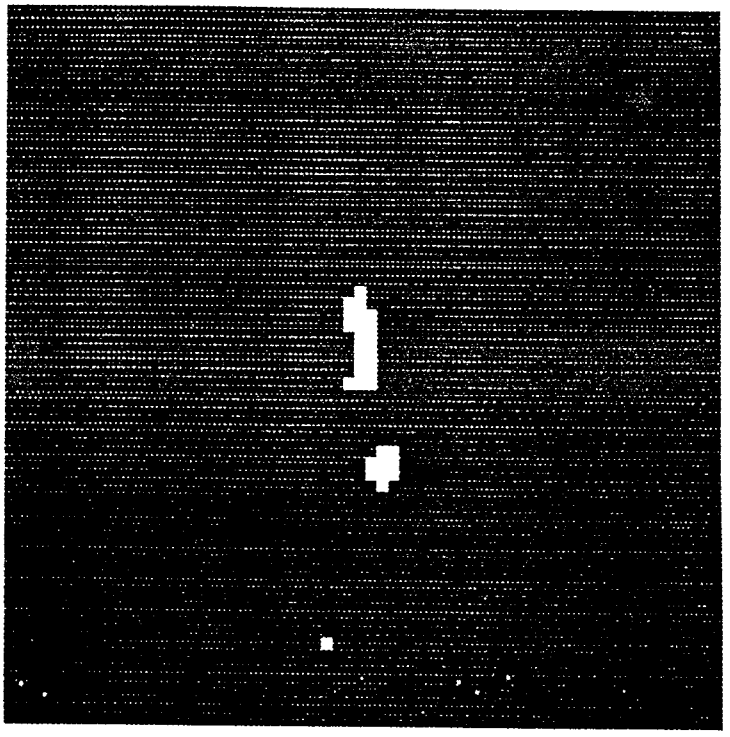
II.4

file: image.fits
dir: /tmp

186.0 286.0 x



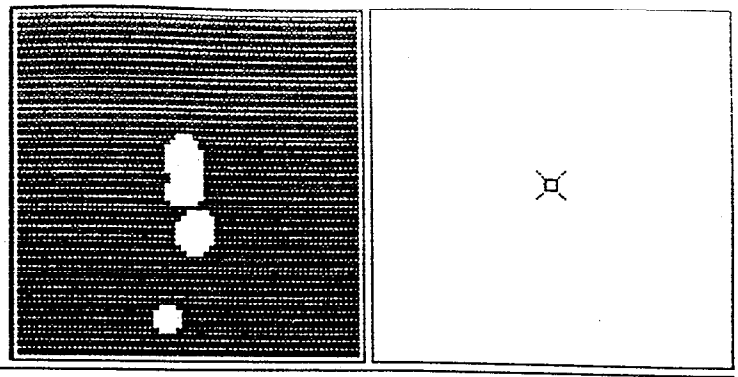
erode(b,S5)



II.5

file: image.fits
dir: /tmp

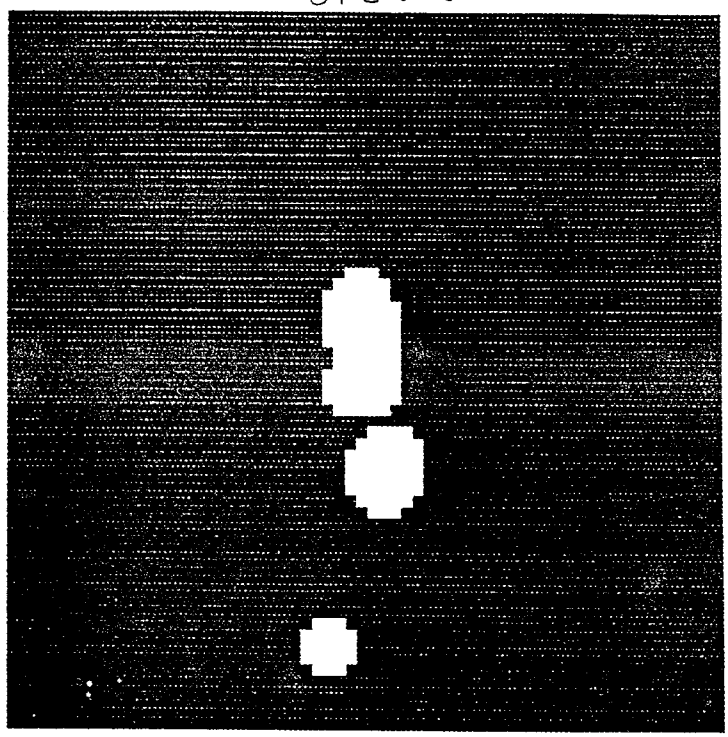
148.0 286.0 x



b = doubge .000149 x3
c = dilate(erode(b,S5),S5)

S5 = 0 1 1 1 0
1 1 1 1 1
1 1 1 1 1
1 1 1 1 1
0 1 1 1 0

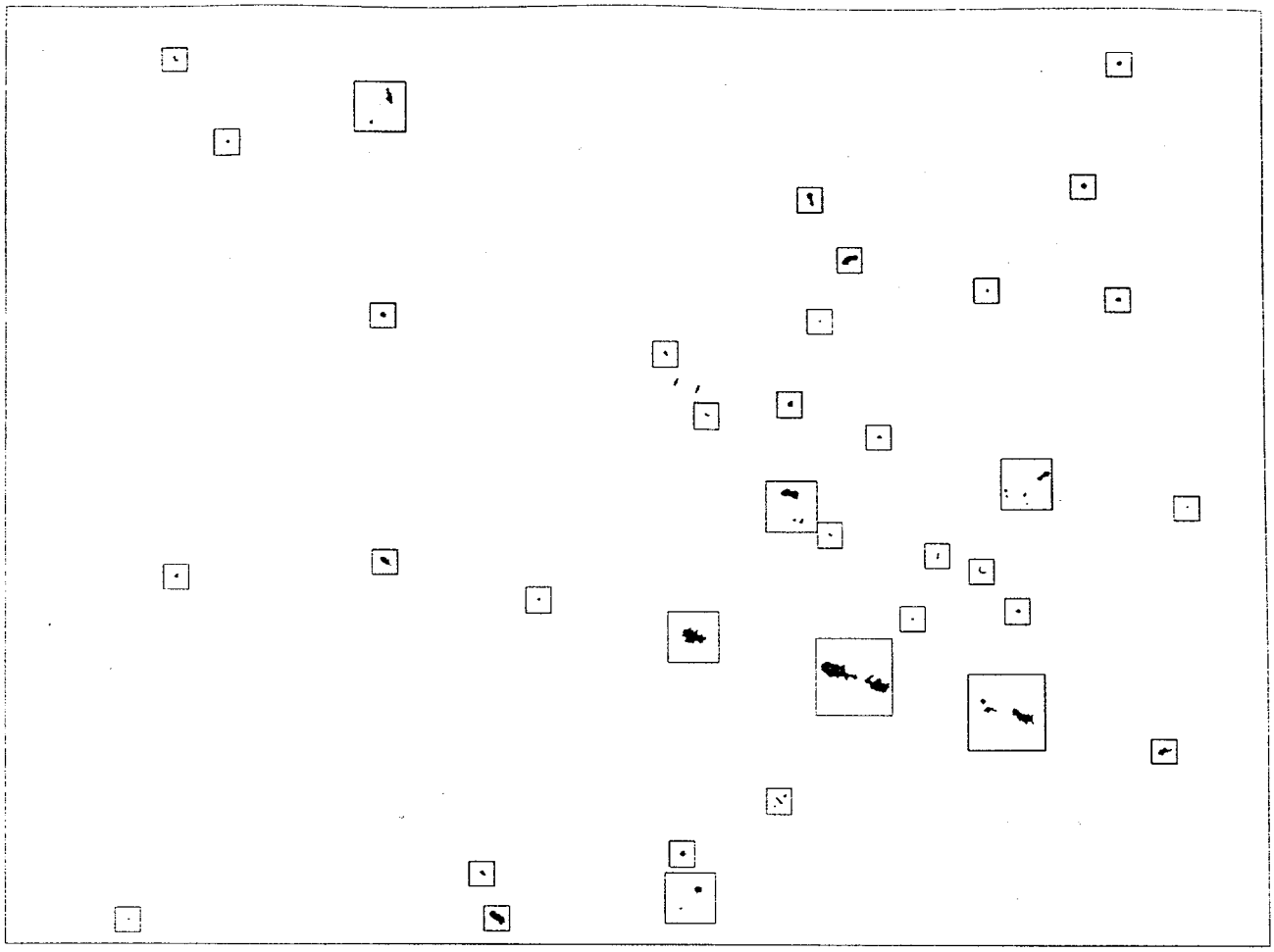
OPENING



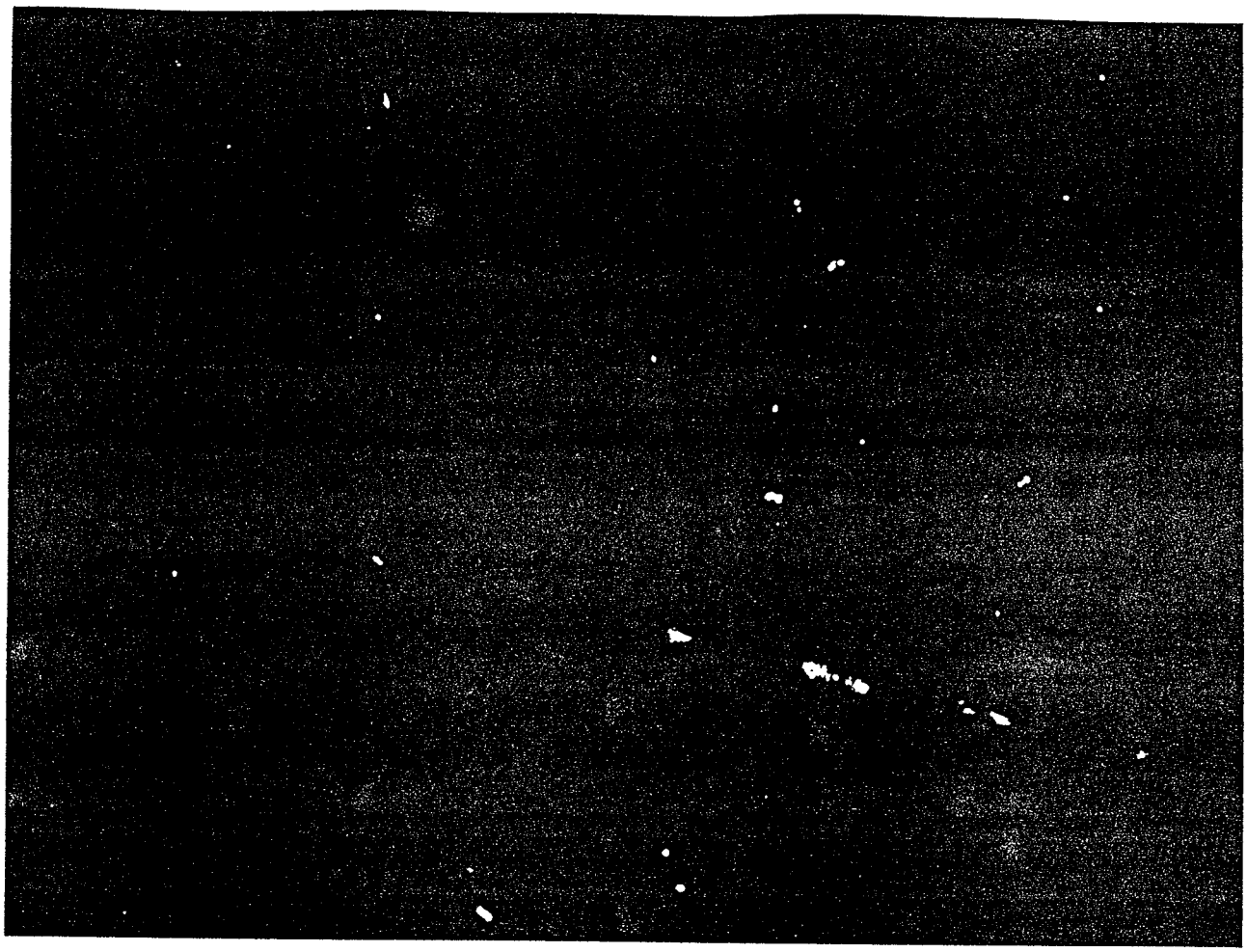
smoothes contour
breaks narrow isthmuses
eliminates small islands
and sharp peaks, capes

II. 6

FIT11060+3017 with r_gap =33 pixels



FIT11060+3017A



TRAINING SET

Visual Classifications:

233	Bent
426	Somewhat Bent
1255	Non-Bent
721	? (too noisy to tell)
10	Misc.
2645	Total

TEST SET

Visual Classifications:

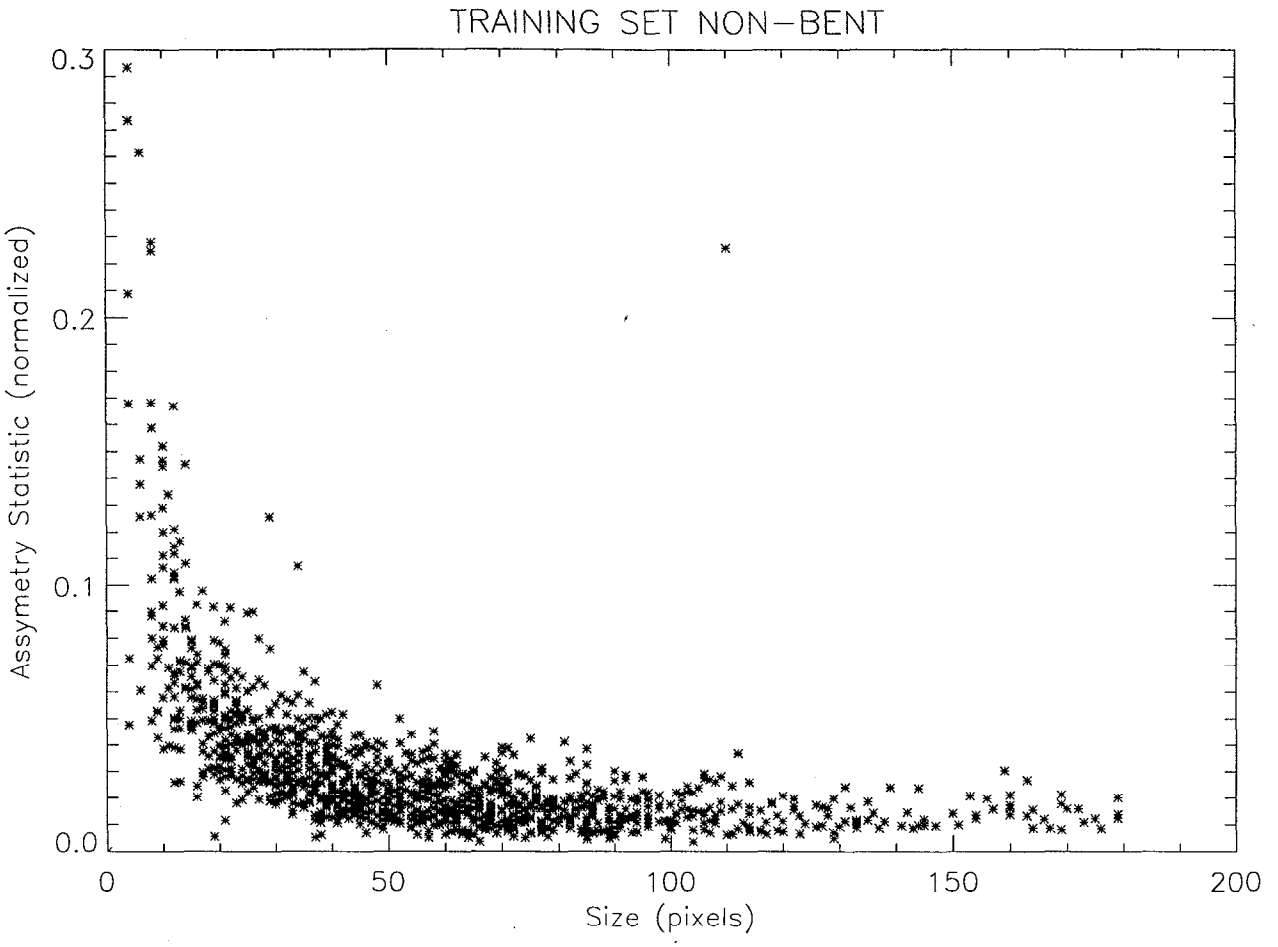
52	Bent
165	Somewhat Bent
612	Non-Bent
227	? (too noisy to tell)
3	Misc.
1059	Total

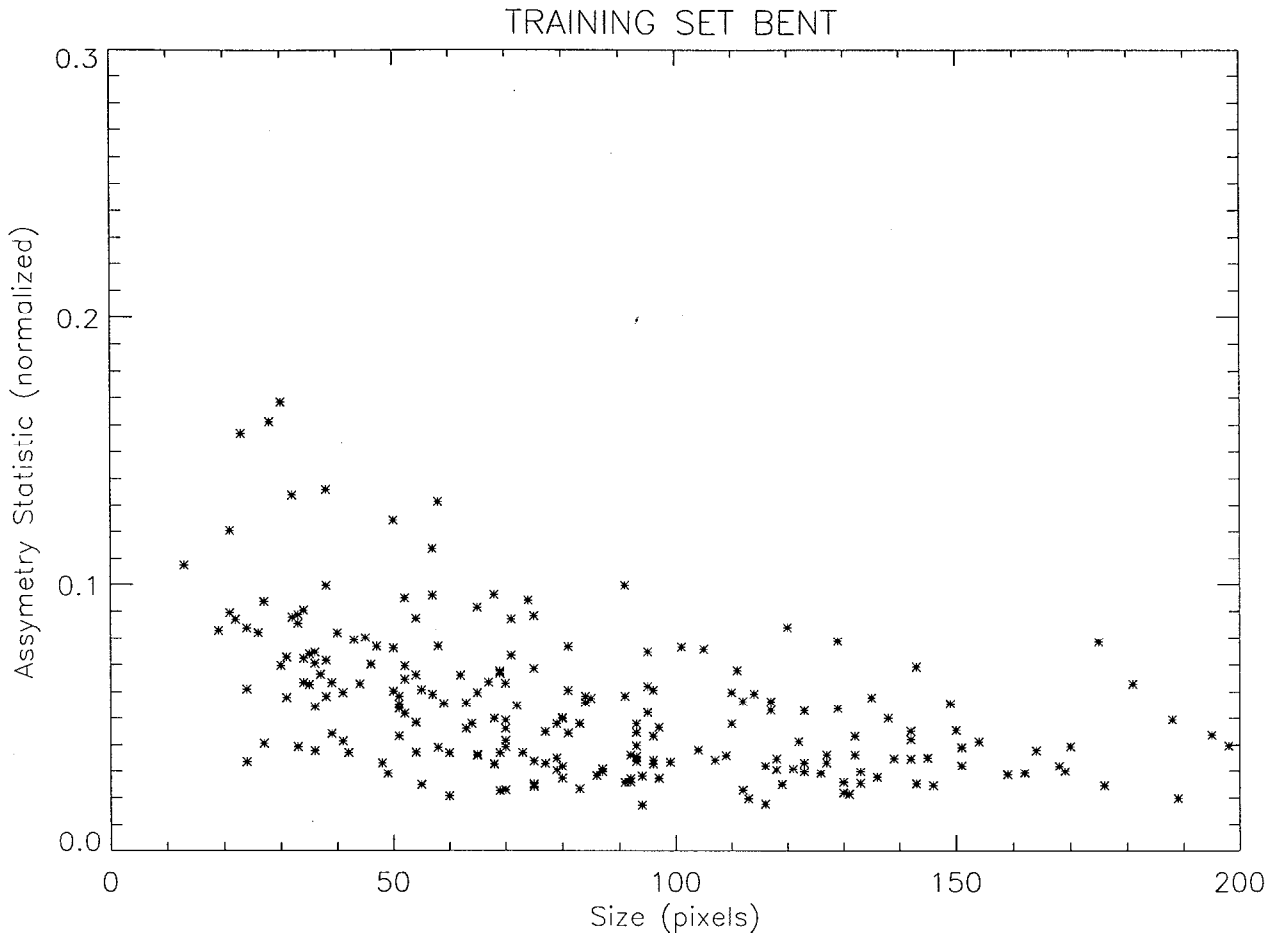
PATTERN RECOGNITION - INVARIANCE for FEATURES

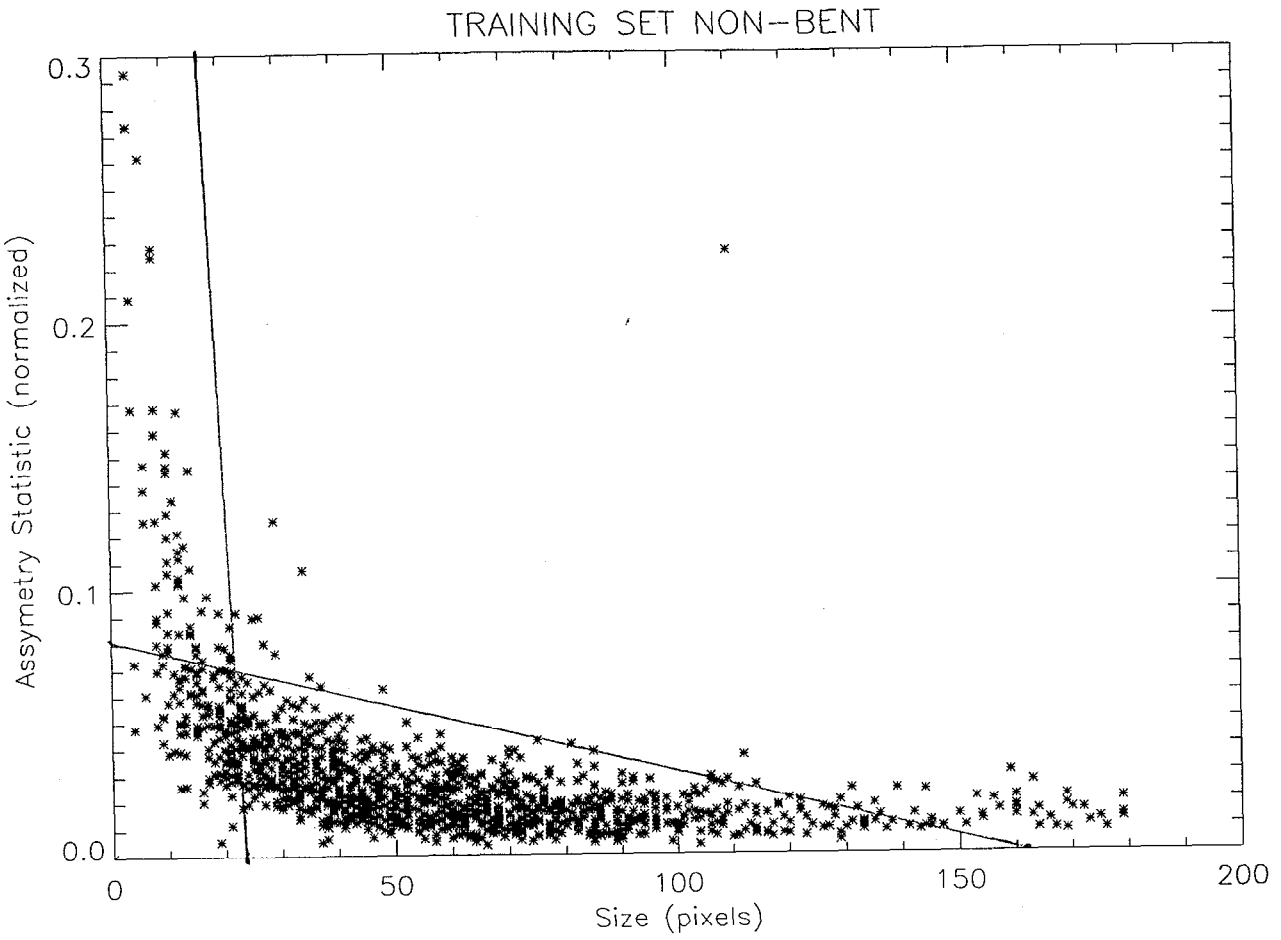
Translation

Rotation

Scale - the 'Problem'


TRAINING SET NON-BENT
Assymetry Statistic (normalized)
0.3
0.2
0.1
0.0
0
50
100
150
200
Size (pixels)

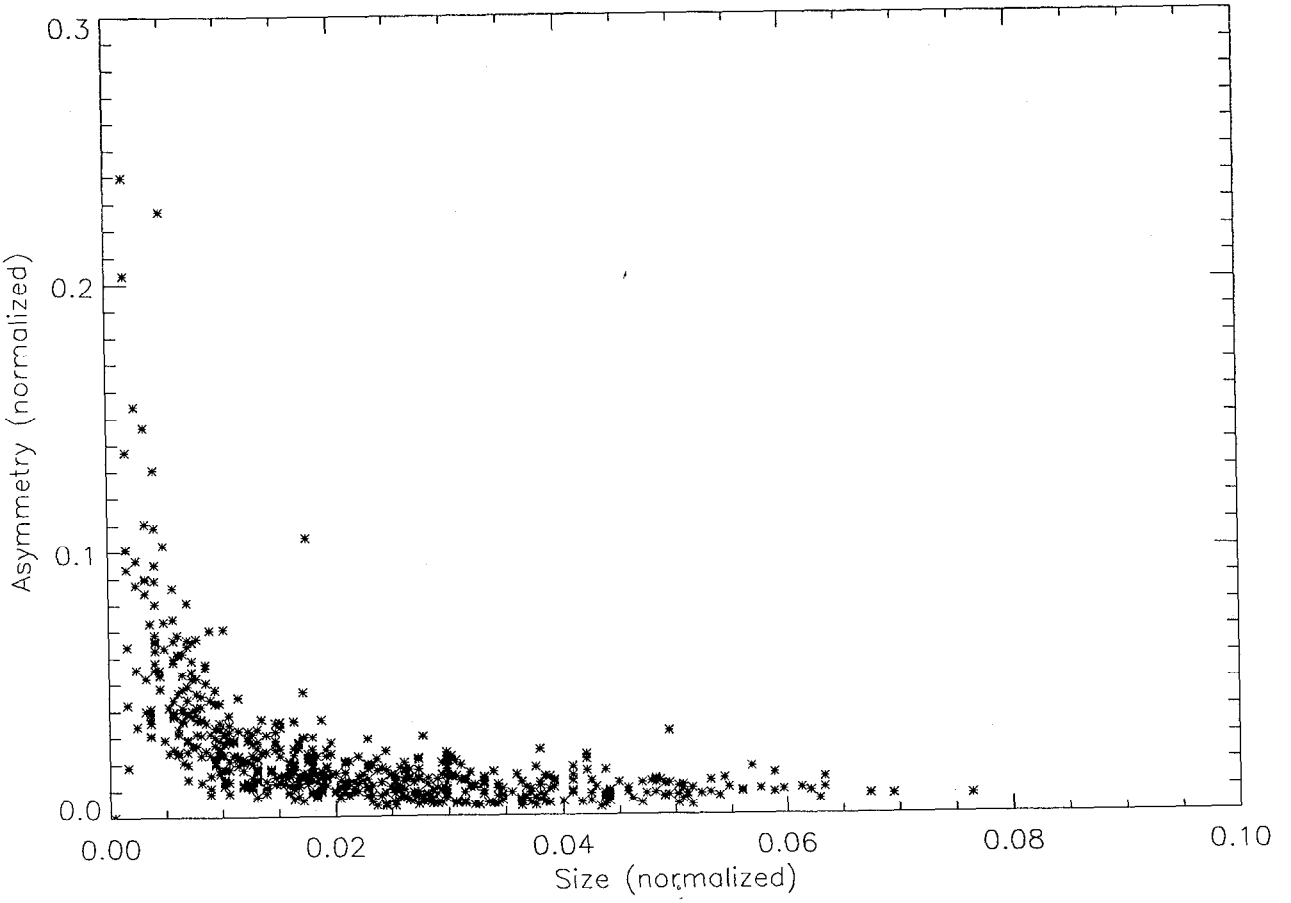

TRAINING SET BENT
Assymetry Statistic (normalized)
0.0
0.1
0.2
0.3
0
50
100
150
200
Size (pixels)



Decision Tree Boundaries

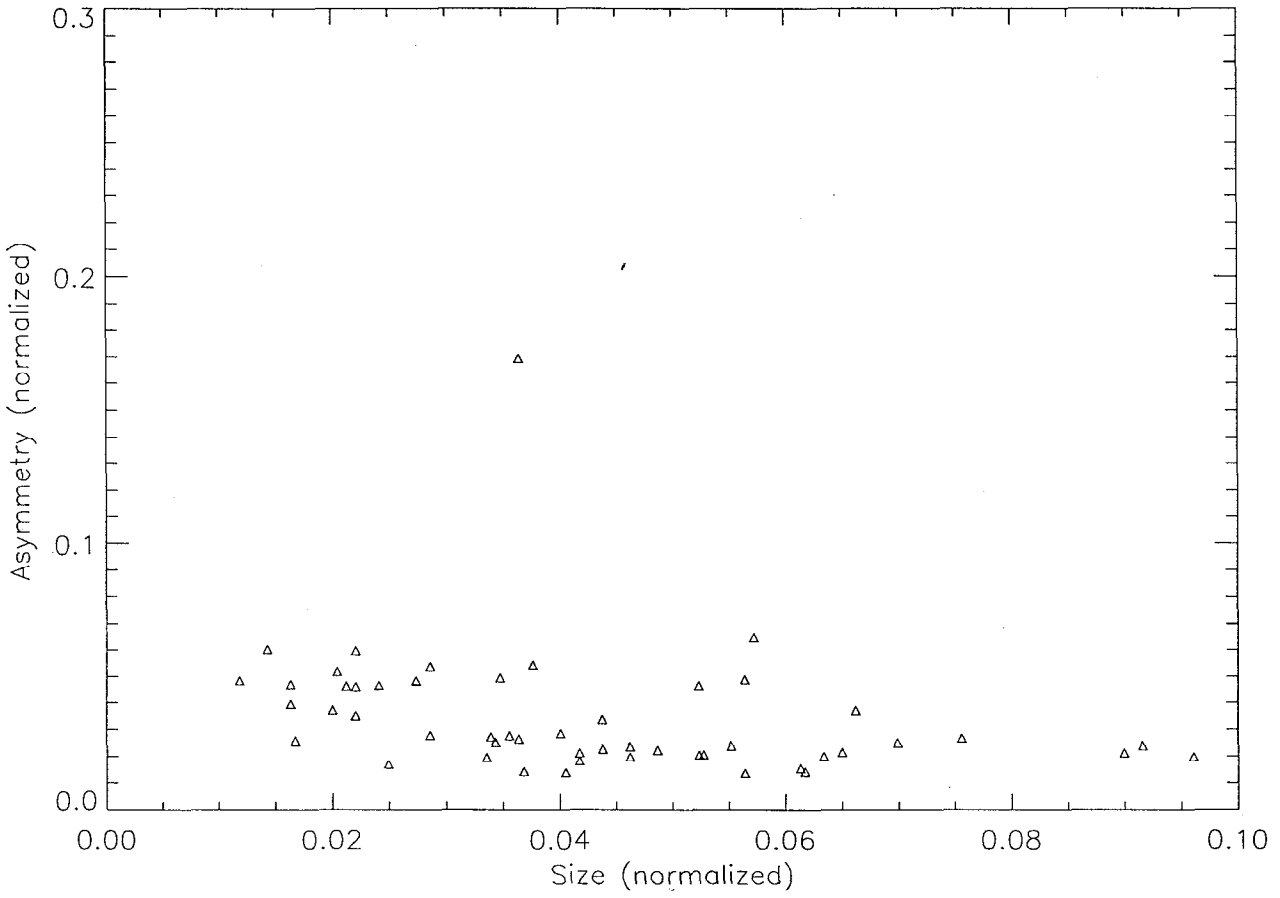
Test Set

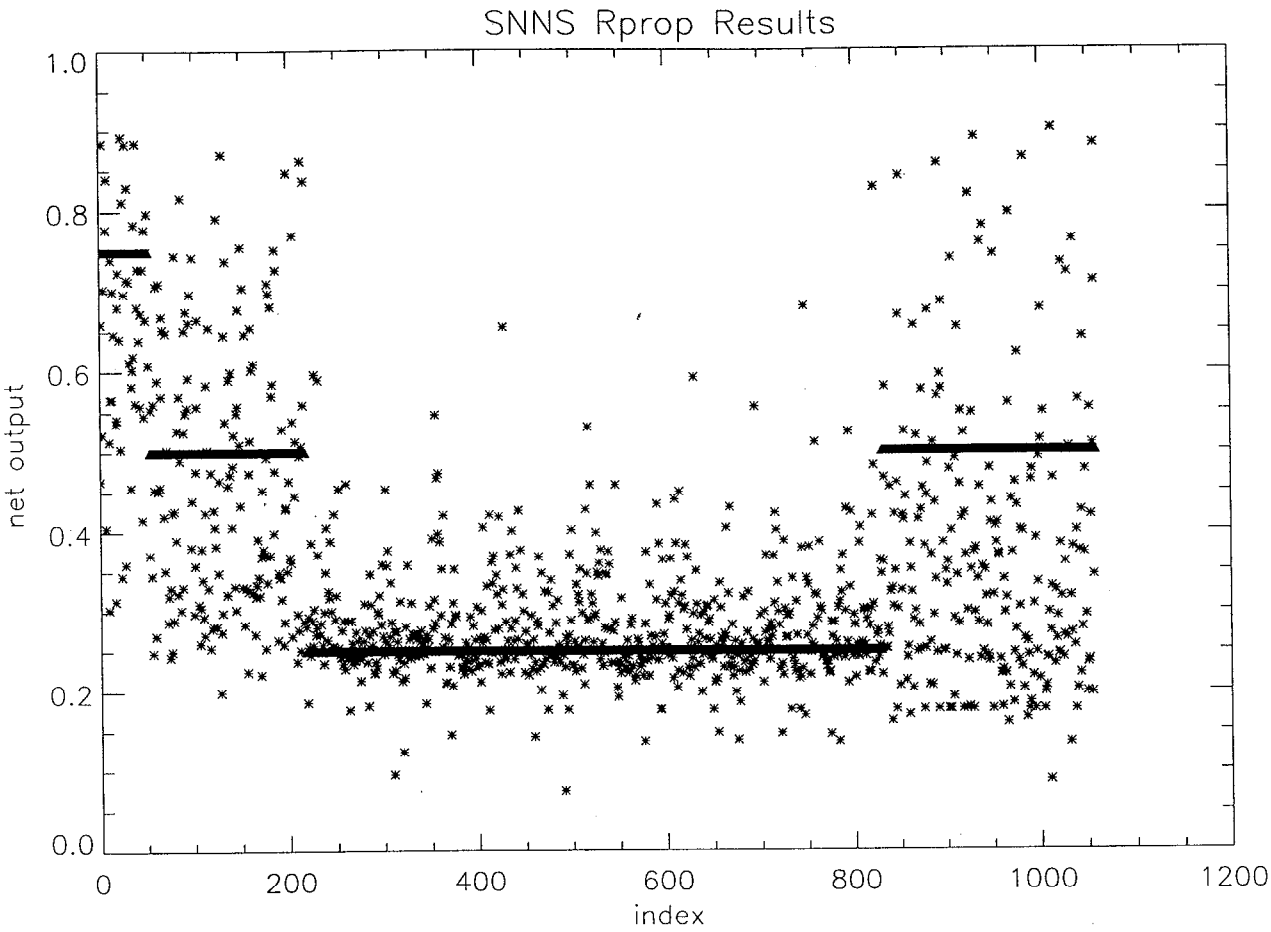
Visual Classification NonBents



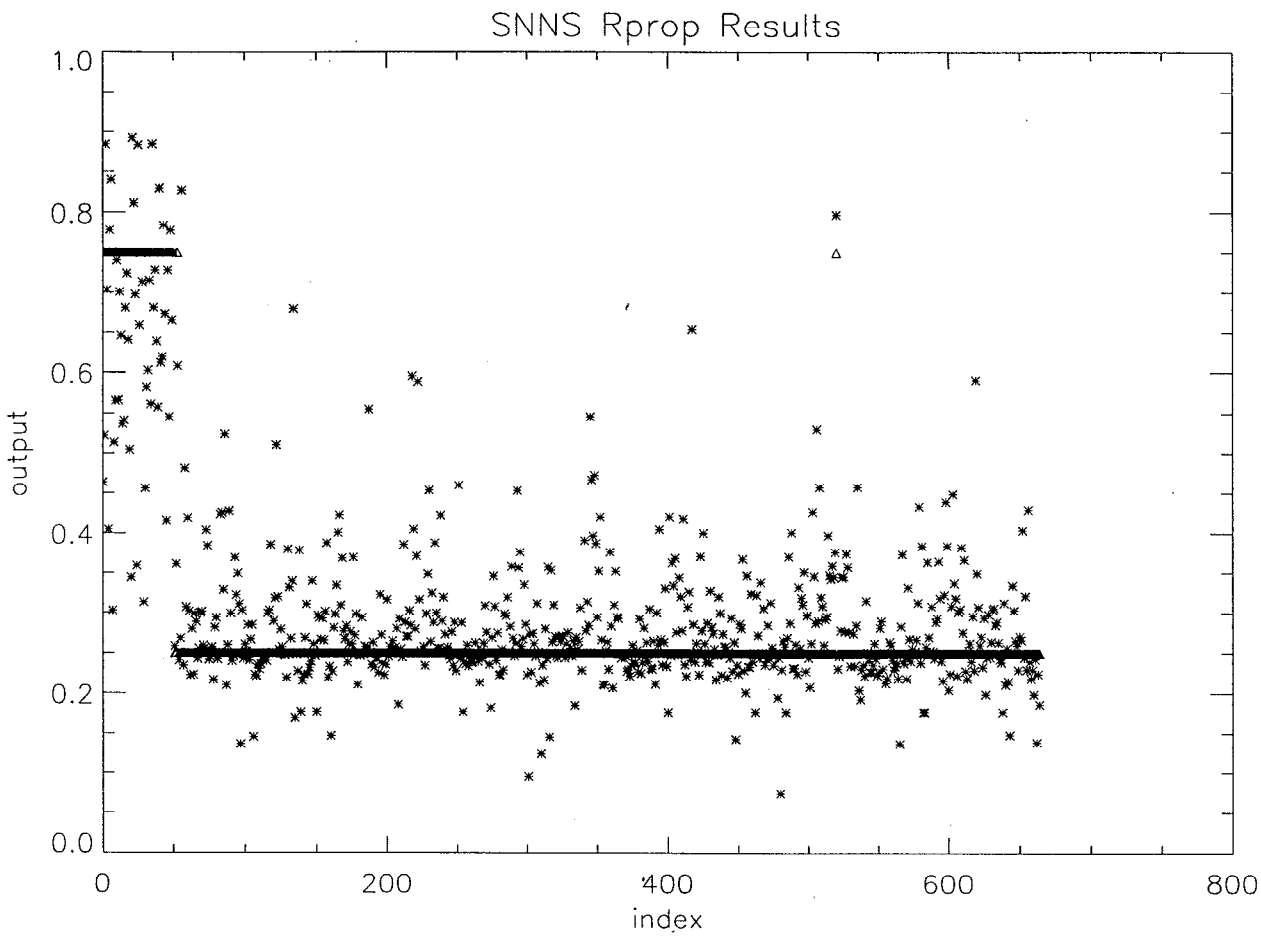
Test Set

Visual Classification Bents

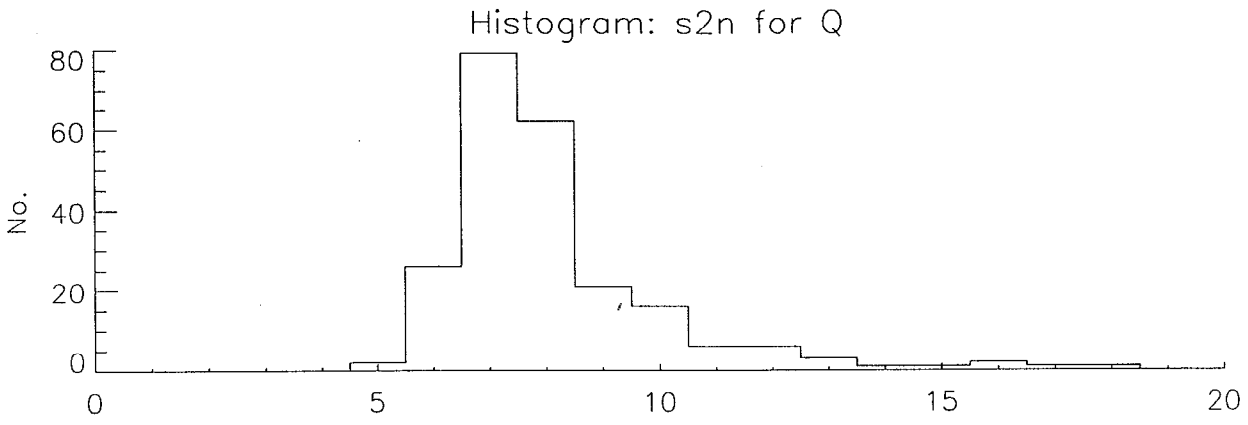


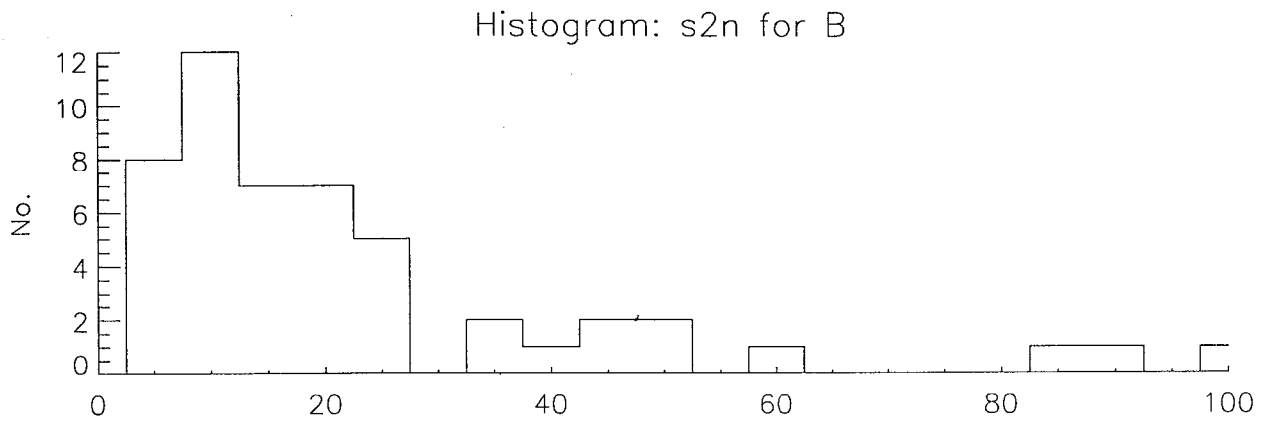


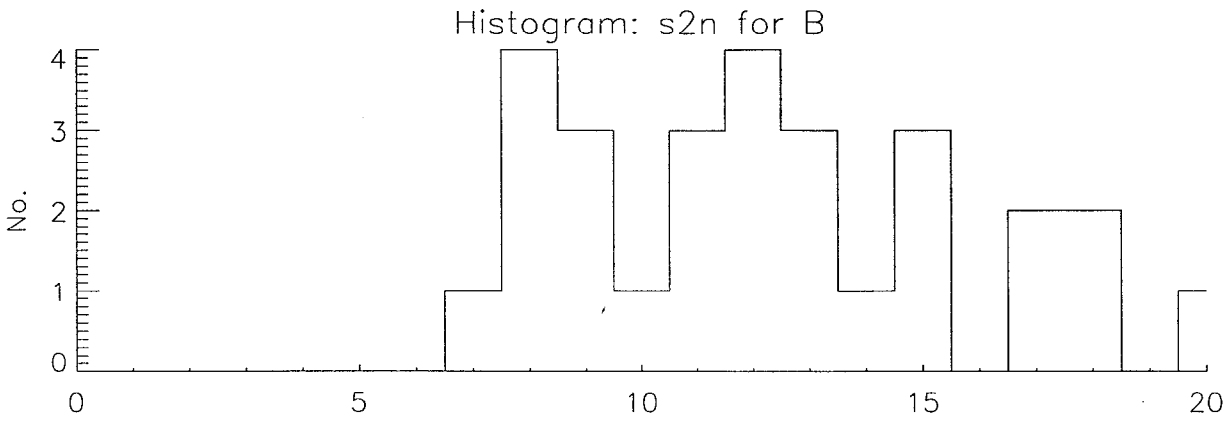
B BS Non Bent Q

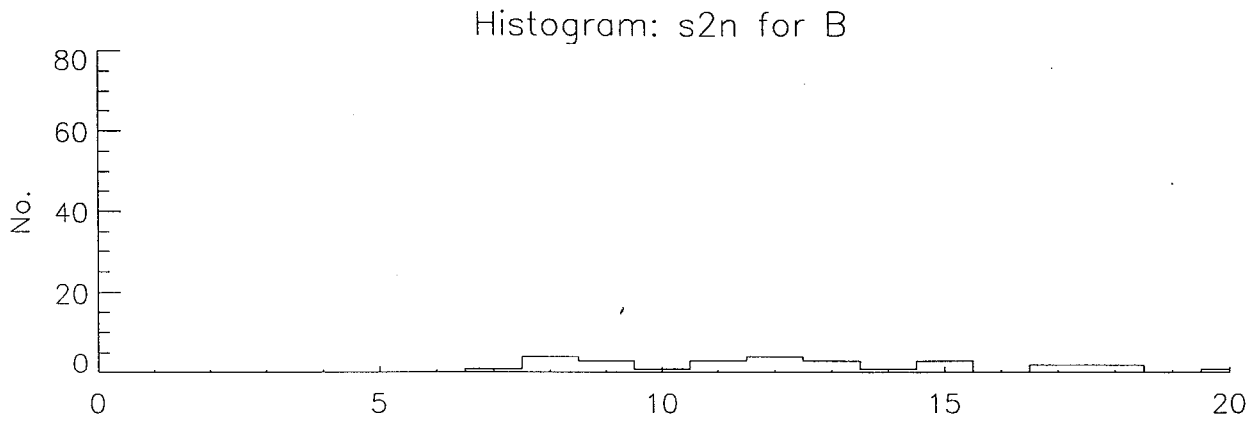


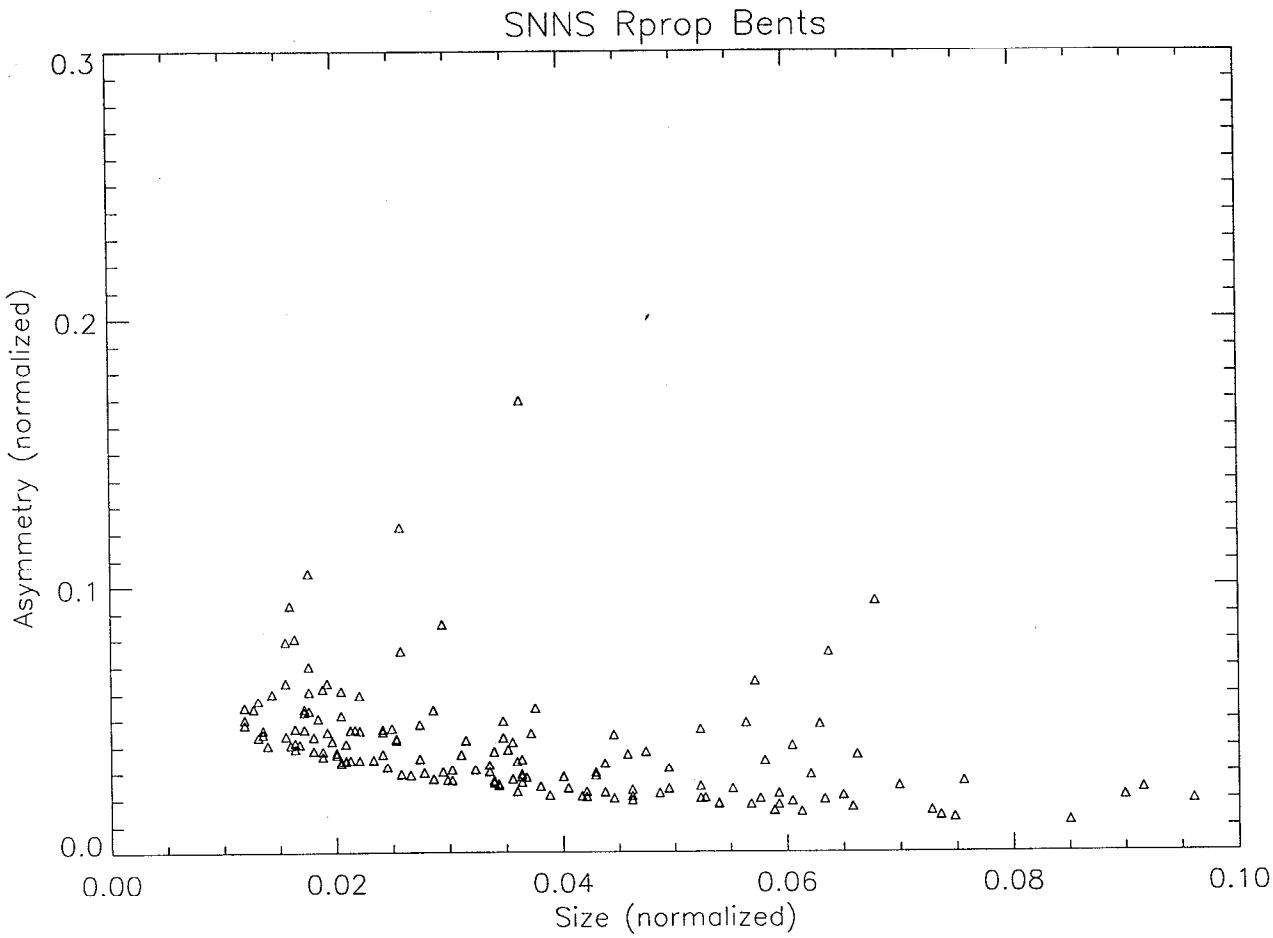
Bent Non Bent

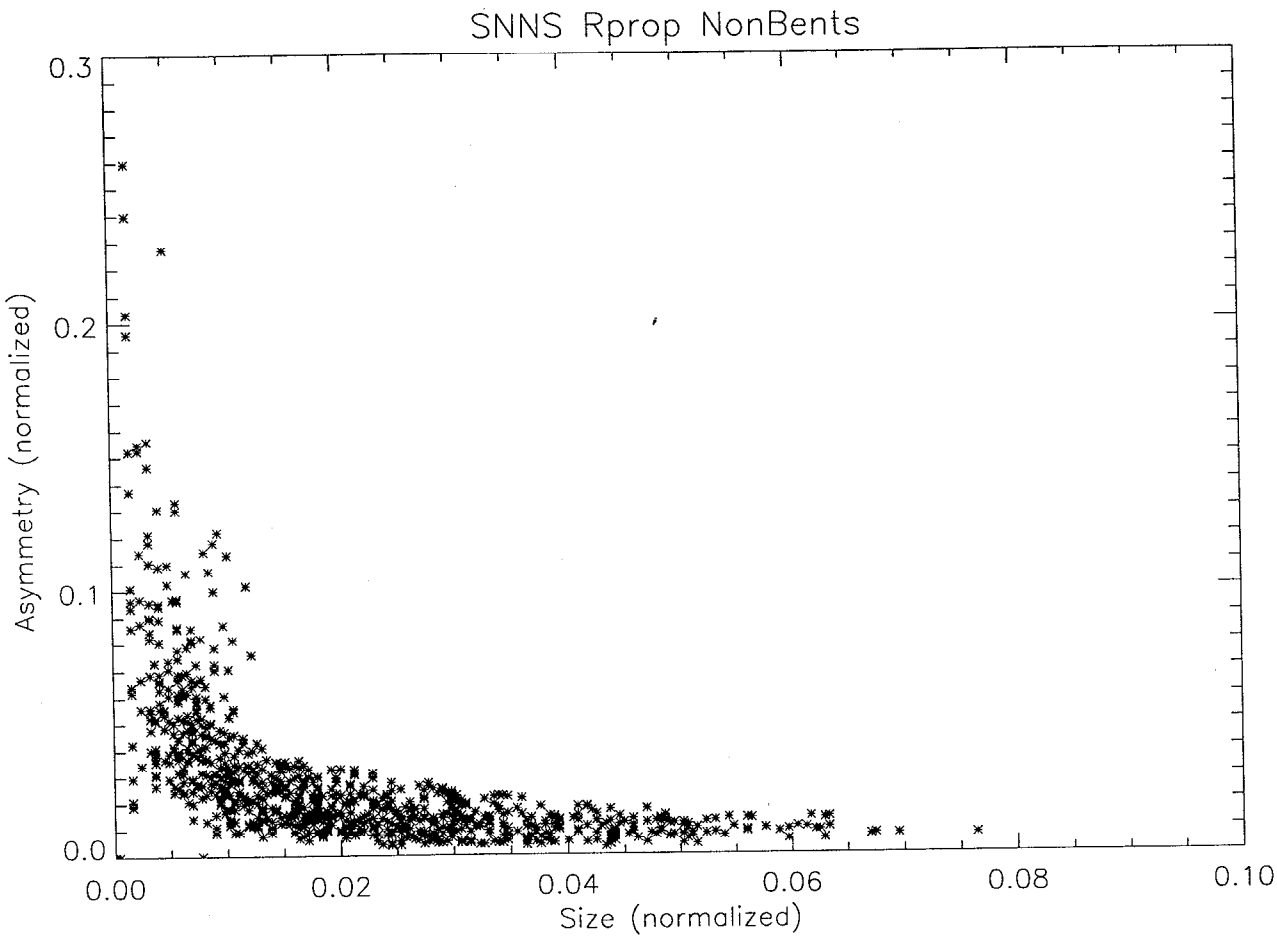

Histogram: s2n for Q
No.
80
60
40
20
0
0
5
10
15
20


Histogram: s2n for B
No.
0
2
4
6
8
10
12
0
20
40
60
80
100


Histogram: s2n for B
No.
0
1
2
3
4
0
5
10
15
20


Histogram: s2n for B
No.
80
60
40
20
0
0
5
10
15
20


SNNS Rprop Bents
Asymmetry (normalized)
0.0
0.1
0.2
0.3
0.00
0.02
0.04
0.06
0.08
0.10
Size (normalized)

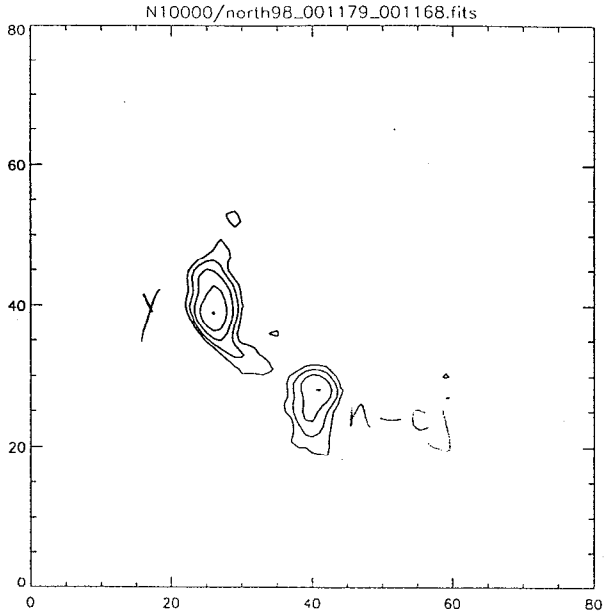

SNNS Rprop NonBents
Asymmetry (normalized)
0.3
0.2
0.1
0.0
0.00
0.02
0.04
0.06
0.08
0.10
Size (normalized)

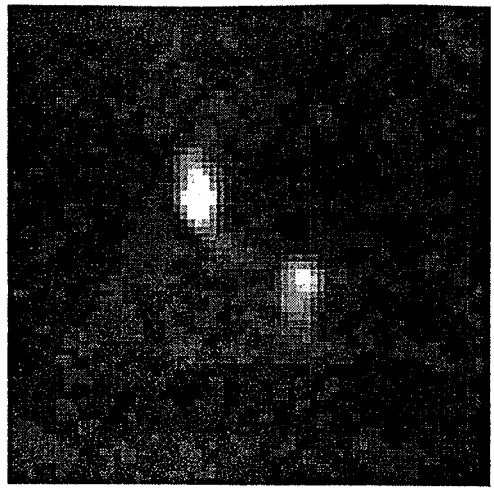
	BENTS FOUND/TOTAL BENTS	NON-BENTS(F.P.)/BENTS FOUND
OC1 decision tree catalog features	.71	.75
OC1 decision tree with image stats	.88	.54
SNNS Rprop	.85	.25

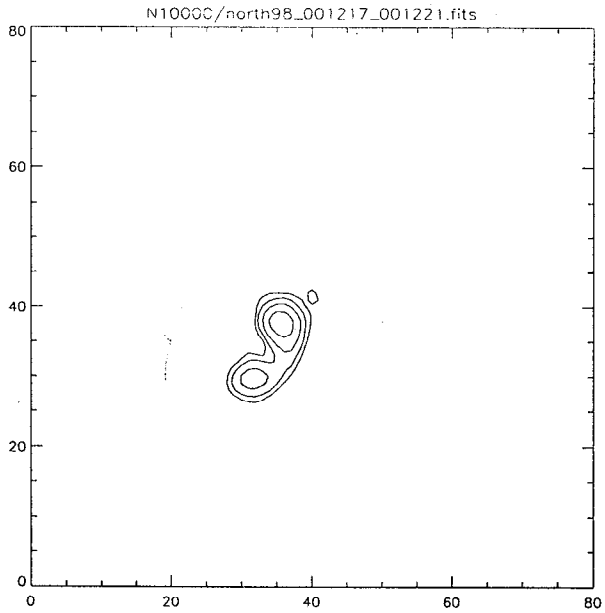
Visual Classifications of 115 OC1 Selected Non-Asymmetric Sources

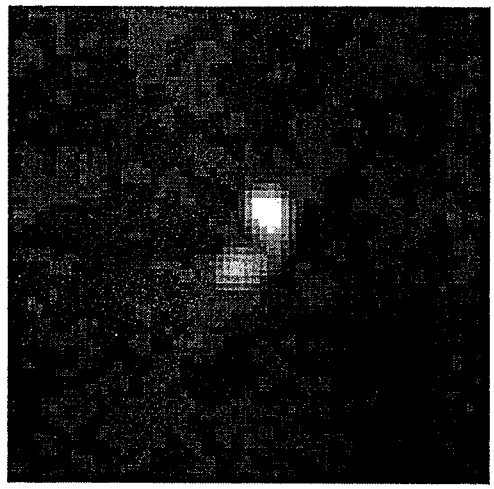
27 Yes
21 Maybe
15 Core-Jet
52 No

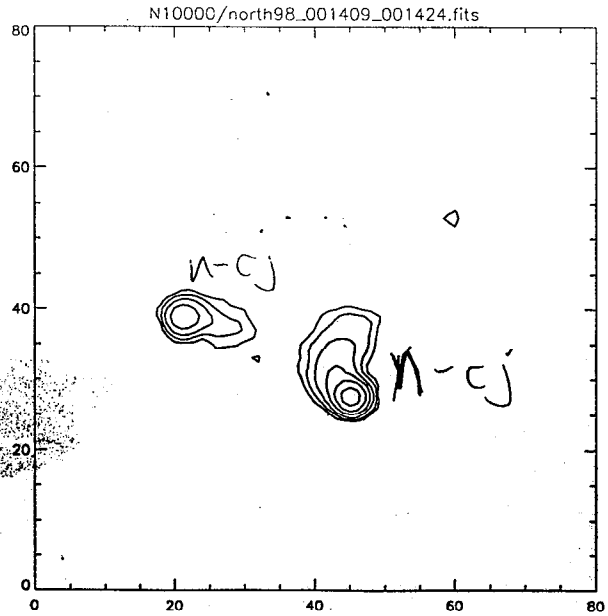
(generated by [illegible]
w/test 149.tot.ocl.reo

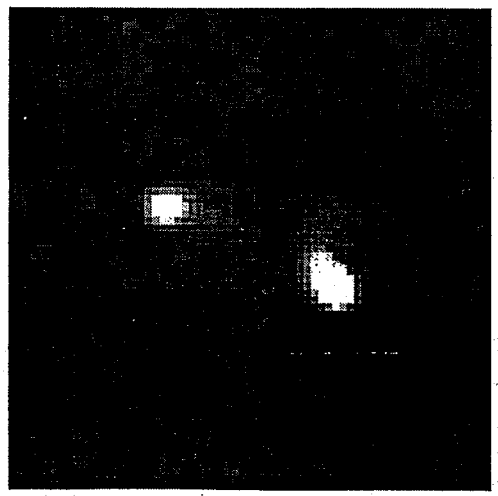


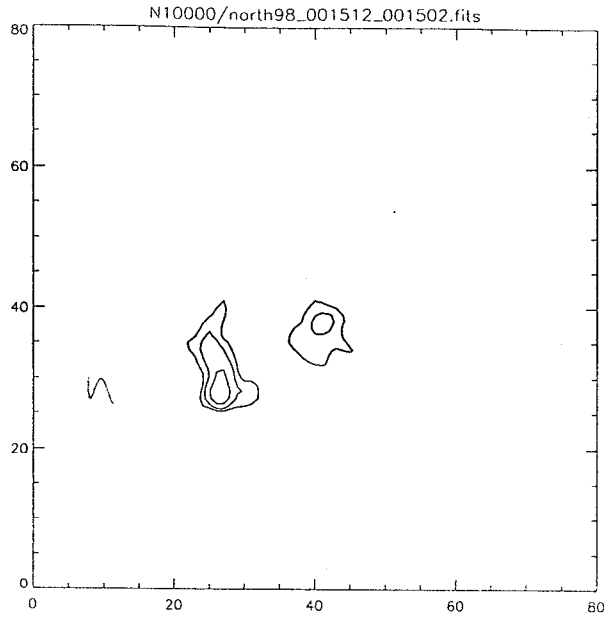


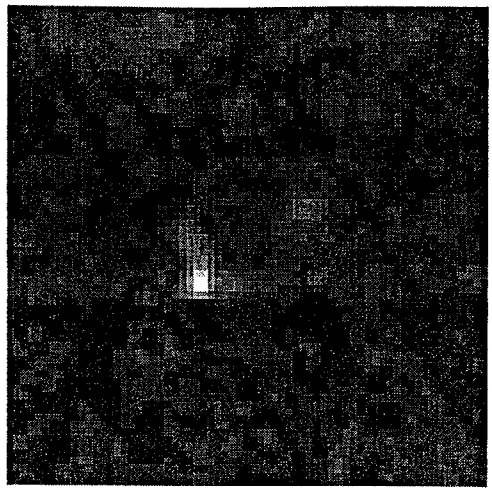


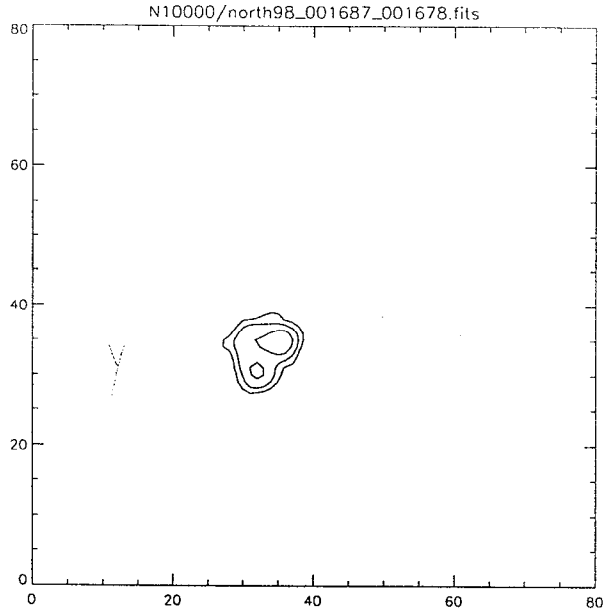


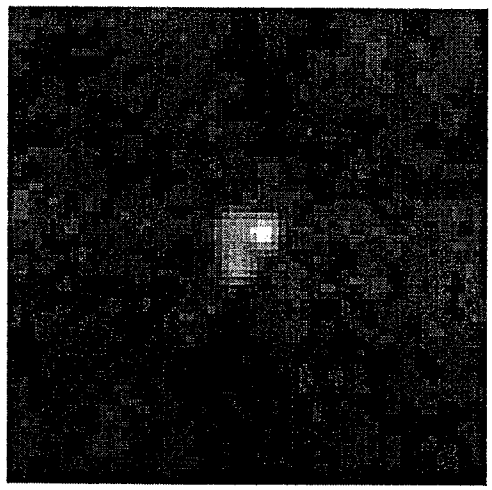


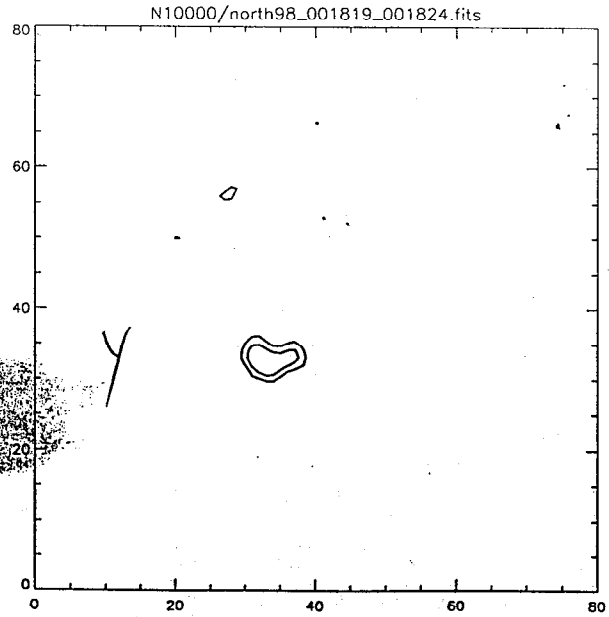


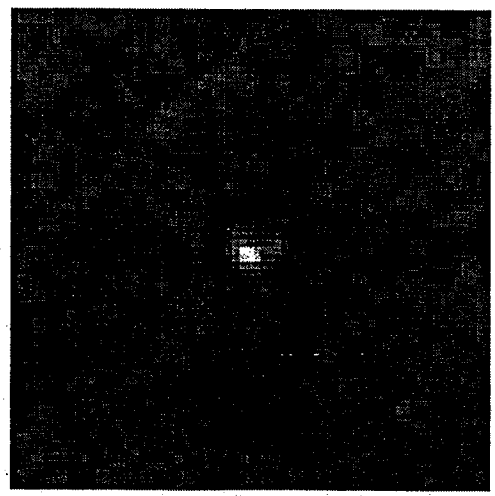

N10000/north98_001512_001502.fits
80
60
40
20
0
0
20
40
60
80
n



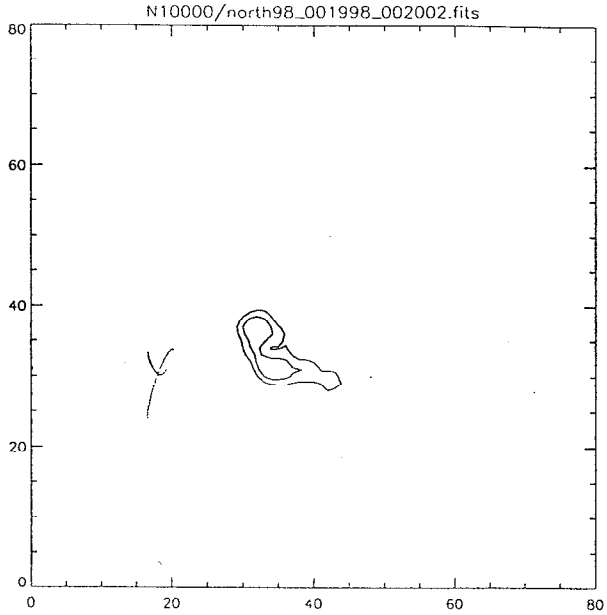

N10000/north98_001687_001678.fits
80
60
40
20
0
0
20
40
60
80
Y

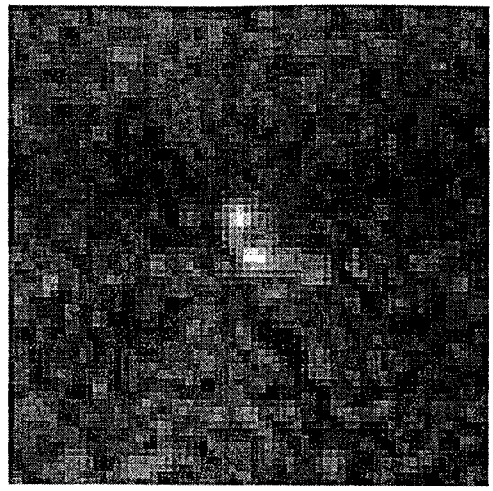



N10000/north98_001819_001824.fits
80
60
40
20
0
0
20
40
60
80
Y

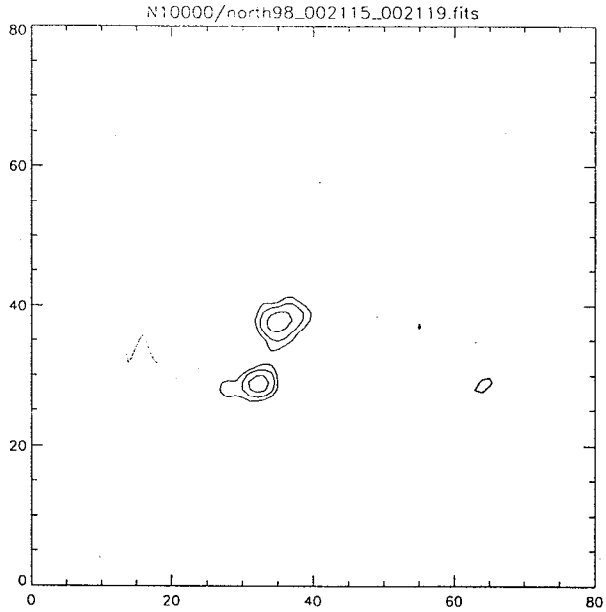


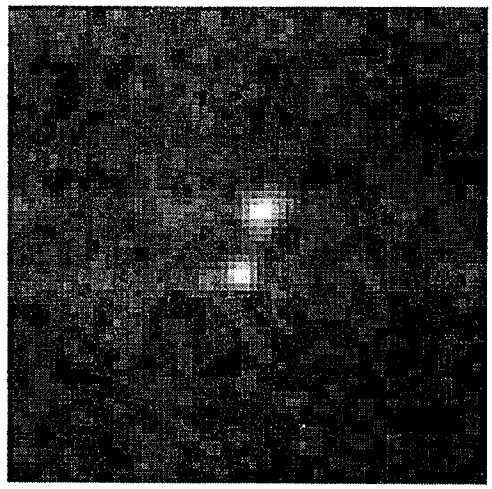
N10000/north98_001998_002002.fits



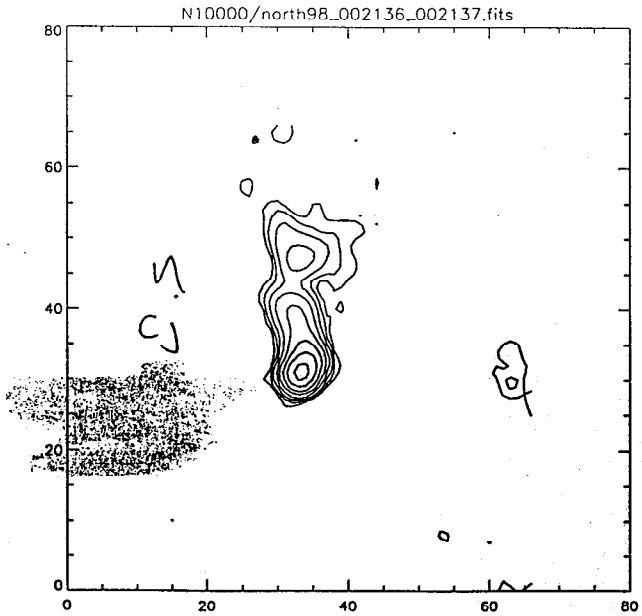


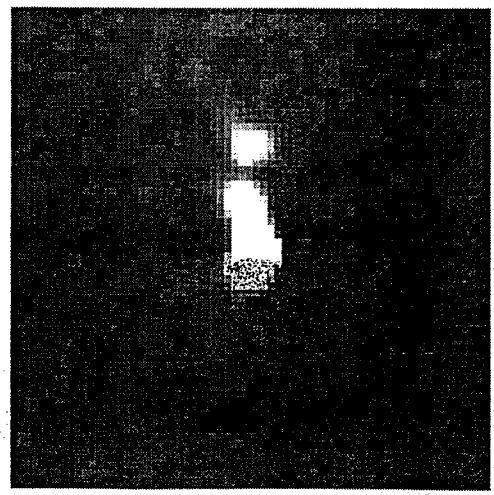
N10000/north98_002115_002119.fits



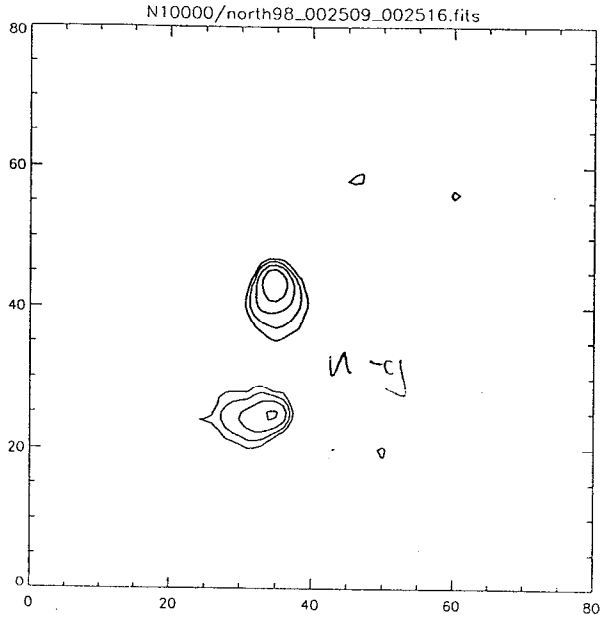


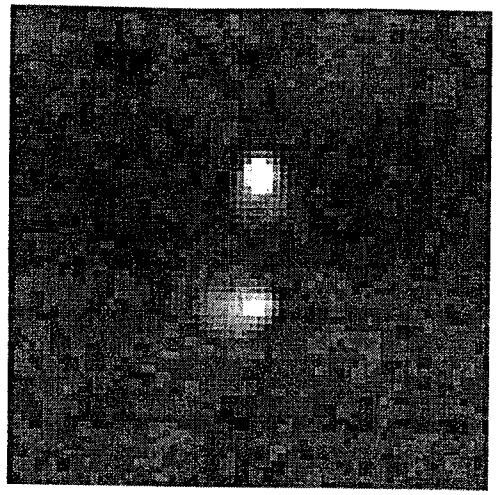
N10000/north98_002136_002137.fits



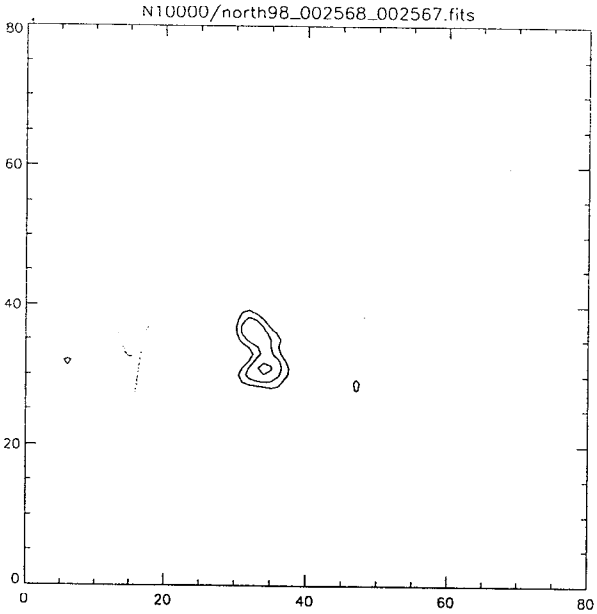


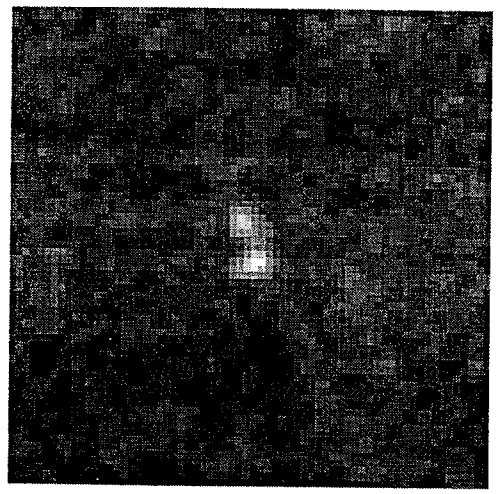
N10000/north98_002509_002516.fits



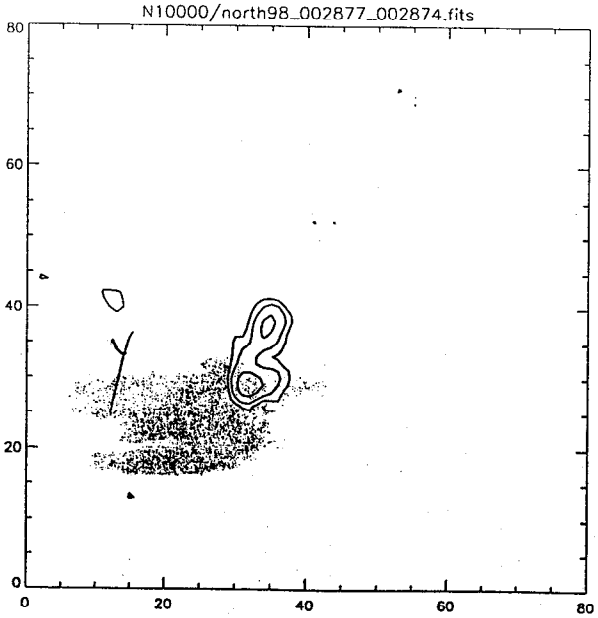


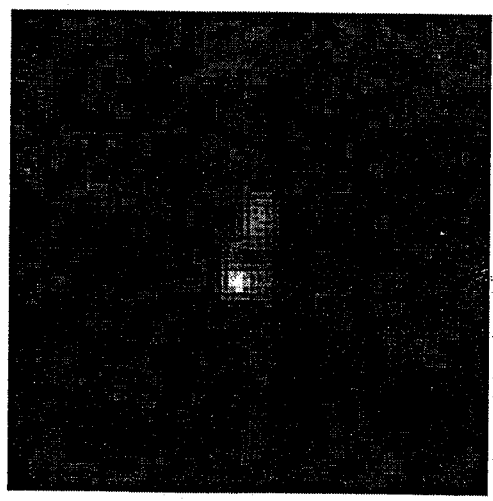
N10000/north98_002568_002567.fits



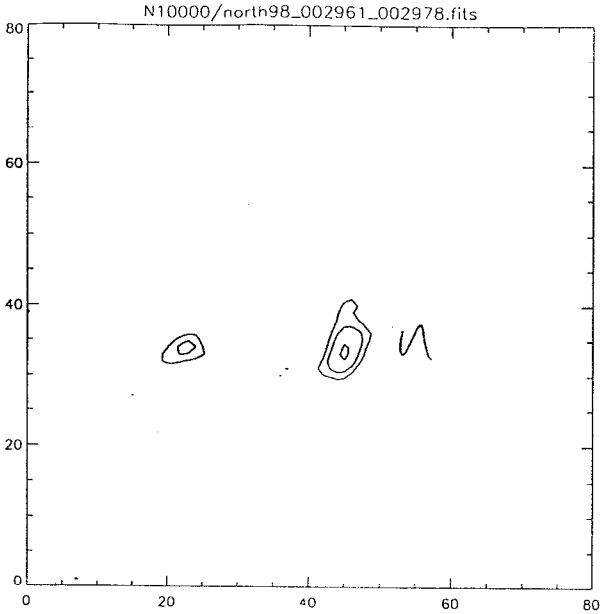


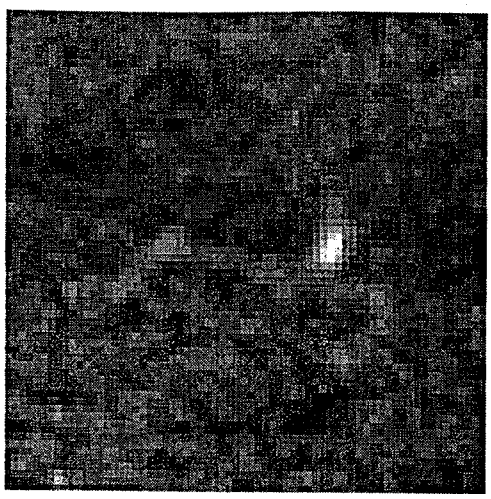
N10000/north98_002877_002874.fits



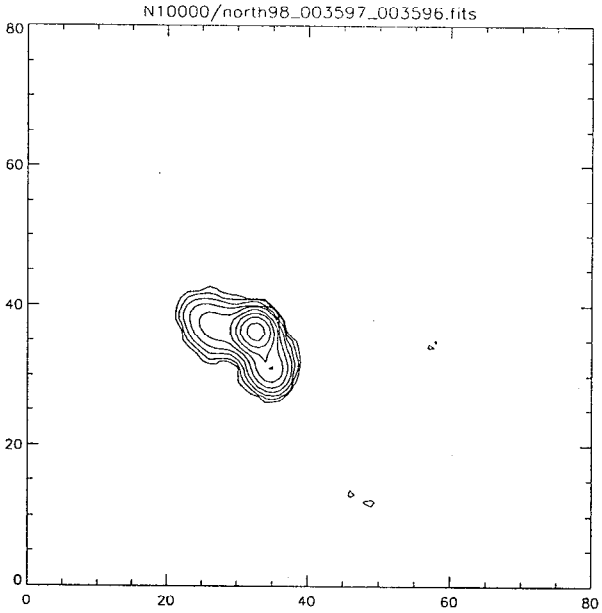


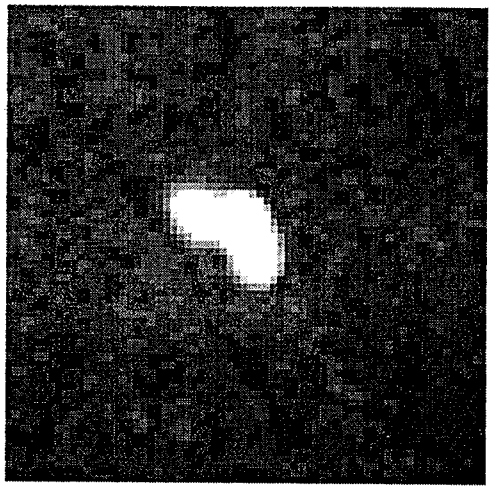
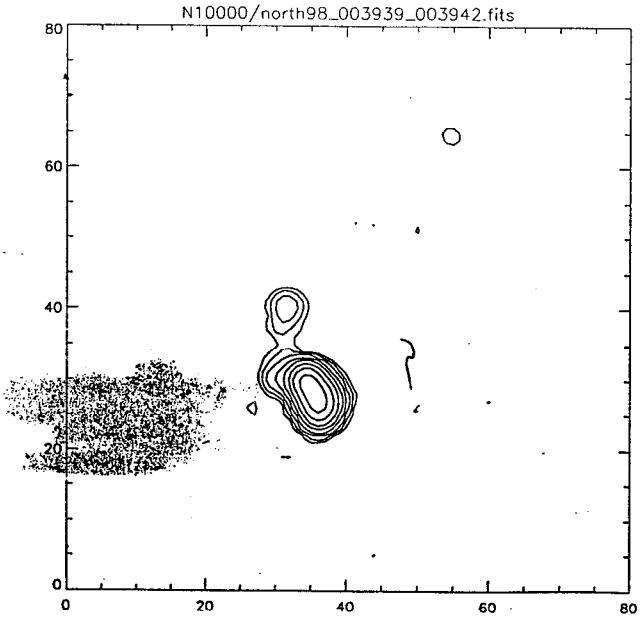
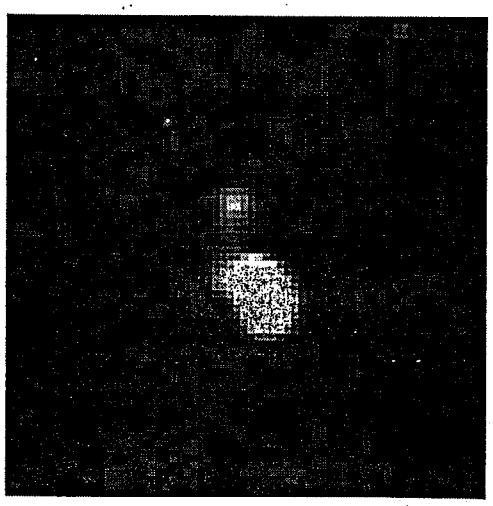
N10000/north98_002961_002978.fits





N10000/north98_003597_003596.fits





N10000/north98_003939_003942.fits