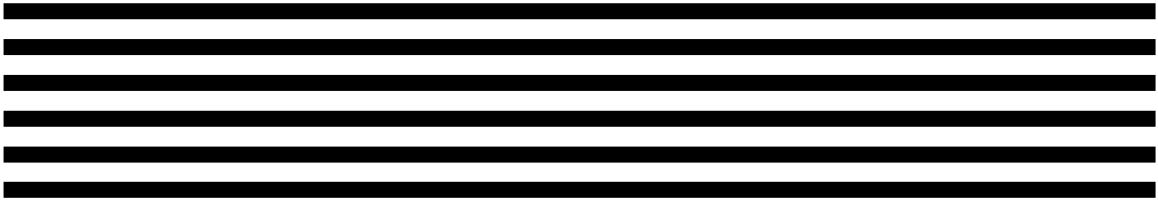
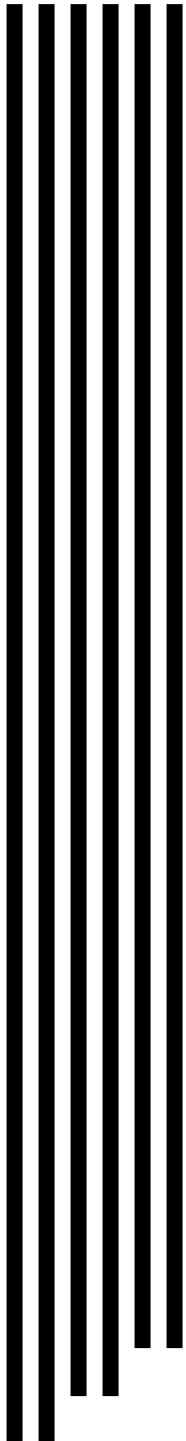




CIMAT



Herramientas para el reconocimiento de radiofuentes extendidas en grandes rastreos del cielo

TESIS

Que para obtener el grado de
Maestro en Ciencias con Especialidad en
Computación y Matemáticas Industriales

Presenta
Martín Silerio Vázquez

Directores de Tesis
Dr. Johan Jozef Lode Van Horebeek
Dr. Heinz Joachim Andernach
Dr. Carlos Alanías Rodríguez Rico

Guanajuato, Guanajuato. 9 de Agosto de 2012.

Integrantes del jurado

Presidente: Dr. Rogelio Ramos Quiroga

Secretario: Dr. Rogelio Hasimoto Beltrán

Vocal: Dr. Carlos Alanías Rodríguez Rico

Lector especial: Dr. Johan Jozef Lode Van Horebeek

Lector especial: Dr. Heinz Joachim Andernach

Asesores:

Dr. Johan Jozef Lode Van Horebeek

Dr. Carlos Alanías Rodríguez Rico

Dr. Heinz Joachim Andernach

Sustentante:

Martín Silerio Vázquez

*A mis papás,
Nely y Nicho.*

Agradecimientos

A mis tres asesores por la excelente dirección de este trabajo. Al Dr. Johan Van Horebeek por su visión, su entusiasmo, su interminable paciencia y por haber hecho mucho más de lo que le correspondía como asesor y tutor académico. A los Dres. Heinz Andernach y Carlos Rodríguez Rico por todo el tiempo dedicado y los conocimientos compartidos, por su completa disposición para ayudar y por haberme mostrado un poco de la gran belleza que posee la ciencia que los apasiona.

A mis sinodales, los Dres. Rogelio Hasimoto Beltrán y Rogelio Ramos Quiroga por sus valiosas observaciones y recomendaciones.

A mis papás, mis hermanos y mi pequeña sobrina por su amor y apoyo incondicional. Gracias a ellos he llegado hasta aquí.

A mis profesores, amigos y compañeros por compartir sus conocimientos a lo largo de la maestría. Un agradecimiento muy especial a dos de mis mejores amigos en la vida, Guillermo y Carmen, quienes me brindaron apoyo total e incondicional durante los momentos más difíciles de esta etapa.

A los coordinadores del posgrado, Arturo Hernández Aguirre, Mariano Rivera Meráz y Rogelio Hasimoto Beltrán por su ayuda y asesoría durante mi estancia en el CIMAT. Gracias por no dejar de creer en mí.

Al CONACyT, por haberme proporcionado el apoyo económico necesario para la obtención de este grado.

Contenido

1 Panorama general	1
1.1 Galaxias y radiogalaxias	1
1.2 Rastreos (<i>surveys</i>) y catálogos astronómicos	5
1.3 El problema del reconocimiento de objetos	6
1.3.1 Las radiogalaxias más grandes conocidas	7
1.4 Ejemplos de métodos automatizados para reconocer objetos	8
1.4.1 Trabajo de Proctor (2006)	8
1.4.2 Trabajo de Kirshner <i>et al.</i> (2002)	12
2 Problema de interés	17
2.1 Definición y descripción del problema	17
2.2 Bases de datos a utilizarse	19
2.2.1 Atlas de imágenes	20
2.2.2 Catálogo de radiofuentes	21
2.3 Objetivos específicos	22
3 Método propuesto	25
3.1 Construcción de una base de datos maestra de interés	25
3.2 Lista de características	32
3.2.1 Longitud angular	32
3.2.2 Índice de asimetría (asimetría I)	36

3.2.3	Índice de asimetría sobre la imagen binaria (asimetría B)	37
3.2.4	Índice de asimetría sobre la imagen binaria usando apertura (asimetría A)	37
3.2.5	Índice de gordura	38
3.2.6	Índice de delgadez central	39
3.2.7	Índice de rectitud	39
3.2.8	Intensidad promedio	40
3.2.9	Índice de contraste de intensidad	41
3.3	Poder discriminatorio de las características	41
3.4	Clasificadores basados en las características	42
3.5	Radiofuentes extendidas con potencial de ser RGG no registradas	48
A	Sistemas de coordenadas astronómicas	51
A.1	Sistema de coordenadas ecuatoriales	51
A.2	Sistema de coordenadas Galácticas	53
B	Descripción del software	55
B.1	Barra de menú	57
B.1.1	Archivo	57
B.1.2	Catálogo	58
B.1.3	Contornos	59
B.1.4	Tensores	60
B.1.5	Operadores morfológicos	61
B.1.6	<i>Zoom</i>	62
B.2	Ruido	63
B.2.1	Radiofuentes extendidas	65
B.2.2	Simetrías	67
Bibliografía		69

Capítulo 1

Panorama general

En este capítulo introducimos los conceptos necesarios para el planteamiento y desarrollo de esta tesis. En la sección 1.1 se describen los objetos astronómicos de interés: las radiogalaxias. En la sección 1.2 se describen las bases de datos disponibles para el manejo de las imágenes del cielo. En la sección 1.3 se define el problema general de detección de objetos astronómicos. Por último, en la sección 1.4 se resumen dos trabajos de aplicación de reconocimiento estadístico de patrones para resolver el problema de manera automática.

1.1 Galaxias y radiogalaxias

La astronomía es la ciencia que estudia los cuerpos celestes del Universo. Entre estos cuerpos se encuentran planetas y sus satélites, cometas, estrellas, así como los sistemas de estrellas, gas y polvo llamados galaxias. Una galaxia es un sistema masivo de estrellas, planetas, gas, polvo cósmico, etc. unido gravitacionalmente. De acuerdo a observaciones ópticas (es decir, lo que se observa en un telescopio común) las galaxias se han clasificado en distintos tipos: espirales, elípticas e irregulares.

En la actualidad es posible observar el universo en todas las bandas del espectro electromagnético. Estas bandas son radio, infrarrojo, óptico, ultravioleta, rayos x y rayos gamma. La frecuencia de observación determina el tipo de información obtenida del objeto. En la

Figura 1.1 observamos una galaxia vista en tres diferentes frecuencias. Las frecuencias de observación en radio o radiofrecuencias van desde algunos MHz hasta alrededor de los 300 GHz.

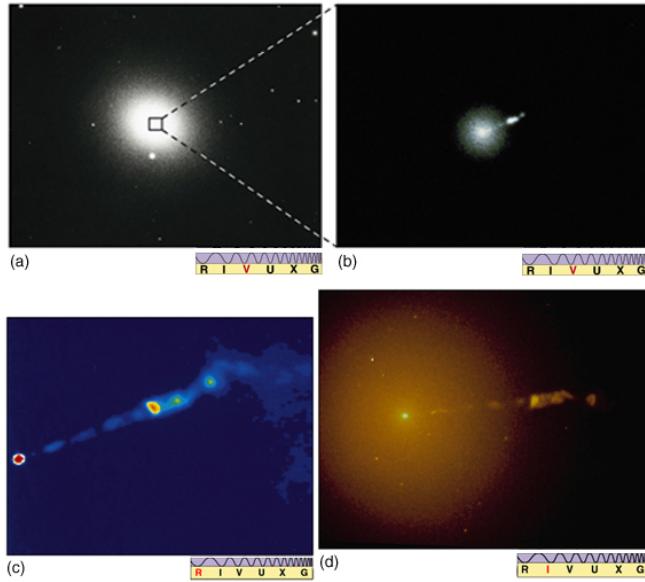


Figura 1.1: Galaxia elíptica gigante M87 observada en diferentes longitudes de onda. (a) Observación en óptico obtenida con un tiempo de exposición grande. (b) Observación en óptico con un tiempo de exposición más pequeño y campo menor. (c) Observación en radio. (d) Observación en infrarojo (Chaisson & McMillan 2001).

Al hacer observaciones en radio es más fácil observar objetos muy alejados que con el óptico, pero resulta más difícil estimar la distancia a la que se encuentran dichos objetos y por lo tanto su tamaño absoluto. Las observaciones de las galaxias en las diferentes bandas del espectro han permitido observar otras características y con base en estas se ha extendido la clasificación. En la Tabla 1.2 se enlistan las características que permiten definir la diferencia entre galaxias activas y normales. Dicha tabla también muestra que hay una subclasificación dentro de las galaxias activas. En particular serán de nuestro interés las galaxias que emiten en frecuencias de radio de manera significativa y tienen una morfología específica, conocidas como radiogalaxias. En la Figura 1.3 observamos un ejemplo de radiogalaxia, Cygnus A. Esta galaxia es un ejemplo prototípico de la morfología que tiene las radiogalaxias.

1.1. Galaxias y radiogalaxias

Differences between active galaxy types and normal galaxies.

Galaxy Type	Active Nuclei	Emission Lines		X-rays	Excess of		Strong Radio	Jets	Variable	Radio loud
		Narrow	Broad		UV	Far-IR				
Normal	no	weak	none	weak	none	none	none	none	no	no
Starburst	no	yes	no	some	no	yes	some	no	no	no
Seyfert I	yes	yes	yes	some	some	yes	few	no	yes	no
Seyfert II	yes	yes	no	some	some	yes	few	yes	yes	no
Quasar	yes	yes	yes	some	yes	yes	some	some	yes	10%
Blazar	yes	no	some	yes	yes	no	yes	yes	yes	yes
BL Lac	yes	no	none/faint	yes	yes	no	yes	yes	yes	yes
OVV	yes	no	stronger than BL Lac	yes	yes	no	yes	yes	yes	yes
Radio galaxy	yes	some	some	some	some	yes	yes	yes	yes	yes

Tabla 1.2: Tabla comparativa entre tipos de galaxias activas y tipos de galaxias normales (Wikipedia 2012).

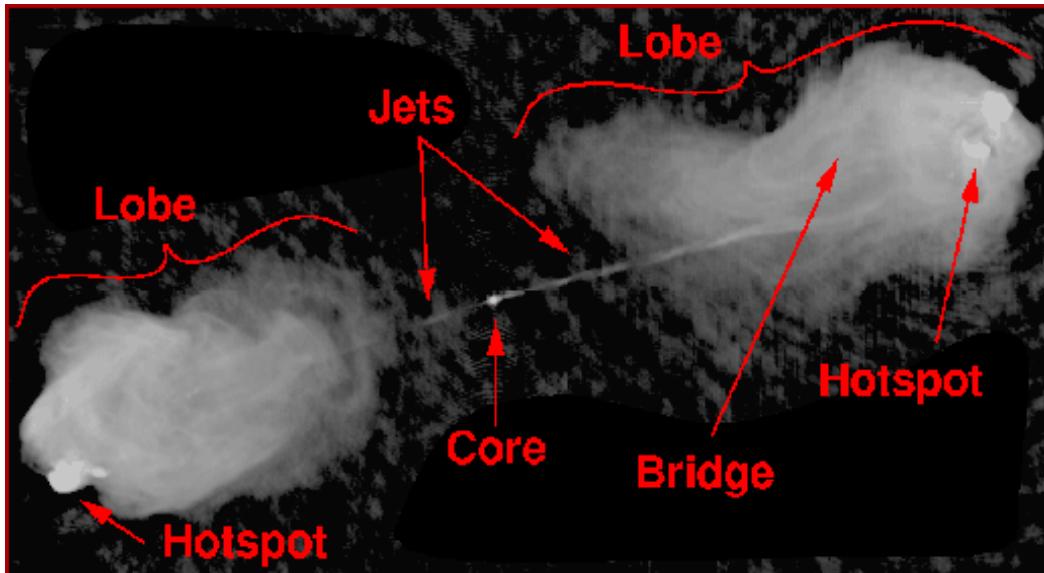


Figura 1.3: Galaxia Cygnus A observada en radio. Los niveles de gris indican diferentes brillos superficiales o una frecuencia de 5 GHz ($\lambda = 6$ cm): un tono más claro corresponde a un brillo más alto, y el negro corresponde a la ausencia de emisión en radio. (Leahy *et al.* 2000).

En una radiogalaxia se identifican principalmente tres partes: núcleo, chorros y lóbulos. En la figura 1.3 están señaladas estas partes con sus nombres en inglés (*cores*, *jets* y *lobes*, respectivamente).

El núcleo en radio coincide con el núcleo de la galaxia óptica y es donde un hoyo negro supermasivo (mayor a un millón de masas solares) ocasiona la formación de un disco de acreción (ver Figura 1.4) que a su vez da origen a dos chorros muy colimados en las dos direcciones perpendiculares a este disco de acreción. A lo largo de estos ejes propaga un plasma de protones y electrónes con velocidades relativistas (muy cercanas a la velocidad de la luz, $c = 3 \times 10^8$ m/s). La colimación de los chorros de partículas proviene de un campo magnético (por orden de 10^{-3} Gauss, y decreciendo con mayor distancia del núcleo) que permiten que las galaxias lleguen a grandes distancias de la galaxia óptica. La emisión en radio que registramos en la Tierra es debido al proceso sincrotrón, la radiación emitida por una partícula cargada en aceleración. Esta emisión en la imagen en radio se observa como dos segmentos colimados (no siempre rectos, ni siempre visibles de ambos lados) partiendo del núcleo, los cuales son llamados chorros o *jets*. Por último, en los puntos llamados *hot spots* donde los chorros se encuentran con el medio intergaláctico, las partículas (y por lo tanto su emisión registrada) se esparcen hacia todas direcciones, formando los denominados lóbulos de la radiogalaxia (ver por ejemplo (Miley 1980)).

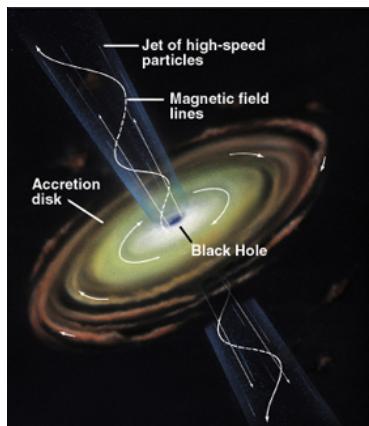


Figura 1.4: Modelo del núcleo de una radiogalaxia (Chaisson & McMillan 2001).

1.2 Rastreos (*surveys*) y catálogos astronómicos

Un rastreo astronómico es una serie de observaciones realizadas en cierta región del espectro electromagnético, el cual cubre grandes áreas contiguas del cielo. El objetivo de los rastreos astronómicos es crear una base de datos la cual pueda usarse como punto de partida para la selección y posterior observación de objetos de interés. Típicamente, cada posición del cielo es observada sólo por un intervalo pequeño de tiempo, y en muchos casos esto da una idea sobre si vale la pena dedicarles más tiempo de observación. En la Tabla 1.5 tenemos algunos ejemplos de rastreos astronómicos de mayor tamaño en óptico y en radio.

Nombre (sigla)	Tipo	Frecuencia	θ	Región ¹	Referencia
SDSS	Óptico	~600 THz	~ 1''	35% del cielo ($b > 30^\circ$)	(York 2000)
NVSS	Radio	1.4 GHz	45''	Hemisferio norte ($\delta > -40$)	(Condon <i>et al.</i> 1998)
SUMSS	Radio	843 MHz	45''	Hemisferio sur ($\delta < -30$)	(Bock <i>et al.</i> 1999)
FIRST	Radio	1.4 GHz	5''	Variable (30% del cielo)	(Becker <i>et al.</i> 1995)
WENSS	Radio	325 MHz	50''	Hemisferio norte ($\delta > +28.5$)	(de Ruiter 1998)
WISH	Radio	352 MHz	50''	Hemisferio sur ($-9 < \delta < -26$)	(de Breucks 2002)

Tabla 1.5: Ejemplos de rastreos astronómicos. θ es la resolución angular del rastreo.

Un catálogo astronómico es una lista de objetos astronómicos, muchas veces agrupados con base en ciertas características que tienen en común (morfología, tipo, etc.). En varios rastreos tomados en radio el catálogo está compuesto de radiofuentes discretas, que son objetos derivados de ajustes de gaussianas sobre la imagen en zonas donde hay emisiones claras de intensidad. Dichos ajustes respetan ciertas restricciones, por ejemplo, no pueden

¹Si el lector no está familiarizado con los sistemas de coordenadas astronómicas se le recomienda revisar el Apéndice A.

exceder cierto tamaño relacionado con la resolución angular.

Una radiofuent es cualquier objeto del cielo del que se percibe emisión de radio. En el contexto de los catálogos de rastreos en radio se llaman radiofuentes discretas a cada uno de los ajustes gaussianos antes mencionados. A una radiofuent extendida se le pudieron haber ajustado una o varias gaussianas. Cuando a una radiofuent le corresponden dos o más radiofuentes discretas del catálogo decimos que la radiofuent está resuelta en el catálogo. A las radiofuentes que les corresponde sólo una radiofuent discreta del catálogo les llamamos radiofuentes no resueltas cuando el tamaño del ajuste gaussiano no sobrepasa el haz de la antena o del interferómetro. En adelante nos referiremos con la palabra “radiofuent” a cualquier región del cielo donde se percibe emisión de radio y con “fuente” a los ajustes gaussianos de los catálogos.

1.3 El problema del reconocimiento de objetos

El problema consiste en usar las imágenes del rastreo y la información del catálogo para reconocer ciertos objetos de interés. Tradicionalmente la inspección visual juega un papel muy importante en esta tarea. Se apoya en varias herramientas computacionales como Aladin (Bonnarel *et al.* 2002), Fv (Pence & Chai 2009) y DS9 (Joye & Mandel 2003). Estos programas permiten visualizar las imágenes de distintas maneras: cambiando la paleta de colores, localizando picos o cambiando el contraste. También es posible realizar cálculos sobre la función de intensidad $I : \mathbb{R}^2 \rightarrow \mathbb{R}$ derivada de la imagen, por ejemplo las curvas de nivel.

Las curvas de nivel o contornos son una de las herramientas más usadas. Con ellas es posible localizar fácilmente objetos que sobresalen en extensión o intensidad. Aunque las curvas de contorno proporcionan menos información que la gráfica tridimensional de la sábana definida por I , son más cómodas de observarse y más fáciles de obtener. En la Figura 1.6 vemos un ejemplo de curvas de contorno de una imagen en radio graficadas sobre la imagen correspondiente en óptico.

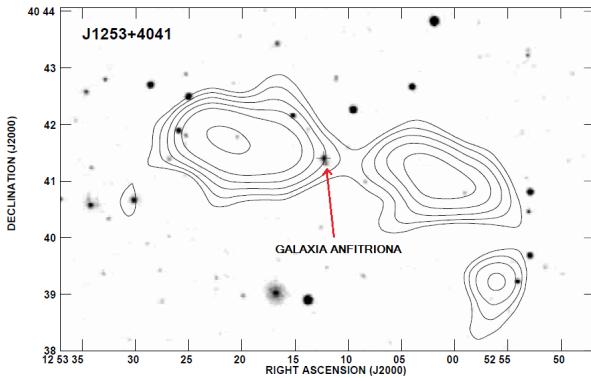


Figura 1.6: Imagen en radio (contornos) obtenida del NVSS sobre imagen óptica (escala de grises) obtenida del DSS (Machalski *et al.* 2001).

Uno de los factores más importantes para la identificación de radiogalaxias es lograr la localización de la llamada galaxia anfitriona en el espectro óptico. En la Figura 1.6 está señalada la galaxia anfitriona de dicho ejemplo. Mientras no haya evidencia de la existencia de la galaxia anfitriona el objeto observado en radio no es más que un candidato a ser una radiogalaxia. Al estudiar la galaxia anfitriona es posible determinar la distancia a la que se encuentra y, por lo tanto, su tamaño absoluto.

La inspección visual de los rastreos es muy tardada debido a la magnitud de la información con la que se cuenta. Para facilitar y acelerar el proceso se han desarrollado técnicas semi-automáticas. A continuación se da un reseña de dos trabajos con resultados importantes obtenidos mediante inspección visual. En las secciones siguientes se resumen los artículos publicados (Proctor 2006) y (Kirshner *et al.* 2002) que usan técnicas automatizadas para la detección de radiogalaxias y que serán relevantes para la propuesta de esta tesis.

1.3.1 Las radiogalaxias más grandes conocidas

Los trabajos de (Willis *et al.* 1974) y (Machalski *et al.* 2008) son excelentes ejemplos de los alcances de la inspección visual. En ellos se trabaja sobre las tres radiogalaxias más grandes identificadas hasta ahora.

En (Willis *et al.* 1974) se hace un estudio profundo de las características de 3C236 y DA240, las dos radiogalaxias más grandes identificadas hasta entonces. Sus tamaños absolutos de 5.7 Mpc y 2 Mpc no tenían precedentes. Se comparan sus propiedades con las de otras radiogalaxias y se discute bastante acerca del patrón de evolución de estas galaxias.

Es necesario aclarar que los tamaños reportados dependen del valor del parámetro de Hubble (H_0). Este valor está relacionado con la velocidad de expansión del universo. A través de los años las estimaciones de este parámetros han mejorado. En (Willis *et al.* 1974) se consideró $H_0 = 50 \text{ km/s/Mpc}$, mientras hoy en día suele usarse $H_0 = 75 \text{ km/s/Mpc}$.

En (Machalski *et al.* 2008) se reporta el descubrimiento de la radiogalaxia más grande identificada hasta ahora: la J1420-0545. Su tamaño proyectado absoluto es 4.69 Mpc usando $H_0 = 71 \text{ km/s/Mpc}$. Los tamaños proyectados de las galaxias 3C236 y DA240 con $H_0 = 71 \text{ km/s/Mpc}$ son respectivamente 4 Mpc y 1.4 Mpc.

1.4 Ejemplos de métodos automatizados para reconocer objetos

1.4.1 Trabajo de Proctor (2006)

En (Proctor 2006) se utilizan algoritmos del área de reconocimiento estadístico de patrones para clasificar ciertos objetos del rastreo astronómico FIRST (Faint Images of the Radio Sky at Twenty Centimeters (Becker *et al.* 1995)). La frecuencia de observación del FIRST es 1.4 GHz y su resolución angular es igual a $5''$. Existe una versión previa del artículo de 2006 (Proctor 2003) y un artículo reciente muy relacionado (Proctor 2011). No nos adentraremos mucho en los detalles que hacen diferentes a los artículos entre sí. Sin embargo, los tres artículos comparten las ideas que sirvieron de motivación para la realización de esta tesis.

El problema bajo consideración es la detección de cierto tipo de radiogalaxias con tres radiofuentes discretas del tipo llamado *bent double*; consisten en dos chorros (*jets*) o lóbulos (*lobes*) extendidos hacia lados opuestos de un núcleo central.

1.4. Ejemplos de métodos automatizados para reconocer objetos

Como primer paso, se extraen del catálogo grupos triples de radiofuentes discretas que están separadas entre sí a lo más $0.96'$. Se busca identificar cuáles grupos triples son *bent doubles* y cuáles no.

En la Figura 1.7 podemos observar varios ejemplos de grupos triples. En el primer renglón de la imagen vemos ejemplos típicos de radiogalaxias compuestas por tres fuentes. En estos los chorros están perfectamente alineados. En el segundo renglón vemos ejemplos donde los chorros tienen varios grados de desviación con respecto a una recta (emisión hacia lados opuestos). Esta clase de objetos son los denominados *bent doubles*. En el tercer renglón tenemos ejemplos de grupos de tres fuentes que no son *bent doubles*. Por último, en el cuarto renglón tenemos ejemplos de grupos ambiguos. Esto es, debido a la poca resolución o al ruido en la imagen, resulta imposible determinar visualmente si es o no un *bent double*.

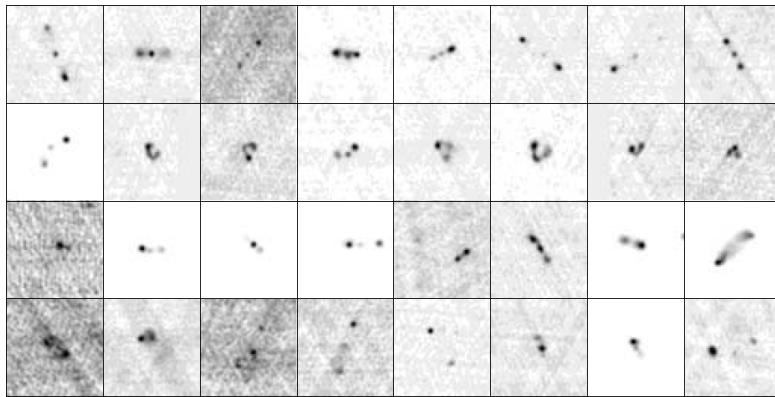


Figura 1.7: Ejemplos de grupos con tres fuentes (Proctor 2006).

Una muestra aleatoria de 2823 grupos fue seleccionada del total de grupos triples, los cuales son alrededor de 15000. Cada triple fue visualmente asignado a una de las clases *bent double*, no *bent double* o caso ambiguo: 147 *bent doubles*, 1395 no *bent doubles* y 1281 casos ambiguos. Se tomó como conjunto de entrenamiento los triples catalogados como *bent double* o no *bent double* restringidos a que su SNR (*Signal to Noise Ratio*) es mayor a 8.5. Se obtuvieron 115 *bent doubles* y 939 no *bent doubles*. Como conjunto de prueba se usaron los 2823 triples seleccionados en la muestra aleatoria.

Para cada grupo triple se propone medir 21 características. Estas se calculan usando los parámetros del ajuste gaussiano listados en el catálogo. Cada triple forma un triángulo cuyos vértices son los centros de las fuentes del grupo. En la Figura 1.8 observamos cómo se etiquetan los lados de dichos triángulos. Se determina cuál es el lado de mayor tamaño, y dicha distancia se denomina d_{\max} . La fuente opuesta al lado más grande se asume como el núcleo propuesto del posible *bent double*. Las medidas de los dos lados restantes se identifican por d_{mid} y d_{\min} según corresponda.

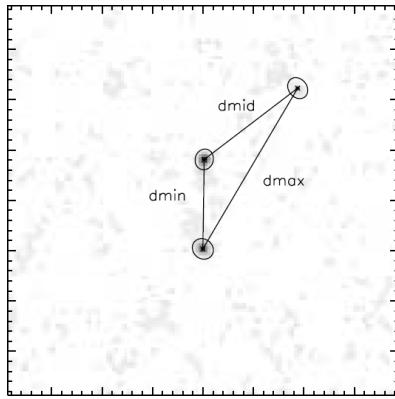


Figura 1.8: Construcción del triángulo asociado a un grupo triple (Proctor 2006). d_{\max} es la longitud del lado más grande del triángulo, d_{\min} es la longitud del lado más pequeño y d_{mid} es la longitud del lado restante.

En la Tabla 1.9 enumeraremos las 21 características propuestas. El tamaño de la silueta mencionado en la tabla es calculado para cada una de las fuentes. Está definido como la cantidad de pixeles que están por encima de cierto umbral bajo el modelo ajustado para dicha fuente. Es decir, el modelo ajustado asocia a la posición (x, y) la densidad de flujo

$$S(x, y) = S_p \exp \left[- \left(\frac{x^2}{2\sigma_x^2} + \frac{y^2}{2\sigma_y^2} \right) \right],$$

donde S_p , σ_x y σ_y son los parámetros de la gaussiana ajustada a la fuente en cuestión. T_{ss} en la Tabla 1.9 es la suma de los tamaños de silueta de las tres fuentes del triple y R_{ss} es la razón entre los tamaños de silueta de las fuentes que no se etiquetan como núcleo. La razón se toma como el tamaño de silueta menor entre el mayor, por lo que $R_{ss} \leq 1$.

1.4. Ejemplos de métodos automatizados para reconocer objetos

Número	Descripción
1	Longitud del lado intermedio del triángulo (d_{mid}).
2	Razón entre la longitud del lado menor y la longitud del lado intermedio (d_{min}/d_{mid}).
3	Razón entre la suma de las longitudes de los lados intermedio y menor, y la longitud del lado mayor ($(d_{mid} + d_{min})/d_{max}$).
4	Razón entre los tamaños de silueta de las fuentes que no se etiquetan como núcleo (del menor entre el mayor) (R_{ss}).
5	Suma de los tamaños de silueta de las tres fuentes (T_{ss}).
6	Razón entre la distancia del punto medio del lado más pequeño y la fuente opuesta, y la longitud del lado menor.
7	Valor absoluto del coseno del ángulo entre el eje mayor de la fuente opuesta al lado intermedio y el lado menor.
8	Razón entre la raíz cuadrada del tamaño de silueta del núcleo y la longitud de su lado opuesto.
9	Razón entre la raíz cuadrada del tamaño de silueta de la fuente opuesta al lado intermedio y la longitud del lado intermedio.
10	Tamaño de silueta más grande de entre las fuentes que no se etiquetan como núcleo.
11	Valor absoluto del coseno del ángulo entre el eje mayor de la fuente opuesta al lado menor y el lado intermedio.
12	Flujo integrado del núcleo.
13	Máximo de los flujos integrados de las tres fuentes.
14	Flujo integrado de la fuente opuesta al lado intermedio.
15	Razón entre los flujos integrados de las fuentes que no se etiquetan como núcleo (del menor entre el mayor).
16	Razón entre el flujo pico y el flujo integrado de la segunda fuente más intensa.
17	Razón entre el flujo pico y el flujo integrado de la fuente menos intensa.
18	Razón entre el flujo integrado de la fuente más intensa y la suma de los flujos integrados de las tres componentes.
19	Razón entre el flujo integrado de la fuente menos intensa y la suma de los flujos integrados de las tres componentes.
20	Razón entre el flujo integrado del núcleo y la suma de los flujos integrados de las tres componentes.
21	Razón entre el flujo integrado de la fuente opuesta al lado menor y la suma de los flujos integrados de las tres componentes.

Tabla 1.9: Lista de características a medir sobre los triples propuesta en (Proctor 2006).

Las características en la Tabla 1.9 son usadas para ajustar clasificadores usando el conjunto de entrenamiento definido anteriormente. Se usan árboles de decisión y redes neuronales. Después de varios análisis, se obtiene que los clasificadores propuestos logran obtener un error de clasificación menor a 20% usando entre 15% y 30% del conjunto de entrenamiento. Esto indica que el desempeño del método propuesto es considerablemente bueno.

1.4.2 Trabajo de Kirshner *et al.* (2002)

En (Kirshner *et al.* 2002) se aborda el mismo problema que en (Proctor 2006): la clasificación de grupos triples del FIRST en *bent doubles* y no *bent doubles*. Se trabaja con un pequeño conjunto de entrenamiento que consiste en 128 ejemplos de *bent doubles* y 22 ejemplos de no *bent doubles*. Dichos ejemplos fueron tomados de entre cerca de 15,000 grupos triples que se derivaron del catálogo del FIRST. Cada ejemplo fue analizado por tres astrónomos y fue clasificado como *bent double* si al menos dos de los tres astrónomos estuvieron de acuerdo en que debe clasificarse así. En la Figura 1.10 tenemos algunos ejemplos de estas clasificaciones.

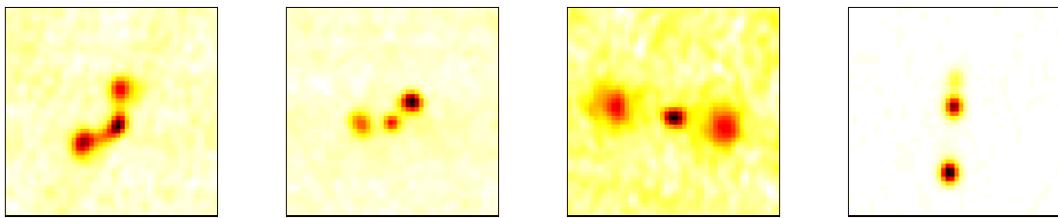


Figura 1.10: Ejemplos de triples clasificados como *bent doubles* (las dos imágenes a la izquierda) y no *bent doubles* (las dos imágenes a la derecha) (Kirshner *et al.* 2002).

La idea es determinar para cada triple qué tan probable es que sea un *bent double*. Eso requiere identificar cuál de las fuentes es el núcleo, cuál el lóbulo 1 y cuál el lóbulo 2. A diferencia de (Proctor 2006), aquí se consideran todos los posibles escenarios de identificación, es decir, cada fuente del triple puede ser el núcleo, el lóbulo 1 o el lóbulo 2. Denotamos un triple mediante $\mathcal{C} = \{c_1, c_2, c_3\}$. Sea S_3 el conjunto de permutaciones de tres objetos y Π una variable aleatoria que toma valores en S_3 . Cada elemento de S_3 (o posible valor de Π)

corresponde a un ordenamiento de \mathcal{C} con la siguiente interpretación: si escribimos

$$\Pi(\mathcal{C}) = (\Pi(\mathcal{C})_1, \Pi(\mathcal{C})_2, \Pi(\mathcal{C})_3)$$

entonces $\Pi(\mathcal{C})_1$ (el primer elemento en el orden) es la fuente que se toma como el núcleo, $\Pi(\mathcal{C})_2$ la fuente que se toma como el lóbulo 1 y $\Pi(\mathcal{C})_3$ la fuente que se toma como el lóbulo 2. La Figura 1.11 muestra gráficamente los seis escenarios posibles.

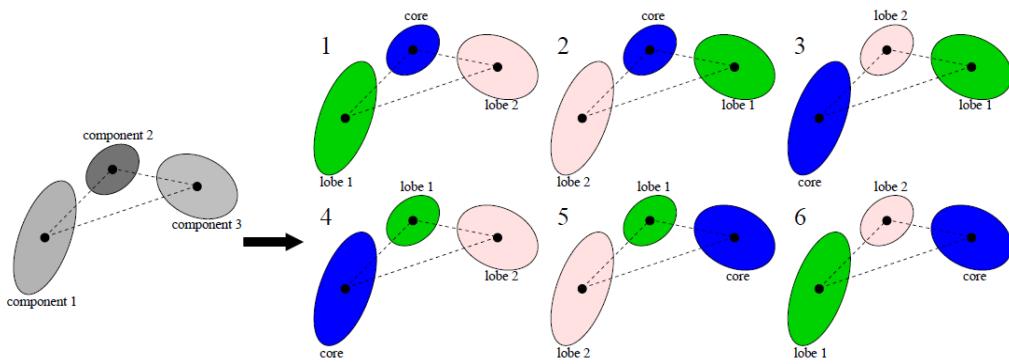


Figura 1.11: Ordenamientos posibles para un candidato a *bent double*. Los órdenes 1 y 2 son las opciones más viables (Kirshner *et al.* 2002).

Para cada triple \mathcal{C} se calculó un vector de características $F(\mathcal{C})$ y se buscó la probabilidad de observar $F(\mathcal{C})$ suponiendo que \mathcal{C} es un *bent double*: $P(\text{observar } F(\mathcal{C}) \mid \mathcal{C} \text{ es bent double})$. Esto permite clasificar un triple \mathcal{C} como *bent double* o no dependiendo de si esta probabilidad es mayor que cierta cota inferior elegida por el usuario y permite ordenar triples de acuerdo a dicha probabilidad. Sabemos:

$$P(\text{observar } F(\mathcal{C}) \mid \mathcal{C} \text{ es bent double}) := \sum_{\pi \in S_3} P(F(\mathcal{C}) \mid \Pi = \pi) P(\Pi = \pi).$$

Para definir $F(\mathcal{C})$ se propone medir tres características para cada fuente del triple. Por lo tanto, dado el ordenamiento $\Pi = \pi$ de \mathcal{C} tendremos nueve características:

$$f_{mj}(\Pi(\mathcal{C})), m = 1, \dots, 3, j = 1, \dots, 3$$

y considerando el triángulo formado por las tres fuentes (ver Figura 1.12):

- $f_{1j}(\Pi(\mathcal{C}))$: Logaritmo del ángulo formado en el vértice $\Pi(\mathcal{C})_j$.
- $f_{2j}(\Pi(\mathcal{C}))$: Logaritmo de la razón entre las longitudes de lados que coinciden en el vértice $\Pi(\mathcal{C})_j$.
- $f_{3j}(\Pi(\mathcal{C}))$: Logaritmo de la razón de las intensidades de las fuentes que no son $\Pi(\mathcal{C})_j$.

Los autores suponen independencia condicional, o sea $P(f_{11}(\Pi(\mathcal{C})), \dots, f_{33}(\Pi(\mathcal{C})) \mid \Pi = \pi)$ es igual al producto $\prod_{\substack{m=1, \dots, 3 \\ j=1, \dots, 3}} P(f_{mj}(\Pi(\mathcal{C})) \mid \Pi = \pi)$, y cada f_{mj} sigue una distribución normal con promedio y varianza por estimar.

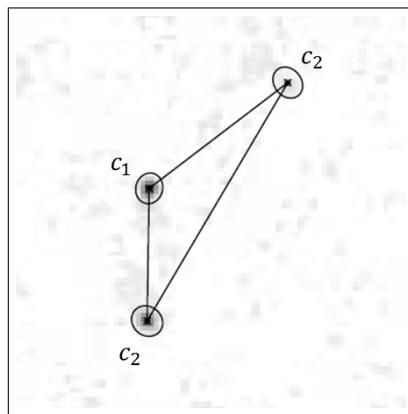


Figura 1.12: Triángulo formado por las fuentes del triple usado para el cálculo de las características (Proctor 2006).

Para estimar los valores de los parámetros de las gaussianas, se considera el valor de Π para cada triple como una variable latente y se usa el algoritmo EM.

Se encontró que el algoritmo EM converge relativamente rápido (entre 20 y 30 iteraciones). Después de la estimación de los parámetros es posible calcular la probabilidad de que nuevos objetos sean *bent doubles*. En las Figuras 1.13 y 1.14 observamos triples con probabilidades altas y bajas respectivamente.

Para probar la eficiencia de la clasificación se usó validación cruzada obteniendo un área bajo la curva ROC (*Receiver Operating Characteristic*) igual a 0.9336, lo que indica que se hace muy bien la detección de *bent doubles*.

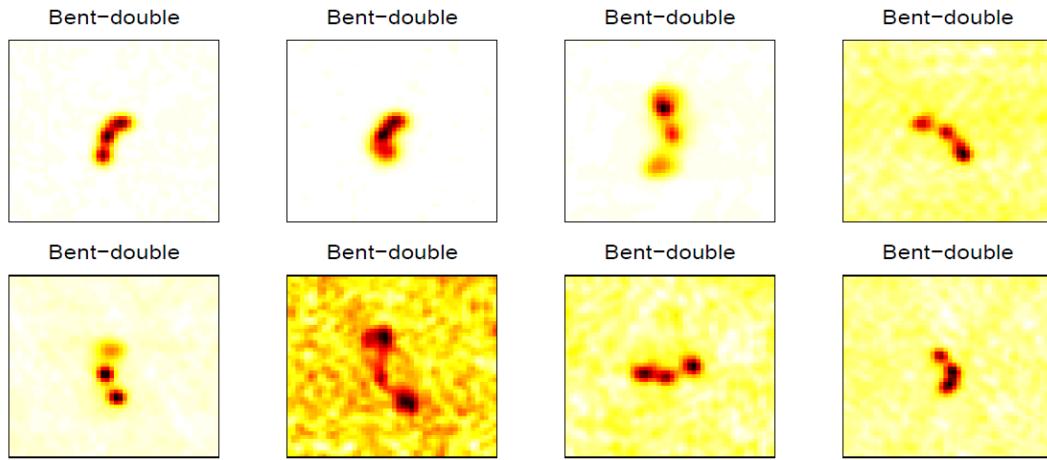


Figura 1.13: Objetos con probabilidad alta de ser *bent double*. El texto en la parte superior de cada imagen indica la clasificación dada por los astrónomos.

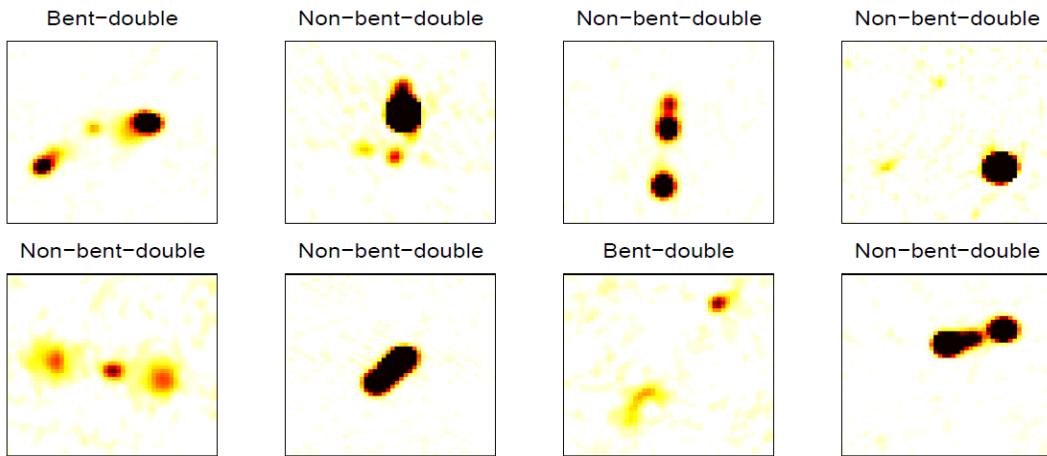


Figura 1.14: Objetos con probabilidad baja de ser *bent double*. El texto en la parte superior de cada imagen indica la clasificación dada por los astrónomos.

Capítulo 2

Problema de interés

En la sección 2.1 se define el problema específico que se aborda en esta tesis. En la sección 2.2 se describen a detalle las bases de datos sobre las que se trabajará. Por último, se enlistan los objetivos particulares del trabajo en la sección 2.3.

2.1 Definición y descripción del problema

Ya que el uso de técnicas de reconocimiento estadístico de patrones en astronomía es relativamente reciente, las posibilidades para hacer un aporte son muchísimas. Para esta tesis se buscó definir un problema acotado que pudiera concluirse en el tiempo con el que se contaba y que lograra hacer un aporte pequeño al área.

El objetivo principal de esta tesis es desarrollar herramientas computacionales para el reconocimiento de radiofuentes extendidas en el rastreo NVSS (NRAO VLA Sky Survey). Este rastreo se realizó con el observatorio NRAO VLA (National Radio Astronomy Observatory Very Large Array, ver Figura 2.1). Se tiene un interés particular en radiogalaxias gigantes. Se considera gigante una galaxia si su tamaño sobrepasa un megapársec (Mpc). El Mpc es una unidad de longitud usada en astronomía equivalente a 3261600 años luz (3.0857×10^{19} km). Una radiogalaxia gigante es al menos 30 veces más grande que nuestra galaxia.

En adelante usaremos la sigla RGG para referirnos a las radiogalaxias gigantes. Es



Figura 2.1: Antenas que conforman el NRAO VLA. Están colocadas en la configuración “D”, la cual fue usada para realizar el NVSS.

importante señalar que el prefijo “radio” no puede omitirse, ya que una “galaxia gigante” es una galaxia de dimensiones grandes al observarse en el óptico. Una galaxia gigante puede o no tener radiogalaxia asociada, la cual por lo general no es una RGG.

A diferencia de (Proctor 2006) y (Kirshner *et al.* 2002), en esta tesis se trabaja con NVSS y no con FIRST. Otra diferencia es que la distribución bidimensional de la intensidad de las imágenes será tomada en cuenta *a priori*, en conjunto con las radiofuentes discretas enlistadas en el catálogo, mientras en (Proctor 2006) y (Kirshner *et al.* 2002) solo se utilizó *a posteriori* para verificar las estructuras en radio.

Los rastreos FIRST y NVSS fueron desarrollados en el mismo observatorio, en la misma frecuencia de observación. Ambos proyectos comenzaron en 1993; el NVSS está concluido y el FIRST sigue en desarrollo. Las dos diferencias principales entre el FIRST y el NVSS son el porcentaje del cielo cubierto y la resolución angular. El NVSS cubre el 82% del cielo, mientras que el FIRST por el momento solo cubre el 30%. La resolución angular del FIRST es de $5''$ y la del NVSS es $45''$. Como consecuencia de esto un pixel de una imagen del NVSS tiene como contraparte en el FIRST una subimagen de tamaño 9×9 pixeles. En la Figura 2.2 podemos ver una comparación.

Ya que se tiene interés en radiofuentes extendidas, el NVSS resulta una mejor opción que el FIRST. La razón es que al aumentar la resolución angular es posible observar más detalles,

2.2. Bases de datos a utilizarse

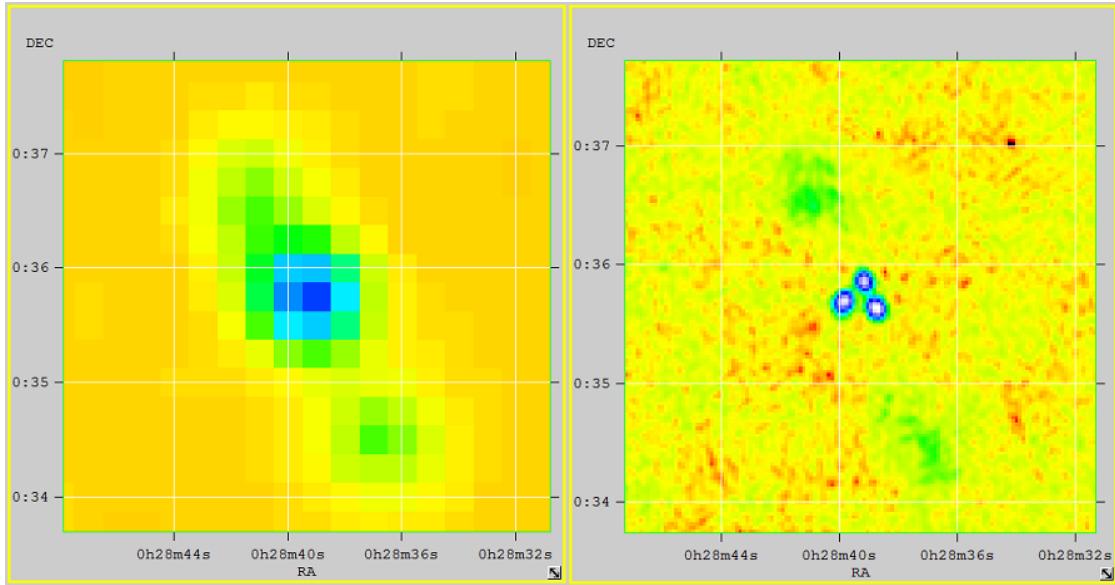


Figura 2.2: Comparación de la resolución de NVSS vs. FIRST del mismo arco de $4'' \times 4''$ del cielo. Imagen generada con (Pence & Chai 2009).

pero dejan de percibirse estructuras grandes. En la Figura 2.3 se puede ver un ejemplo de lo anterior. La radiofuentă captada por el NVSS con contornos en rojo no se percibe en la observación del FIRST (contornos azules). Se observan más detalles, pero es muy difícil notar la relación entre las radiofuentes con contornos en azul contenidas en el contorno rojo. Al igual que en la Figura 2.2 la conexión en intensidad entre los lóbulos norte y sur de la radiogalaxia en NVSS (izquierda) se pierden en FIRST (derecha).

2.2 Bases de datos a utilizarse

Como ya se mencionó, se trabajará con el rastreo NVSS. Este rastreo está tomado en la radiofrecuencia 1.4 GHz y cubre toda la parte norte del cielo a partir de la declinación -40° . En (Condon *et al.* 1998) se encuentra una descripción detallada del rastreo.

Los principales productos del NVSS son:

- Un conjunto de 2326 imágenes, cada una cubriendo un área de $4^\circ \times 4^\circ$. Cada una tiene

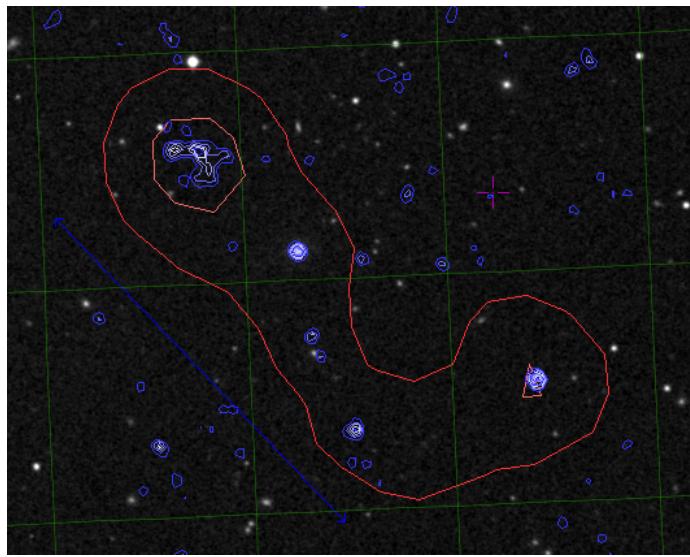


Figura 2.3: Imagen en óptico en blanco y negro. (a) Los contornos en azul corresponden a la observación en radio del FIRST. (b) Los contornos en rojo corresponden a la observación en radio del NVSS.

tres canales: Stokes I, Q y U. Cada imagen fue construida con poco más 100 imágenes pequeñas, correspondientes a cada una de las 217446 exposiciones que fueron necesarios para completar el rastreo. Estas últimas imágenes también están disponibles.

- Un catálogo de fuentes discretas detectadas por ajustes de gaussianas en el conjunto de imágenes; en total son más de 1.8 millones de fuentes (Ver Subsección 2.2.2).

En este trabajo utilizaremos la colección completa de imágenes, así como todo el catálogo de fuentes. Es común en astronomía usar el término “atlas” para referirse a las colecciones de imágenes astronómicas. En adelante usaremos dicho término.

2.2.1 Atlas de imágenes

El atlas del NVSS se compone de 2326 imágenes, cada una de 1024 por 1024 pixeles, más una banda superior de diez pixeles de alto para incluir un *copyright*. Las imágenes vienen en un formato especial, conocido como FITS (*Flexible Image Transport System*). Dicho formato es

el más utilizado en los archivos usados en astronomía, ya que permite integrar información adicional útil, como las coordenadas de la zona que abarca la imagen.

Para un fácil manejo de las imágenes cada una tiene un nombre que permite identificarla sin ambigüedad, el cual fue generado a partir de las coordenadas ecuatoriales del centro de dicha imagen. Por ejemplo, la imagen cuyo centro está dado por $(\alpha, \delta) = (13^h 48^m 0^s, +40^\circ)$ tendrá por nombre I1348P40; la letra que precede la declinación denota el signo de δ , P si $\delta \geq 0$ y M si $\delta < 0$.

En la Figura 2.4 observamos la distribución de los centros de las 2326 imágenes del NVSS. Cerca del polo norte hay menos imágenes para una declinación fija. Esto es claro, al imaginarse cómo se ven los meridianos en un globo terráqueo al aproximarse a los polos. Todas las imágenes tienen el mismo tamaño ($4^\circ \times 4^\circ$), donde 4° en dirección de ascensión recta corresponden a $(4/15)/\cos \delta$ horas de ascensión recta. Por tal razón se requieren menos imágenes para cubrir bandas más cercanas al polo norte. También es importante aclarar que cada imagen se traslape con sus imágenes vecinas y entre más se acerca la declinación a 90° , mayor es el traslape.

El atlas de imágenes en su versión más completa contiene tres canales para cada imagen (I, Q y U). Para los fines de este trabajo será suficiente analizar el canal I, la intensidad de la emisión de la onda de radio en cada zona. Los canales Q y U contienen información sobre la polarización lineal de dicha radiación.

2.2.2 Catálogo de radiofuentes

El catálogo es una tabla que contiene información derivada de las imágenes del NVSS (Condon *et al.* 1998). Las entradas consisten en un ajuste con una función gaussiana bivariada en los picos de la intensidad de la imagen. El proceso para construirlo fue completamente automatizado, por eso no hay garantía de que cada gaussiana corresponda a algún objeto físico. Es posible que se haya hecho un ajuste sobre un pico de intensidad producido por puro ruido en la imagen, por una réplica de una fuente vecina muy fuerte o por un efecto de *aliasing*. El *aliasing* es un fenómeno que causa que señales continuas distintas se vuelvan

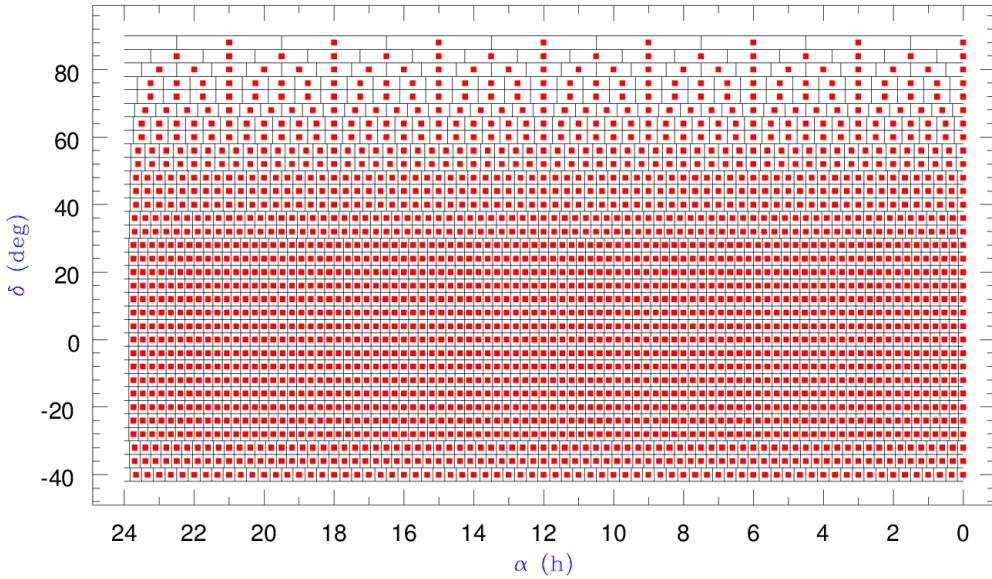


Figura 2.4: Partición de la parte del cielo que cubre el NVSS. Los rectángulos representan cada una de las 2326 imágenes. El punto rojo dentro de cada rectángulo indica la posición del centro de la imagen correspondiente.

iguales después del muestreo digital.

2.3 Objetivos específicos

En el problema general podemos distinguir los siguientes objetivos específicos:

- Desarrollar un ambiente de software que permita visualizar, procesar e interactuar con los productos del NVSS.
- Construir una lista de objetos de donde se seleccionarán los candidatos.
- Proponer varias características que sean informativas para determinar si un candidato es una RGG.
- Validar el potencial de las características usando galaxias extendidas verificadas.

2.3. Objetivos específicos

Describir las características de una radiogalaxia en general es un problema complejo. Las observaciones indican que las radiogalaxias tienen una forma típica, sin embargo, esta forma tiene tres dimensiones. En la Figura 1.4 vemos un modelo de dicha forma. Como observamos solo una proyección de dos dimensiones de esa forma típica, una radiogalaxia puede tener formas variadas en los rastreos.

Así como en (Proctor 2006) y (Kirshner *et al.* 2002) el estudio se centró en una forma especial de radiogalaxias (*bent doubles*), en este trabajo nos limitaremos a buscar identificar galaxias en las que se puedan identificar el núcleo y lóbulos, o solamente los lóbulos. Estas galaxias se parecerán a la que ve el observador B en la Figura 2.5. La galaxia sería muy difícil de detectar desde el ángulo del observador A.

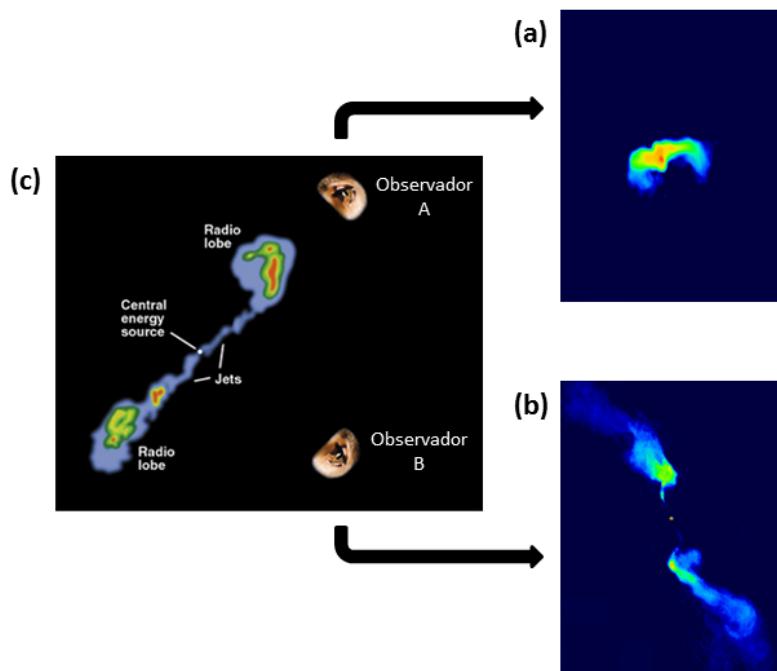


Figura 2.5: (c) Modelo de una radiogalaxia observada desde dos direcciones diferentes (Chaisson & McMillan 2001). (a) Probable observación de frente a lo largo del eje principal (Taylor 2006). (b) Probable observación de canto (Taylor 2006).

En la Figura 2.6 observamos tres ejemplos de RGG. En la columna izquierda vemos las imágenes extraídas del artículo donde fueron reportadas y en la columna derecha vemos

cómo se ven estas galaxias en las imágenes del NVSS.

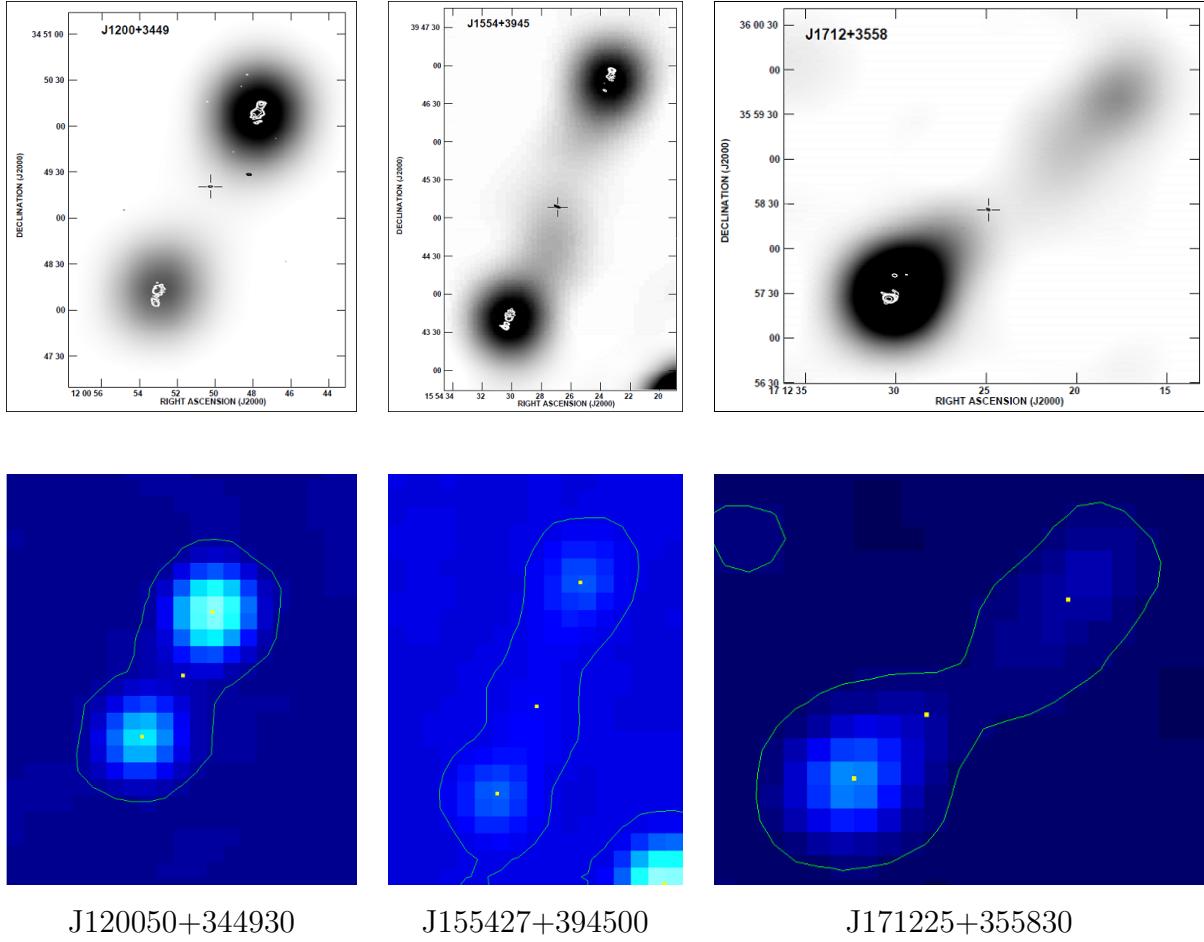


Figura 2.6: Ejemplo de RGG. Imágenes en (Machalski *et al.* 2001) vs. imágenes NVSS extraídas en esta tesis. Los contornos sobreimpuestos en blanco en las imágenes de la primera fila corresponden a observaciones con mayor resolución. En la ltima fila se encuentran las coordenadas de la posición de cada RGG.

Capítulo 3

Método propuesto

Explicamos primero en la sección 3.1 la construcción de una base de datos maestra de regiones que merecen un análisis más profundo. En la sección 3.2 se describen las características que serán usadas para un análisis discriminante posterior. En la última sección presentamos un clasificador usando estas características.

3.1 Construcción de una base de datos maestra de interés

Con una base de datos de imágenes tan grande, no es viable buscar objetos en cualquier región, barriendo completamente cada imagen. Por eso seleccionamos regiones prometedoras usando los criterios descritos a continuación.

Criterio utilizando curvas de nivel del rastreo

En general los astrónomos se guían por las curvas de nivel de la intensidad de las imágenes de un rastreo. Los objetos de interés están contenidos en regiones acotadas por curvas de niveles bajos pero aún significativas con respecto al nivel de ruido. Esto nos llevó al primer criterio de inclusión en la base de datos maestra: limitarnos sólamente a regiones en el interior de

curvas de cierto nivel bajo.

En astronomía comúnmente se trabaja con curvas de nivel arriba de dos o tres veces el nivel estimado de ruido. Las intensidades inferiores a tal nivel no se consideran suficientemente significativas. El nivel de ruido del NVSS se estima en $\sigma = 0.45 \text{ mJy/beam}$. A partir de aquí σ , se refiere a esta cantidad.

El manejo de las curvas de contorno no es sencillo, además de ser costoso computacionalmente. Por eso trabajamos con imágenes binarias U que surgen al aplicar un umbral, es decir

$$U(x) = \begin{cases} 1 & \text{si } I(x) \geq u \\ 0 & \text{si } I(x) < u \end{cases} \quad \text{para todo pixel } x.$$

La frontera de U define la curva de nivel u . Decidimos trabajar con curvas de nivel arriba de $u = 2.5\sigma$, para disminuir la probabilidad de excluir erróneamente regiones interesantes.

Sobre la imagen U definimos una “radiofuentे extendida” como una región conexa de pixeles con $U(x) = 1$. El criterio de conexidad considera dos pixeles como vecinos si son colindantes o si comparten una esquina.

Criterio utilizando las fuentes del catálogo

La base de datos maestra estará compuesta por todas las radiofuentes extendidas que contienen al menos una fuente del catálogo dentro. Al procesar las 2326 imágenes del NVSS se obtuvo que 1630964 radiofuentes extendidas cumplen con este criterio. En la Tabla 3.1 están las frecuencias respecto a la cantidad de fuentes internas.

En la Figura 3.2 tenemos diferentes visualizaciones de un elemento de la base de datos maestra. Las coordenadas de su centro son $(\alpha, \delta) = (11^h 9^m 57^s, -37^\circ 32' 31'')$.

Alrededor del 93.2% de las radiofuentes extendidas contienen sólo una radiofuentе; cerca del 5.8% contiene solo dos. El resto de la base de datos maestra representa sólo el 1% del total. El tamaño de la base de datos maestra es demasiado grande; es del mismo orden que el catálogo de radiofuentes.

Es difícil identificar como una RGG a una radiofuentе extendida con solo una fuente

3.1. Construcción de una base de datos maestra de interés

Cant. fuentes	Frec.						
96	1	38	2	24	10	11	52
85	1	37	1	23	9	10	83
64	1	35	3	22	6	9	100
59	1	34	2	21	7	8	147
55	1	33	2	20	13	7	235
53	1	32	2	19	12	6	395
52	1	31	2	18	12	5	796
47	1	30	2	17	22	4	2883
45	2	29	4	16	24	3	12119
43	1	28	2	15	27	2	93927
42	1	27	4	14	30	1	1519919
40	2	26	10	13	37		
39	1	25	5	12	43		

Tabla 3.1: Frecuencias respecto a la cantidad de fuentes internas en las radiofuentes extendidas de la base de datos maestra.

dentro, salvo que sólo el núcleo o un lóbulo esté detectado como fuente y el resto de la radiogalaxia se quede por debajo del umbral de detección para el catálogo. Por lo anterior, depuramos la base de datos maestra para que cada radiofuentе extendida contenga al menos dos fuentes internas. Es posible que una radiofuentе extendida con sólo una fuente dentro corresponda a una parte de una galaxia gigante fragmentada en varias radiofuentes extendidas, pero la detección de este tipo de objetos está fuera de los alcances de esta tesis.

Criterio eliminando radiofuentes extendidas en zonas contaminadas por ruido

Debido a la presencia de radiofuentes muy fuertes, muchas de ellas cerca del plano Galáctico (con coordenadas galácticas (l, b) con $|b|$ cercano a cero), procedemos a eliminar las radiofuentes extendidas en zonas con exceso de ruido debido a dichas fuentes fuertes. El nivel de ruido promedio en las imágenes es 0.45 mJy/beam. Estimamos el ruido de una zona es-

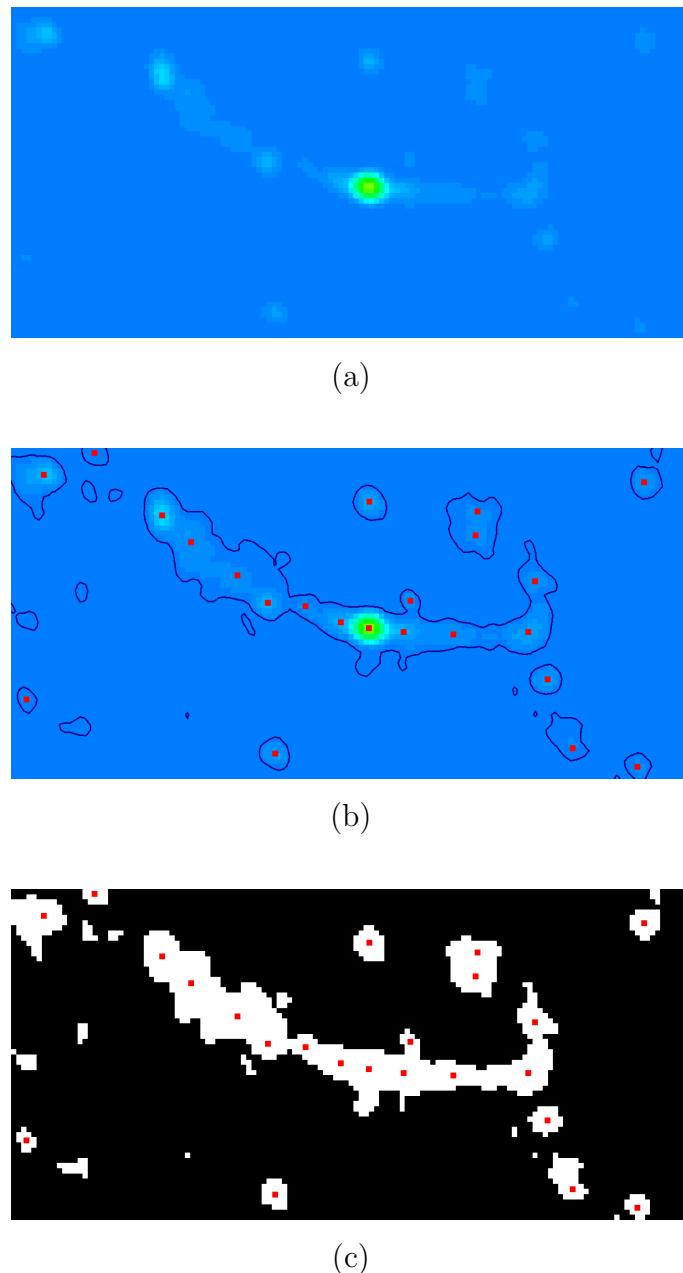


Figura 3.2: Distintas visualizaciones de un objeto en el NVSS. (a) Vecindad local del objeto graficada en falso color. (b) Los puntos rojos marcan la ubicación de las fuentes del catálogo en esa zona; las curvas del nivel 2.5σ aparecen en azul oscuro (c) Imagen después de aplicarle el umbral a nivel 2.5σ

3.1. Construcción de una base de datos maestra de interés

pecífica de una forma sencilla, pero que mostró ser suficientemente eficiente para identificar exceso de ruido. Esta forma se basa en el cálculo del RMS (*Root Mean Square*). Dada una vecindad rectangular N_0 de una imagen del NVSS, estimamos el ruido local con:

$$\sigma_{local} = \sqrt{\frac{1}{\#N} \sum_{x \in N} I(x)^2}, \text{ donde } N = \{x \in N_0 : I(x) < 2.5\sigma\} \quad (3.1)$$

Para estimar el ruido local de cada radiofuente extendida usamos una vecindad rectangular centrada en la radiofuente extendida con el triple de sus dimensiones. En la Figura 3.3 vemos un ejemplo de esto para la radiofuente extendida en $(\alpha, \delta) = (5^h 46^m 12^s, +63^\circ 20' 30'')$. Llamemos R al rectángulo más pequeño que contiene a la radiofuente extendida, el cual aparece en color rojo. El rectángulo azul delimita la vecindad N_0 usada en (3.3); en la figura vemos cómo N_0 triplica las dimensiones de R . Finalmente, los píxeles negros de N_0 integran el conjunto N en (3.3).

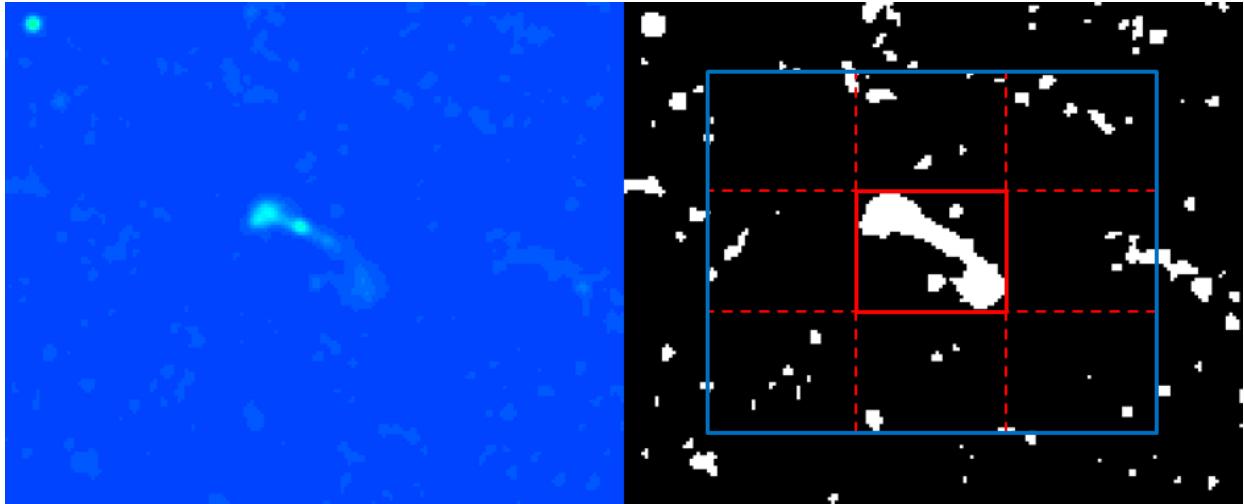


Figura 3.3: (a) A la izquierda tenemos la imagen del NVSS. (b) A la derecha la imagen después de aplicar el umbral.

Para determinar a partir de cuánto ruido una radiofuente extendida debe descartarse hicimos varias inspecciones visuales en imágenes que contienen el plano galáctico y remanentes de supernova. Por ejemplo, visualizamos en las Figuras 3.4 y 3.5 las zonas con ruido local menor y mayor a 0.6 mJy/beam.

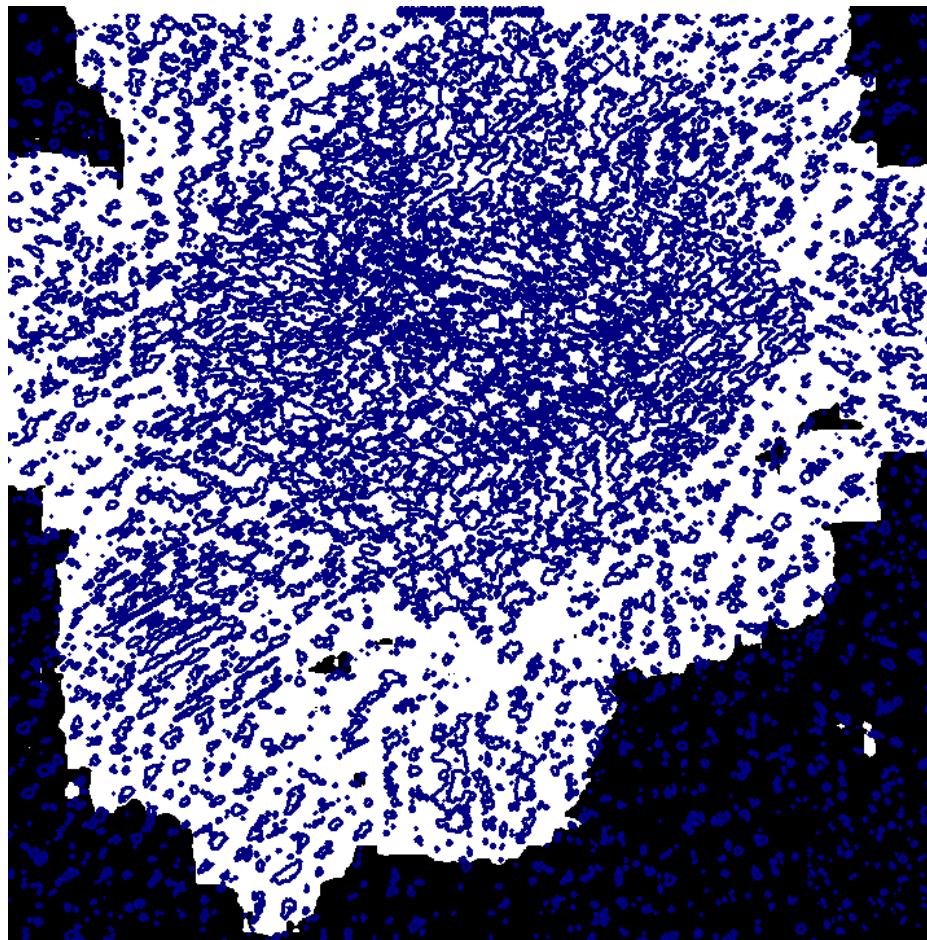


Figura 3.4: Curvas de nivel 2.5σ de la imagen I2000P40 del NVSS. La zona en blanco son las regiones con ruido local mayor a 0.6 mJy.

Con base en varios experimentos similares a los de las Figuras 3.4 y 3.5 concluimos que 0.6 mJy/beam es una buena cota superior para el ruido. Eliminamos de la base de datos maestra todas las radiofuentes extendidas que están en zonas con ruido local mayor a esta cota.

Base de datos resultante

Las radiofuentes extendidas restantes forman el conjunto que estudiaremos. En la Tabla 3.6 está la distribución respecto al número de fuentes internas.

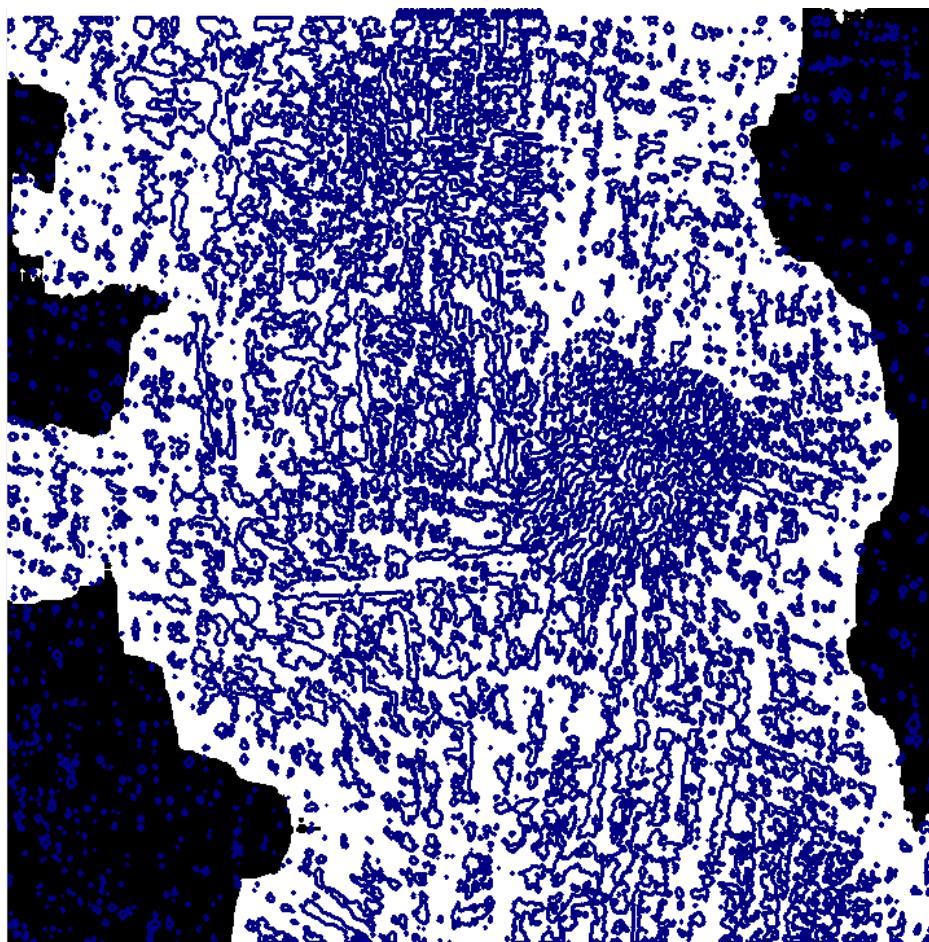


Figura 3.5: Curvas de nivel 2.5σ de la imagen I1724M36 del NVSS. La zona en blanco son las regiones con ruido local mayor a 0.6 mJy/beam.

Cada radiofuente extendida tiene inherentemente la siguiente información:

1. Nombre de la imagen de donde se encuentra.
2. Posición en la imagen.
3. Alto y ancho del rectángulo más pequeño que la contiene.
4. Área en pixeles
5. Cantidad y lista de fuentes internas.

Cant. fuentes	Frec.	Cant. fuentes	Frec.
12	3	6	108
11	2	5	325
10	7	4	1723
9	10	3	10218
8	23	2	88787
7	38		

Tabla 3.6: Cantidades de radiofuentes extendidas para cada número de fuentes internas después de aplicar los criterios de eliminación por exceso de ruido.

Las fuentes internas describen de cierta forma a la radiofuentre extendida. Sin embargo, nuestro objetivo es caracterizar la morfología de la radiofuentre extendida con mediciones adicionales que puedan ser de utilidad para identificar las RGG.

3.2 Lista de características

En esta sección definimos características que proponemos medir para cada radiofuentre extendida, las cuales consideramos como informativas para la tarea de reconocimiento. Es el equivalente en nuestro trabajo a la Tabla 1.9.

3.2.1 Longitud angular

Buscamos medir la longitud angular de las radiofuentes extendidas. Como en muchos casos tienen cierta curvatura, medir la longitud angular tomando como referencia una recta no es acertado. Usamos el *morphological skeleton* (al que nos referiremos simplemente como *skeleton*) para tener una mejor medición. Por simplicidad a partir de aquí nos referiremos a la longitud angular simplemente como longitud.

El *skeleton* reduce la radiofuentre extendida a líneas que resumen su morfología. Extraemos el camino o rama más larga del *skeleton* y medimos su longitud. Esta es la característica usada.

3.2. Lista de características

Contrario al caso continuo, no existe definición única del *skeleton* en el caso discreto. Una forma de obtener el *skeleton* que es común encontrar en la literatura está descrita en el Algoritmo 1. Hace uso de las operaciones erosión y apertura para una imagen binaria I con un elemento estructural D . Se considera una imagen binaria como un conjunto de píxeles y para un pixel x y un conjunto D llamado elemento estructural, D_x denota $\{x + d \mid d \in D\}$. Entonces se tienen las definiciones:

$$\text{Erosión}_D(I) = \{x \mid D_x \subset I\},$$

$$\text{Apertura}_D(I) = \cup_{x \in I, D_x \subset I} D_x.$$

En las Figuras 3.7 y 3.8 observamos ejemplos de erosión y apertura respectivamente, usando como elemento estructural un cuadrado de 3×3 píxeles.

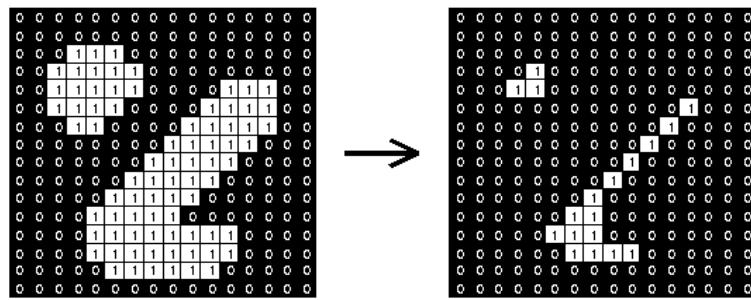


Figura 3.7: Ejemplo de aplicación del operador erosión con elemento estructural igual al cuadrado de 3×3 píxeles (Delmas 2012).

En esta tesis usamos una función programada en MATLAB versión R2012a (MATLAB 2012), la cual es bastante eficiente. La función usada se llama “*bwmorph*” (MATLAB 2011a) y puede hacer varias operaciones morfológicas.

Una vez calculado el *skeleton*, es necesario determinar cuál es el camino más largo en él. Implementamos el Algoritmo 2 para encontrar la geodésica más larga en un conjunto conexo (Soille 1999). En el aparecen:

- La función de propagación $P(I)$: a cada pixel x se le asocia el tamaño del camino mínimo más largo que empieza en x y termina en algún pixel de I .

- Las funciones de distancia geodésica $D_I(x_1)$: a cada pixel x se le asocia el tamaño del camino mínimo de x hasta x_1 .

Para la implementación del Algoritmo 2 usamos la función “`bwdistgeodesic`” (MATLAB 2011b) de MATLAB para calcular las distancias geodésicas y la función de propagación.

Finalmente, el cálculo de la longitud se hizo a nivel de píxeles. Se recorrió el camino más largo calculando las distancias entre pixel y pixel. En la Figura 3.9 podemos ver una ejemplo con el objeto centrado en $(\alpha, \delta) = (7^h 54^m 59^s, -2^\circ 22' 36'')$.

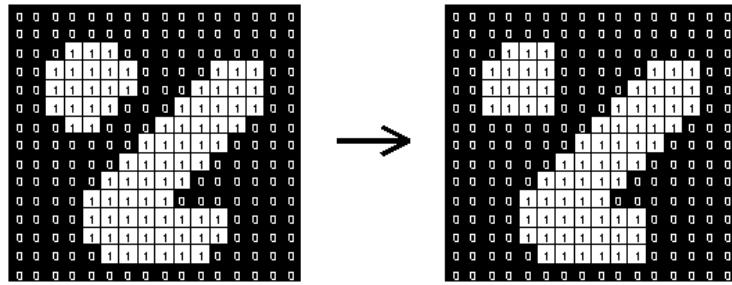


Figura 3.8: Ejemplo de aplicación del operador apertura con elemento estructural igual al cuadrado de 3×3 píxeles (Delmas 2012).

Algoritmo 1 Cálculo del esqueleto

Entrada: Imagen binaria I , elemento estructural D

Salida: *Skeleton* de I

- 1: $S \leftarrow I$
 - 2: **while** Existan píxeles con valor 1 en I **do**
 - 3: $S \leftarrow S \setminus (I \& \neg Apertura_D(I))$
 - 4: $I = Erosion_D(I)$
 - 5: **end while**
 - 6: **return** S
-

3.2. Lista de características

Algoritmo 2 Extracción del camino más largo del esqueleto

Entrada: *Skeleton I*

Salida: Camino más largo de *I*

- 1: Calcular $P(I)$, la función de propagación de *I*.
 - 2: Extraer los dos pixeles x_1 y x_2 tales que $[P(I)](x_i) = \max_{y \in I} [P(I)](y)$
 - 3: Calcular las funciones de distancia geodésica entre x_1 y x_2 : $D_I(x_1)$ y $D_I(x_2)$.
 - 4: Calcular S_I , la suma de $D_I(x_1)$ y $D_I(x_2)$.
 - 5: A lo largo del camino buscado S_I tiene su valor mínimo, el cual es igual a la distancia geodésica entre x_1 y x_2 . Localizamos dicho camino y lo guardamos en la imagen binaria *C*.
 - 6: **return** *C*
-

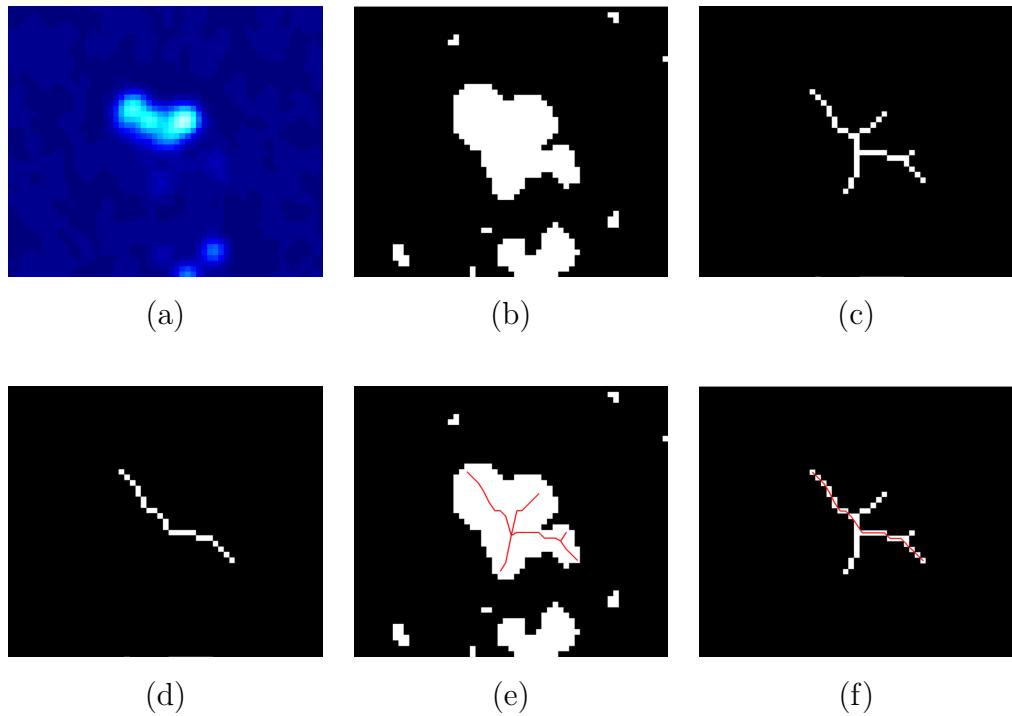


Figura 3.9: (a) Imagen del NVSS (b) Radiofuente extendida (c) *Skeleton* (d) Camino más largo del *skeleton* (e) Radiofuente extendida y *skeleton* (f) *Skeleton* y su camino más largo

3.2.2 Índice de asimetría (asimetría I)

Dado un eje de simetría l , calculamos el índice de asimetría (o error de simetría) de la radiofuentе extendida M respecto a ese eje con:

$$\sum_{x \in M} \frac{|I(x) - I(R_l(x))|}{\mu_I}$$

donde $R_l(x)$ es el pixel reflejado de x respecto a la recta l y μ_I es la intensidad promedio sobre la radiofuentе extendida.

Usamos como eje de simetría la recta definida por:

- La pendiente perpendicular al ángulo dominante en la radiofuentе extendida. Estimamos el ángulo dominante usando un método computacionalmente intensivo. El pseudocódigo está en Algoritmo 3.
- El punto que marca la mitad del camino más largo del *skeleton*.

La Figura 3.10 ilustra diferentes valores de este índice. Observamos en cada recuadro la imagen del NVSS (con las coordenadas de su posición), el eje de simetría, el camino más largo del *skeleton*, las fuentes internas y el contorno correspondiente.

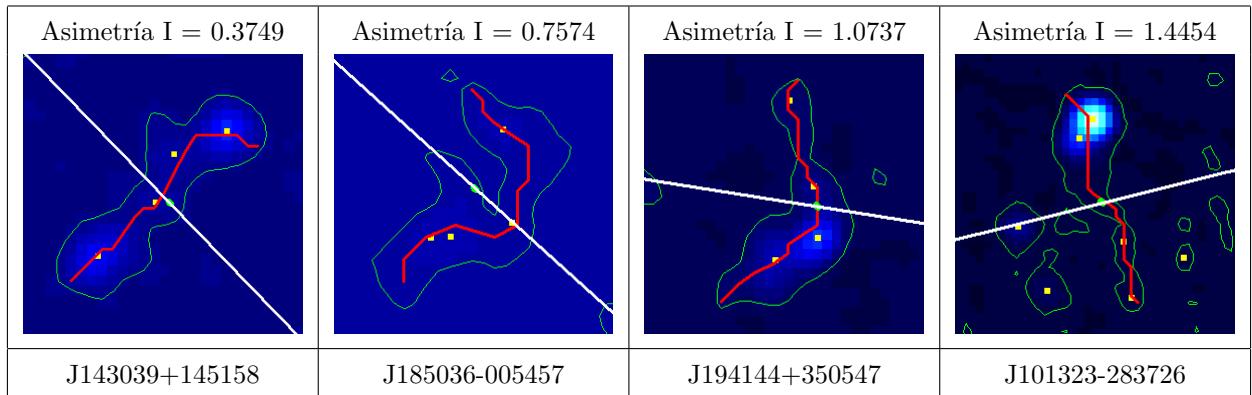


Figura 3.10: Ejemplos de radiofuentes extendidas y su índice de asimetría I.

3.2. Lista de características

Algoritmo 3 Estimación del ángulo dominante de una radiofuentе extendida

Entrada: Radiofuentе extendida M

Salida: Estimación del ángulo dominante de M : θ_M

- 1: Generar aleatoriamente una colección de pares de pixeles $\{(x_1^i, x_2^i)\}$ con $x_1^i, x_2^i \in M$.
 - 2: Calcular los ángulos $\{\theta_i\}$. θ_i es el ángulo de inclinación de la recta definida por x_1^i y x_2^i .
 - 3: $\theta_M \leftarrow \text{Moda}(\{\theta_i\})$
 - 4: **return** θ_M
-

3.2.3 Índice de asimetría sobre la imagen binaria (asimetría B)

Se obtiene igual que la medida anterior, con la diferencia de que la asimetría se calcula sobre la radiofuentе extendida. De esta forma, la asimetría se mide solo sobre la silueta binaria del objeto y no sobre las intensidades. La Figura 3.11 ilustra diferentes valores de este índice. En cada recuadro observamos la imagen binarizada (con las coordenadas su posición), el eje de simetría, el camino más largo del *skeleton*, las fuentes internas y el contorno correspondiente.

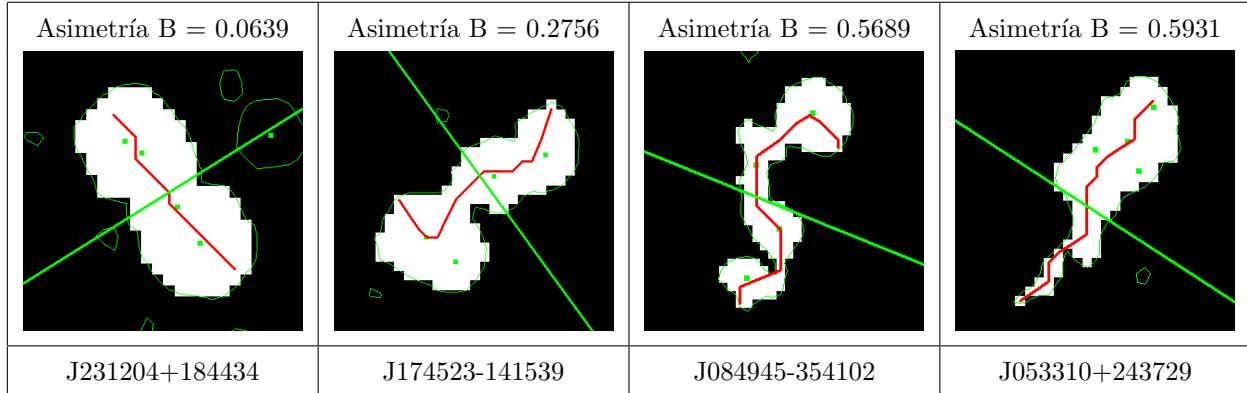


Figura 3.11: Ejemplos de radiofuentes extendidas y su índice de asimetría B.

3.2.4 Índice de asimetría sobre la imagen binaria usando apertura (asimetría A)

En varios ejemplos observamos que las medidas de asimetría anteriores se ven afectadas por zonas pequeñas que no tienen contraparte respecto al eje, pero que afectan de manera significativa la percepción de la simetría global. Para buscar deshacernos del incremento

en el índice causado por zonas pequeñas usamos un operador morfológico. Aplicamos una apertura sobre las zonas que no tienen correspondencia antes de calcular el índice de asimetría justo como en la sección anterior. En las imágenes en la Figura 3.12 los puntos marcados en rojo son los pixeles que si tienen pixel correspondiente respecto al eje de simetría. Los pixeles en blanco son los pixeles que no tienen simétrico respecto al eje y que no fueron afectados por la apertura. También observamos la posición del objetos, el eje de simetría y el contorno correspondiente.

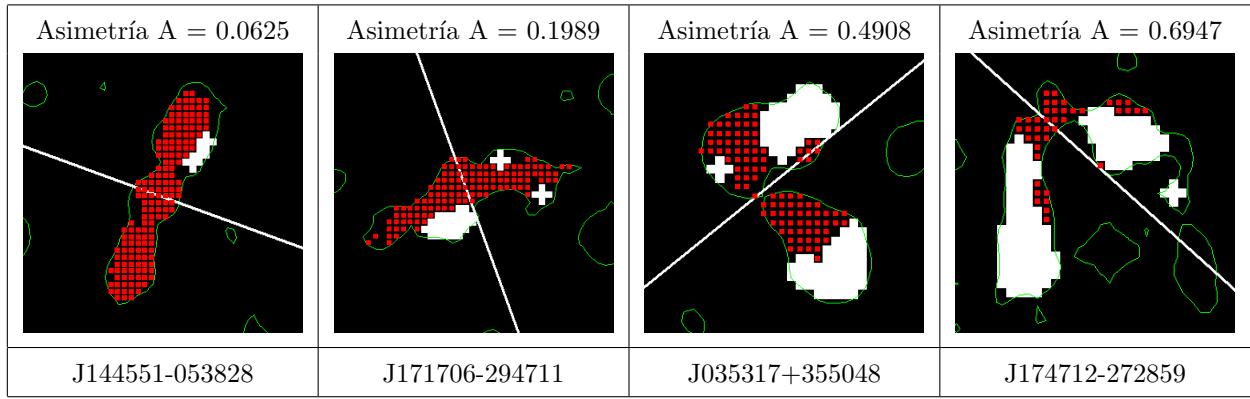


Figura 3.12: Ejemplos de radiofuentes extendidas y su índice de asimetría A.

3.2.5 Índice de gordura

Se requiere una característica que nos indique qué tan alargada es la radiofuentе extendida. Una propuesta sencilla que resultó efectiva es usar la razón entre el área de la radiofuentе extendida y el cuadrado de su longitud. De esta forma objetos con esta razón grande serán de mayor “gordura”. Si este factor es pequeño indica que el área de la radiofuentе extendida está pegada a la curva de donde se extrajo la longitud. La Figura 3.13 ilustra diferentes valores de este índice².

²En este ejemplo y en los correspondientes a las características restantes observamos en cada recuadro la imagen del NVSS y las coordenadas de su posición, el eje de simetría, el camino más largo del *skeleton*, las fuentes internas y el contorno de nivel 2.5σ .

3.2. Lista de características

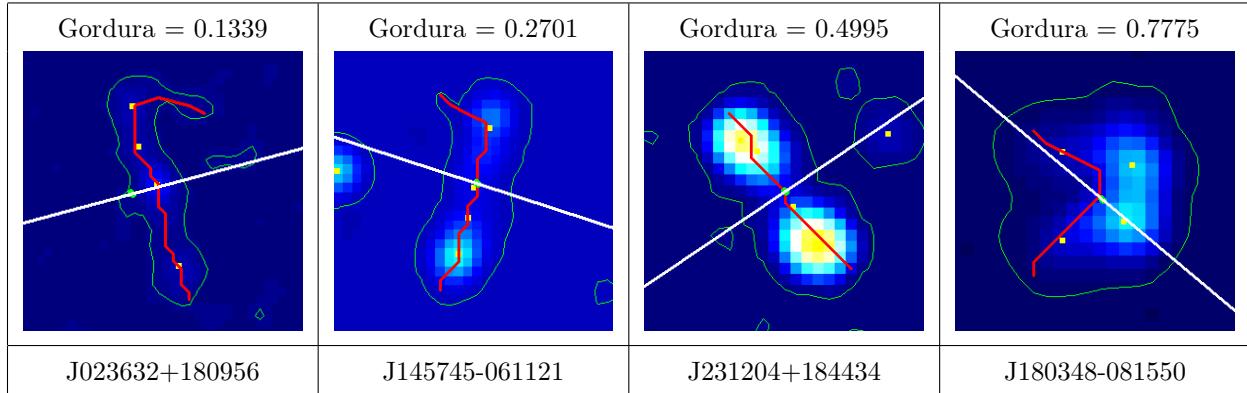


Figura 3.13: Ejemplos de radiofuentes extendidas y su índice de gordura.

3.2.6 Índice de delgadez central

Este índice mide qué tan delgada es la radiofuentе extendida en su parte central. Se calcula dividiendo el área entre el cuadrado del ancho central. El ancho central es la distancia entre los dos puntos donde el eje de simetría corta a la frontera de la radiofuentе extendida. La Figura 3.14 ilustra diferentes valores de este índice.

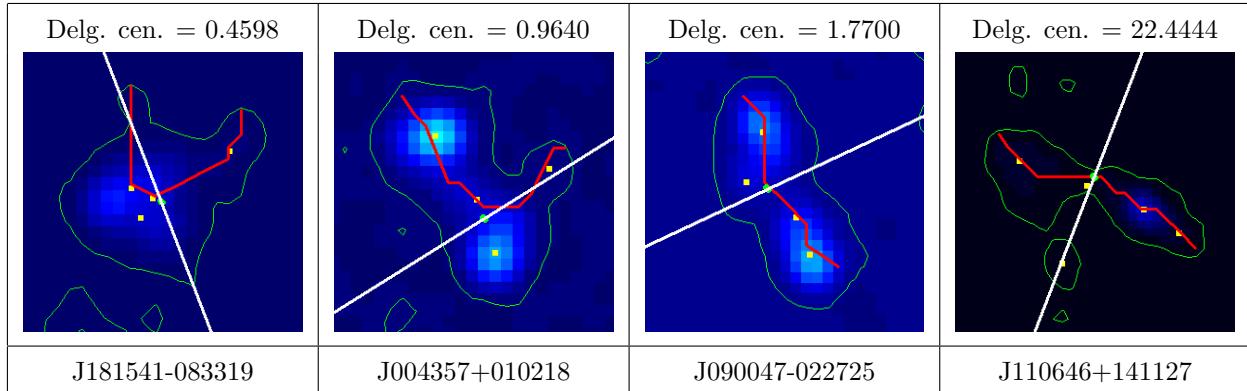


Figura 3.14: Ejemplos de radiofuentes extendidas y su índice de delgadez central.

3.2.7 Índice de rectitud

Para darnos una idea de qué tan recta o derecha está la radiofuentе extendida usamos la razón entre la longitud del camino más largo del *skeleton* y la longitud de la recta que une a los dos extremos. De esta forma, entre más pequeño sea el factor indica una curvatura

más completa. Un factor igual a 1 corresponde a una radiofuentе extendida completamente recta. La Figura 3.15 ilustra diferentes valores de este índice.

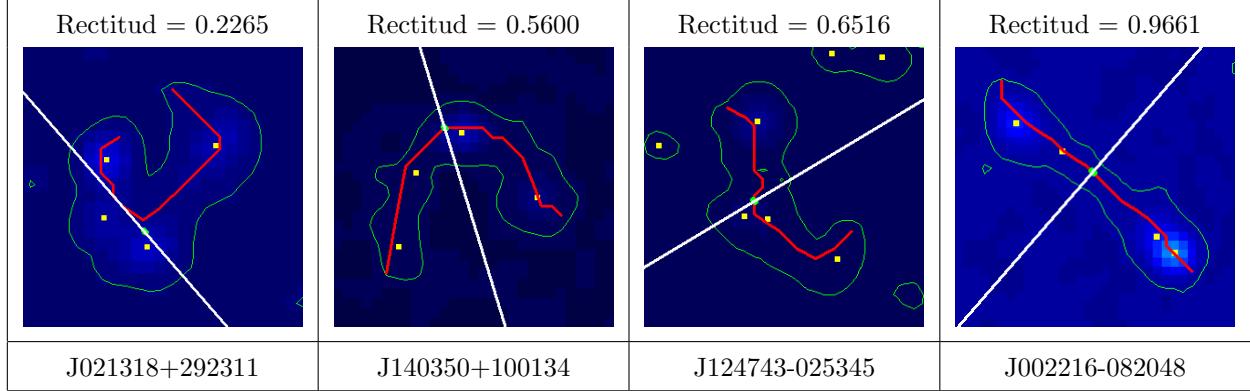


Figura 3.15: Ejemplos de radiofuentes extendidas y su índice de rectitud.

3.2.8 Intensidad promedio

Esta característica busca describir qué tan alta es la intensidad de la radiofuentе extendida. Para esto promediamos las intensidades de los pixeles que forman la radiofuentе extendida. El valor obtenido mide el brillo superficial promedio sobre toda la extensión de la radiofuentе extendida en unidades de Jy/beam. La Figura 3.16 ilustra diferentes valores de esta característica.

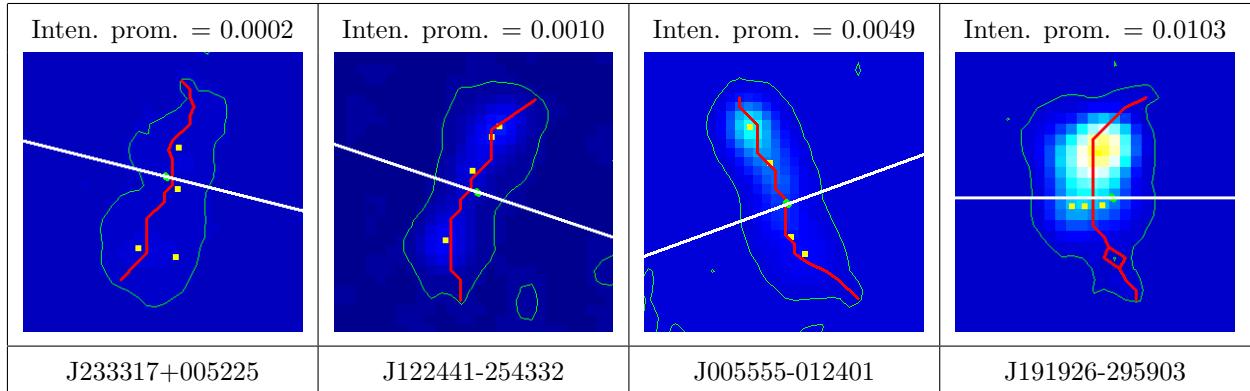


Figura 3.16: Ejemplo de radiofuentes extendidas y su intensidad promedio.

3.2.9 Índice de contraste de intensidad

Esta característica tiene como fin ser un indicador de que tanta intensidad está acumulada en la parte central de la radiofuentе extendida. Para calcular este índice tomamos como referencia el pixel en la mitad del camino más largo del *skeleton*. Encontramos la intensidad máxima en una vecindad del pixel de referencia y la dividimos entre la intensidad promedio. La Figura 3.17 ilustra diferentes valores de este índice.

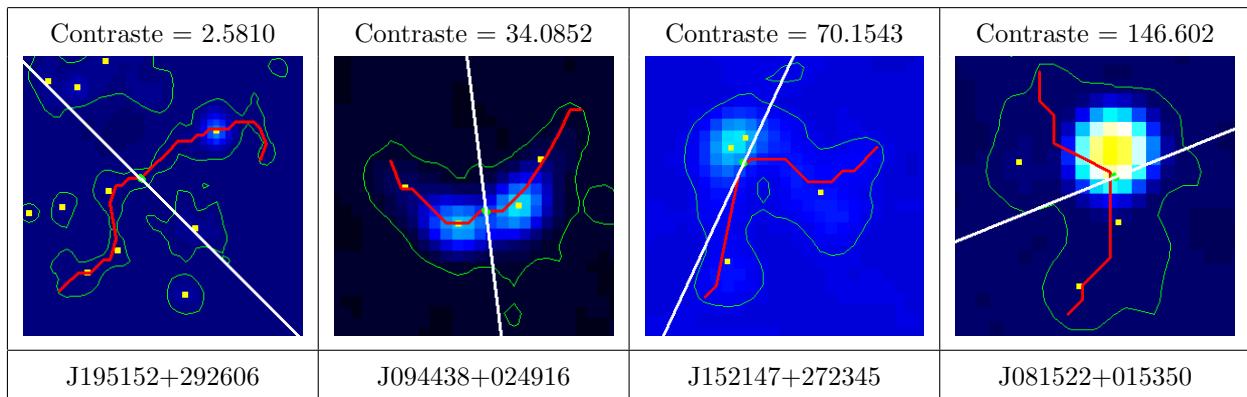


Figura 3.17: Ejemplo de radiofuentes extendidas y su índice de contraste de intensidad.

3.3 Poder discriminatorio de las características

En esta sección investigamos qué tanto las características descritas en la sección anterior ayudan a identificar radiofuentes extendidas que sean candidatas plausibles a ser RGG. Un astrónomo³ inspeccionó visualmente 1000 radiofuentes extendidas de la base maestra (todas con cuatro fuentes asociadas) clasificándolas en las categorías “candidata probable a ser RGG”, “candidata improbable a ser RGG” y “candidata dudosa”. También indicó para cada una si está muy cerca del plano galáctico ($|l| < 15^\circ$) y si la región de la imagen donde se encuentra es muy ruidosa o no. A partir de esta lista construimos el conjunto con el que

³Esta inspección fue realizada por la estudiante Iris del Carmen Santiago Bautista (estudiante de Licenciatura en Física en la Universidad Veracruzana) como parte de una estancia de investigación con el Dr. Carlos A. Rodríguez Rico.

se trabajó. Nos limitamos a los objetos que no están cerca del plano galáctico ($|l| > 15^\circ$) y que no se encuentran en una zona muy ruidosa de acuerdo a la estimación obtenida con (3.1). El resultado fue un conjunto con 92 objetos probables, 193 no probables y 175 con dudas.

A continuación se estudia cada una de las características por separado. Podemos ver las gráficas resultantes en las Figuras 3.18, 3.19, 3.20, 3.21, 3.22, 3.23 y 3.24. Con excepción del índice de intensidad, se nota que la distribución de las características difieren para las clases probable de ser RGG y poco probables de ser RGG, es decir, las características ayudan a predecir la clase correcta.

3.4 Clasificadores basados en las características

En esta sección ajustamos modelos de regresión logística. Si Y indica si una radiofuentre extendida es probable a ser ($Y = 1$) o no ser ($Y = 0$) RGG y $X = (X_1, \dots, X_d)$ es el vector de características de la radiofuentre extendida en cuestión, entonces se estiman α y β tales que

$$P(Y = 1 | X = x) = \frac{1}{1 + e^{-\alpha - \beta' x}}.$$

Después de eliminar las variables no significativas se obtuvo el siguiente modelo:

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-0.5508	1.2568	-0.438	0.661189
Asimetria I	-1.4411	0.4796	-3.005	0.002656 **
Asimetria A	-3.7616	1.2534	-3.001	0.002690 **
Rectitud	5.2800	1.3757	3.838	0.000124 ***
Gordura	-8.7332	1.6054	-5.440	5.33e-08 ***

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 358.51 on 284 degrees of freedom
Residual deviance: 276.92 on 280 degrees of freedom

3.4. Clasificadores basados en las características

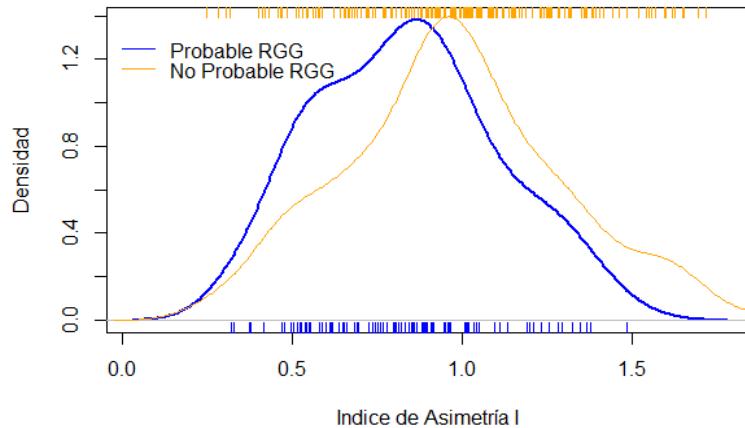


Figura 3.18: Densidades marginales del índice de asimetría I de las radiofuentes extendidas etiquetadas como probables de ser RGG y de las etiquetadas como poco probables de ser RGG.

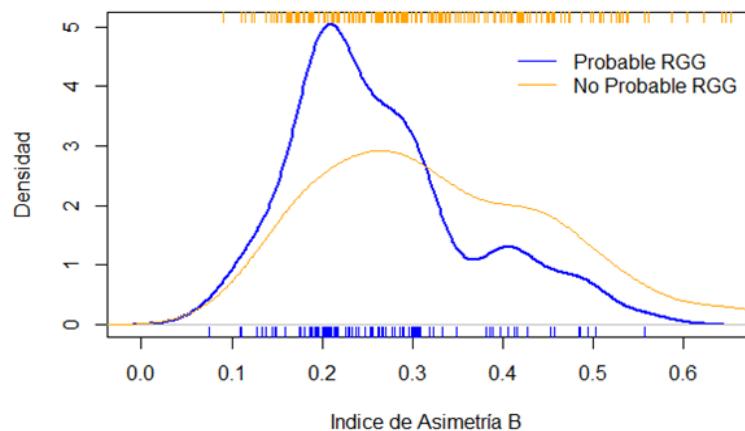


Figura 3.19: Densidades marginales del índice de asimetría B de las radiofuentes extendidas etiquetadas como probables de ser RGG y de las etiquetadas como poco probables de ser RGG.

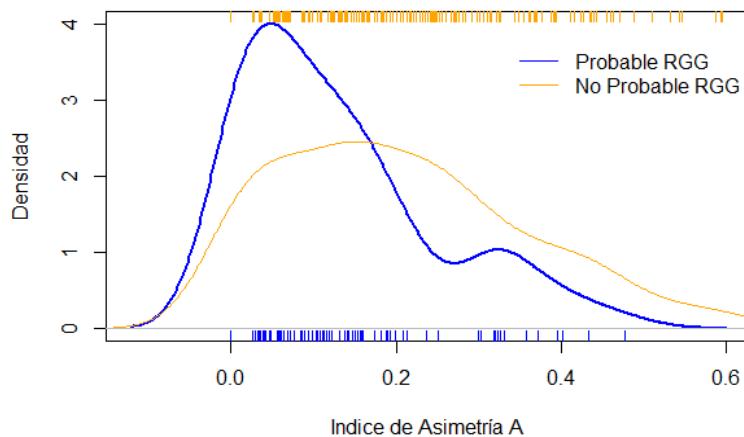


Figura 3.20: Densidades marginales del índice de asimetría A de las radiofuentes extendidas etiquetadas como probables de ser RGG y de las etiquetadas como poco probables de ser RGG.

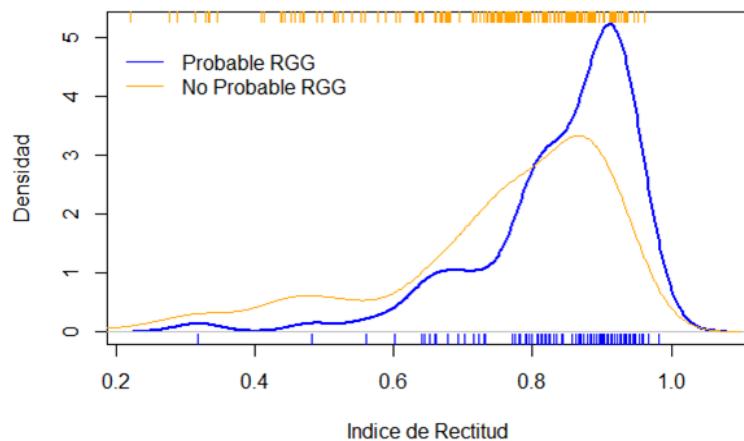


Figura 3.21: Densidades marginales del índice de rectitud de las radiofuentes extendidas etiquetadas como probables de ser RGG y de las etiquetadas como poco probables de ser RGG.

3.4. Clasificadores basados en las características

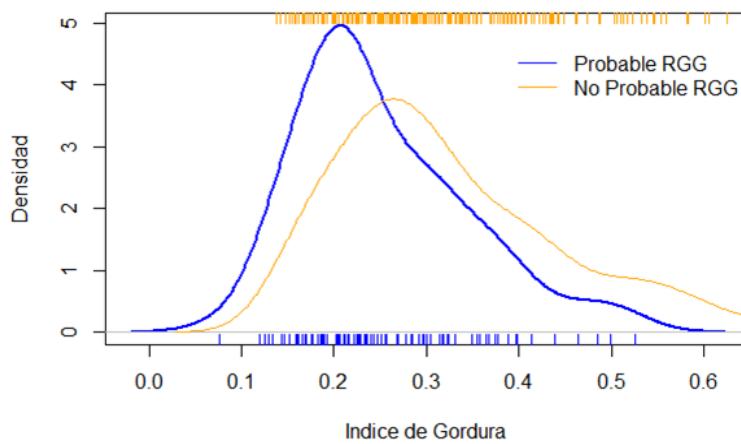


Figura 3.22: Densidades marginales del índice de gordura de las radiofuentes extendidas etiquetadas como probables de ser RGG y de las etiquetadas como poco probables de ser RGG.

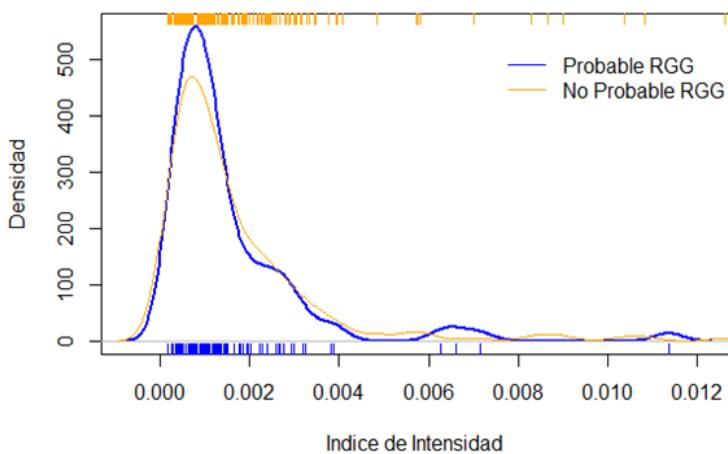


Figura 3.23: Densidades marginales de la intensidad promedio de las radiofuentes extendidas etiquetadas como probables de ser RGG y de las etiquetadas como poco probables de ser RGG.

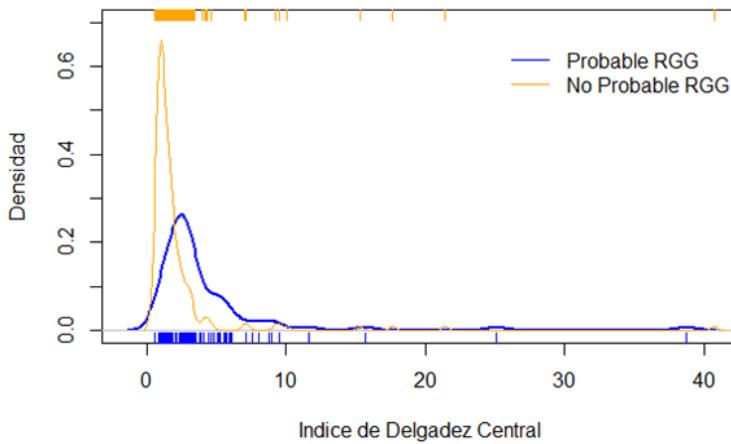


Figura 3.24: Densidades marginales del índice de delgadez central de las radiofuentes extendidas etiquetadas como probables de ser RGG y de las etiquetadas como poco probables de ser RGG.

Se observa un *residual deviance* aceptable. Las cuatro variables significativas son: 1. Índice de simetría I, 2. Índice de simetría A. Índice de rectitud y 4. Índice de gordura. Los signos de los coeficientes son los que se esperan.

Los resultados de la predicción están en la Tabla 3.25. El error total de clasificación es 22%.

		Clasificación humana	
		No prob. RGG	Prob. RGG
Clasificación automática	No prob. RGG	173	40
	Prob. RGG	20	52

Tabla 3.25: Resultados de la predicción usando regresión logística.

Usando un conjunto de datos de prueba el error sube a alrededor del 25%, pero aún por debajo de lo que se obtiene con el clasificador ingenuo bayesiano.

En la Figura 3.26 graficamos las probabilidades estimadas de que $Y = 1$ contra el valor observado para Y ($Y = 0$ para las radiofuentes extendidas poco probables a ser RGG, y $Y = 1$ para las probables a serlo); nos limitamos a los datos de un conjunto de prueba. En azul marcamos una línea de tendencia usando regresión no paramétrica para $E(Y | \text{el valor}$

3.4. Clasificadores basados en las características

estimado para $P(Y = 1 | X)$). Como debe ser, observamos que entre los objetos con una probabilidad estimada baja de ser RGG dominan los clasificados por los astrónomos como poco probables a serlo, y conforme aumenta la estimación sube el porcentaje de objetos que fueron clasificados como probables a ser RGG.

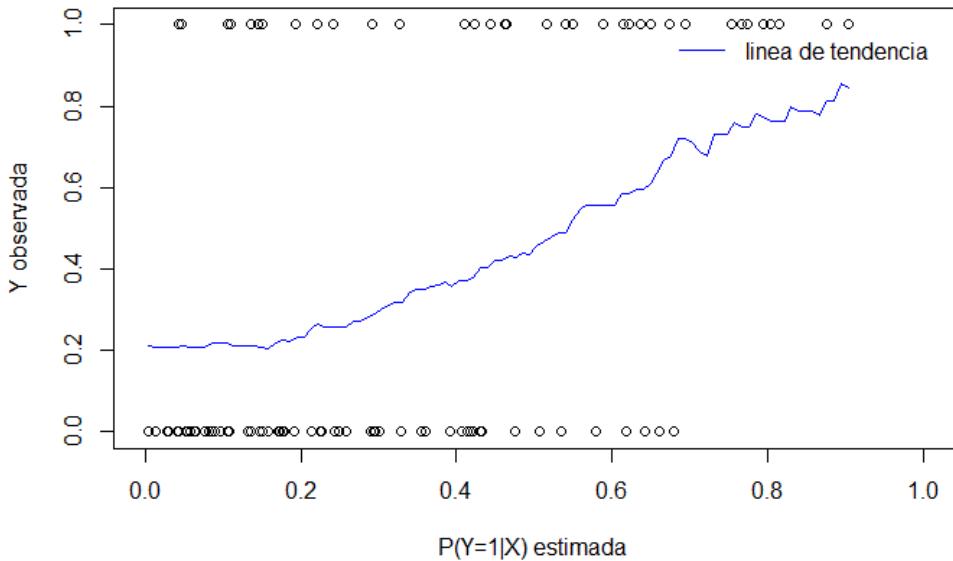


Figura 3.26: Graficación de $P(Y = 1 | X = x)$ bajo el modelo de regresión logística versus los valores de Y observados para un conjunto de prueba (puntos negros). La línea azul es una estimación de la tendencia usando regresión no paramétrica.

Las imágenes en la Figura 3.27 corresponden a los candidatos que fueron clasificados como no probables por los astrónomos, y bajo el modelo construido tienen las probabilidades más altas (peores falsos positivos). En la Figura 3.28 tenemos los candidatos que fueron clasificados como probables por los astrónomos y bajo el modelo construido tienen las probabilidades más bajas (peores falsos negativos).

Las imágenes en las Figuras 3.27 y 3.28 pueden servir para proponer nuevas características, que eventualmente puedan evitar la pobre clasificación de estas radiofuentes extendidas.

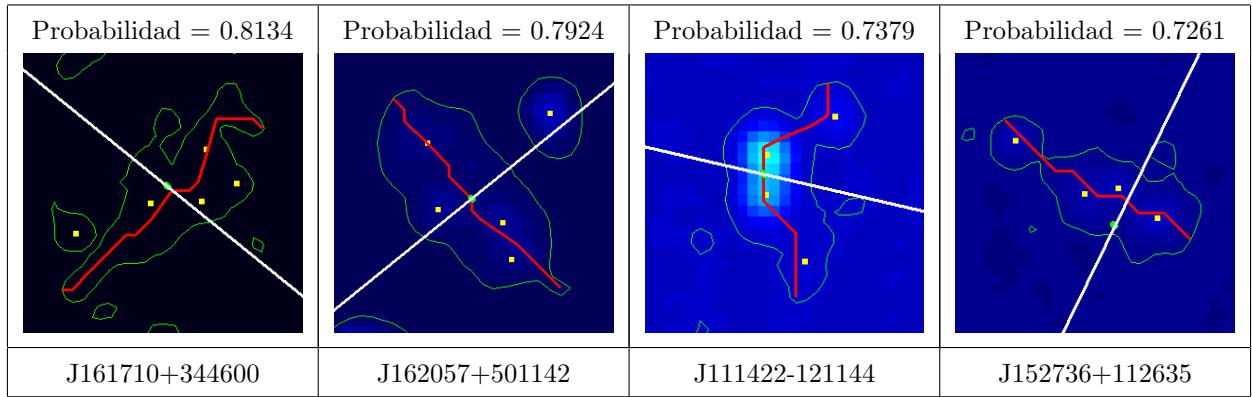


Figura 3.27: Peores falsos positivos bajo el modelo de regresión logística

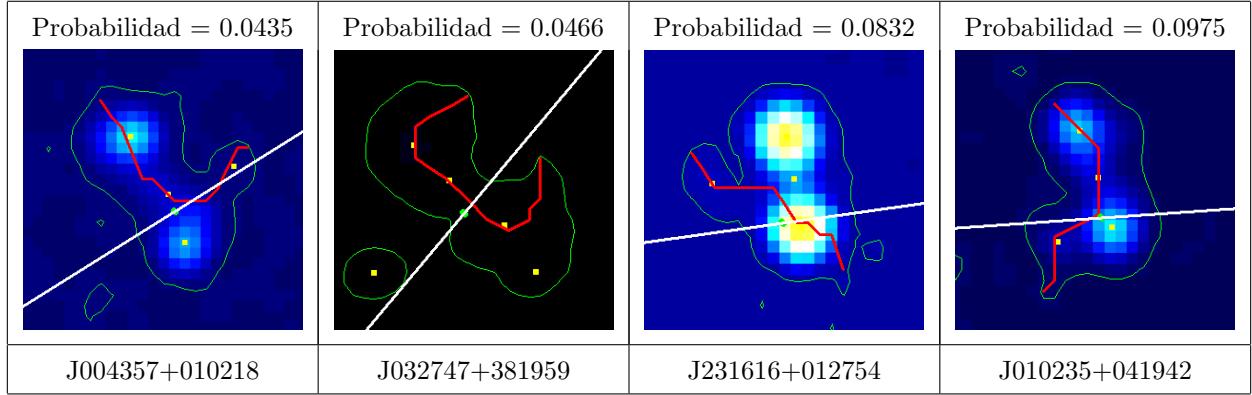


Figura 3.28: Peores falsos negativos bajo el modelo de regresión logística.

3.5 Radiofuentes extendidas con potencial de ser RGG no registradas

Durante la inspección descrita al inicio de la Sección 3.3, además de clasificar los 1000 objetos, también se analizó otra información disponible. Para cada candidato que se clasificó como probable a ser RGG, se estudió su imagen correspondiente en óptico (en caso de estar disponible). Este análisis condujo en varias ocasiones a evidencia de la existencia de la galaxia anfitriona, y en algunos casos no se tiene registro en la literatura de dicha correspondencia.

Lo anterior es interesante ya que el trabajo realizado en esta tesis podría aportar para el registro de ciertas RGG aún no conocidas por la comunidad científica. Cabe aclarar que este es apenas un primer paso; para poder concluir que, efectivamente, los candidatos

3.5. Radiofuentes extendidas con potencial de ser RGG no registradas

en cuestión son RGG aún faltarían observaciones adicionales, consultas y análisis de otras fuentes de información.

En la Figura 3.29 observamos las cuatro radiofuentes extendidas más prometedoras. Para cada una observamos las coordenadas de su posición y el tamaño absoluto correspondiente suponiendo que efectivamente la que se piensa es la galaxia anfitriona.

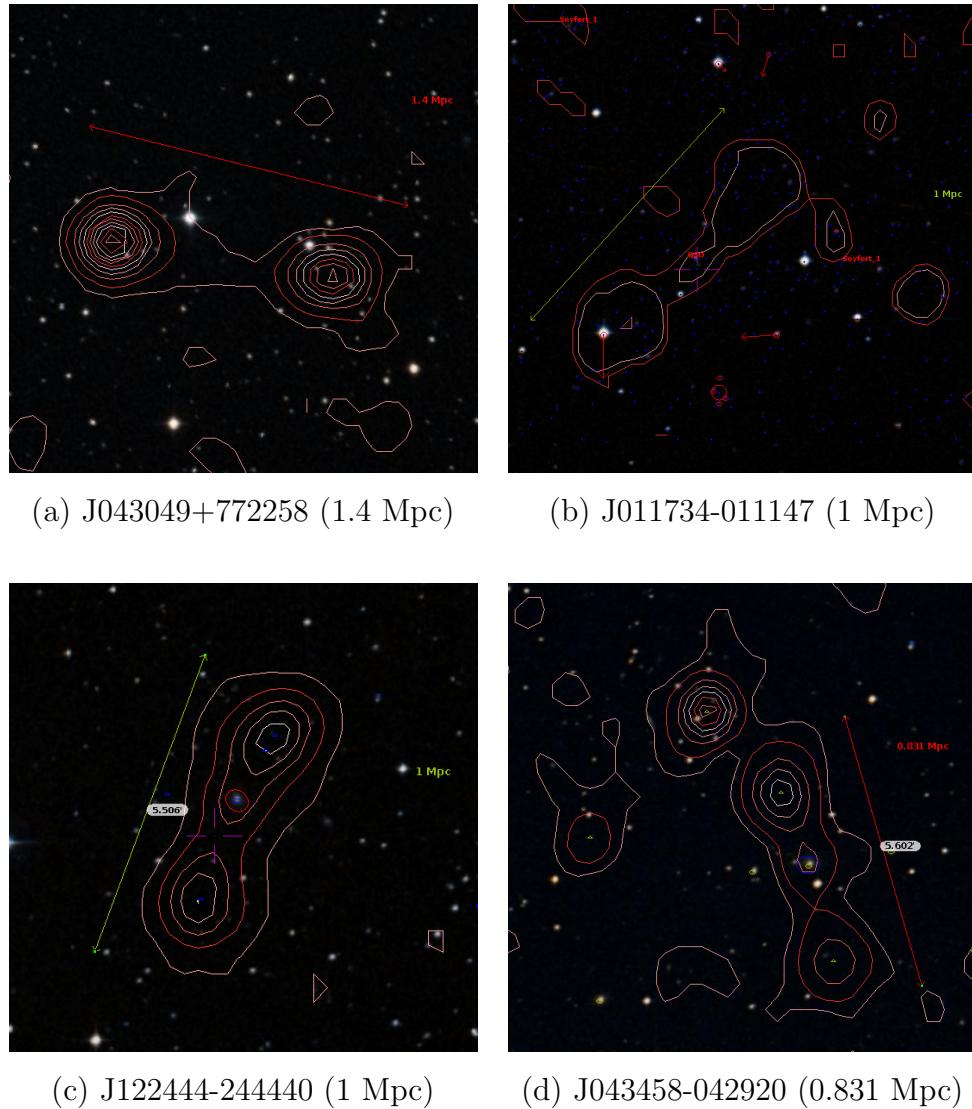


Figura 3.29: Contornos de las radiofuentes extendidas más prometedoras para ser RGG no registradas. Observamos los contornos de las imágenes del NVSS sobre las imágenes correspondientes en óptico.

Apéndice A

Sistemas de coordenadas astronómicas

En astronomía, los sistemas de coordenadas astronómicas o sistemas de coordenadas celestes sirven para especificar posiciones de objetos astronómicos o simplemente la dirección respecto a un punto de referencia. Existen sistemas de coordenadas tanto esféricas como rectangulares. Por lo general el origen y los planos que definen el sistema de coordenadas está asociado a algún objeto astronómico, como pueden ser la Tierra, el Sol o nuestra Galaxia (Vía Láctea).

En esta tesis se usa principalmente el sistemas de coordenadas ecuatoriales. Para algunos aspectos es muy relevante el sistema de coordenadas Galácticas.

A.1 Sistema de coordenadas ecuatoriales

Este sistema de coordenadas es análogo al sistema de coordenadas geográficas sobre la superficie terreste (latitud, longitud). Pensemos en un objeto astronómico, y proyectemos su posición al centro de la Tierra con una recta. El punto donde dicha recta corta a la superficie terreste define las coordenadas ecuatoriales del objeto. El estándar es usar las letras griegas α y δ para referirse a las coordenadas ecuatoriales; α se llama ascensión recta y δ declinación. La ascención recta corresponde a la longitud y la declinación a la latitud en el sistema de coordenadas geográficas. Por lo tanto, la ascención recta se mide en horas, minutos y segundos y la declinación en grados (positivos hacia el norte desde el ecuador y

negativos hacia el sur). En la Figura A.1 observamos la alineación del plano de este sistema de coordenadas.

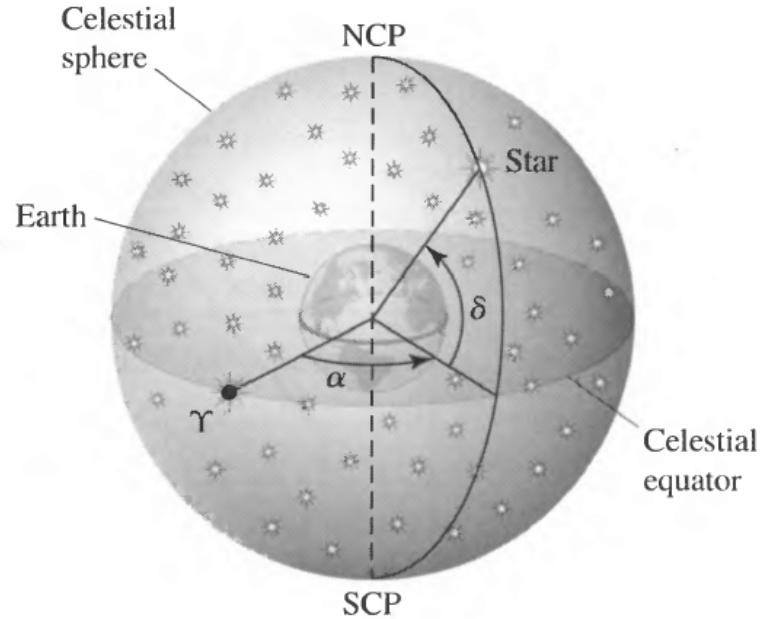


Figura A.1: El sistema de coordenadas ecuatoriales α , δ y γ definen la ascensión recta, la declinación y la posición del equinoccio vernal, respectivamente (Carroll & Ostlie 2007).

Ya que la Tierra cambia su posición constantemente respecto al resto del universo, las coordenadas de un objeto no se mantendrían fijas a través del tiempo. Por lo tanto se define como el cero de ascensión recta el “punto vecinal”, la dirección del Sol en el equinoccio “vernal” (primavera, 21 ± 1 de marzo) de cada año. Sin embargo, por la precesión del eje de rotación de la Tierra esta dirección cambia con una tasa de $\sim 50'/\text{año}$. Por lo tanto, los vectores de α y δ tienen que especificarse junto con el año de referencia. Por lo general se usa el año de referencia estándar, el cual se cambia cada 50 años. J2000 es año de referencia estándar actual y B1950 es el que se usaba anteriormente.

A.2 Sistema de coordenadas Galácticas

Este sistema es similar al de coordenadas ecuatoriales; también son coordenadas esféricas, con la diferencia de que las dos componentes se miden en grados con decimales. Lo que las hace distintas es que el origen de este sistema de coordenadas se encuentra en el Sol y el plano con el que están alineadas coincide con el llamado plano Galáctico (el plano con el que está alineada nuestra galaxia). Las coordenadas se denotan por (l, b) , longitud galáctica y latitud galáctica, respectivamente.

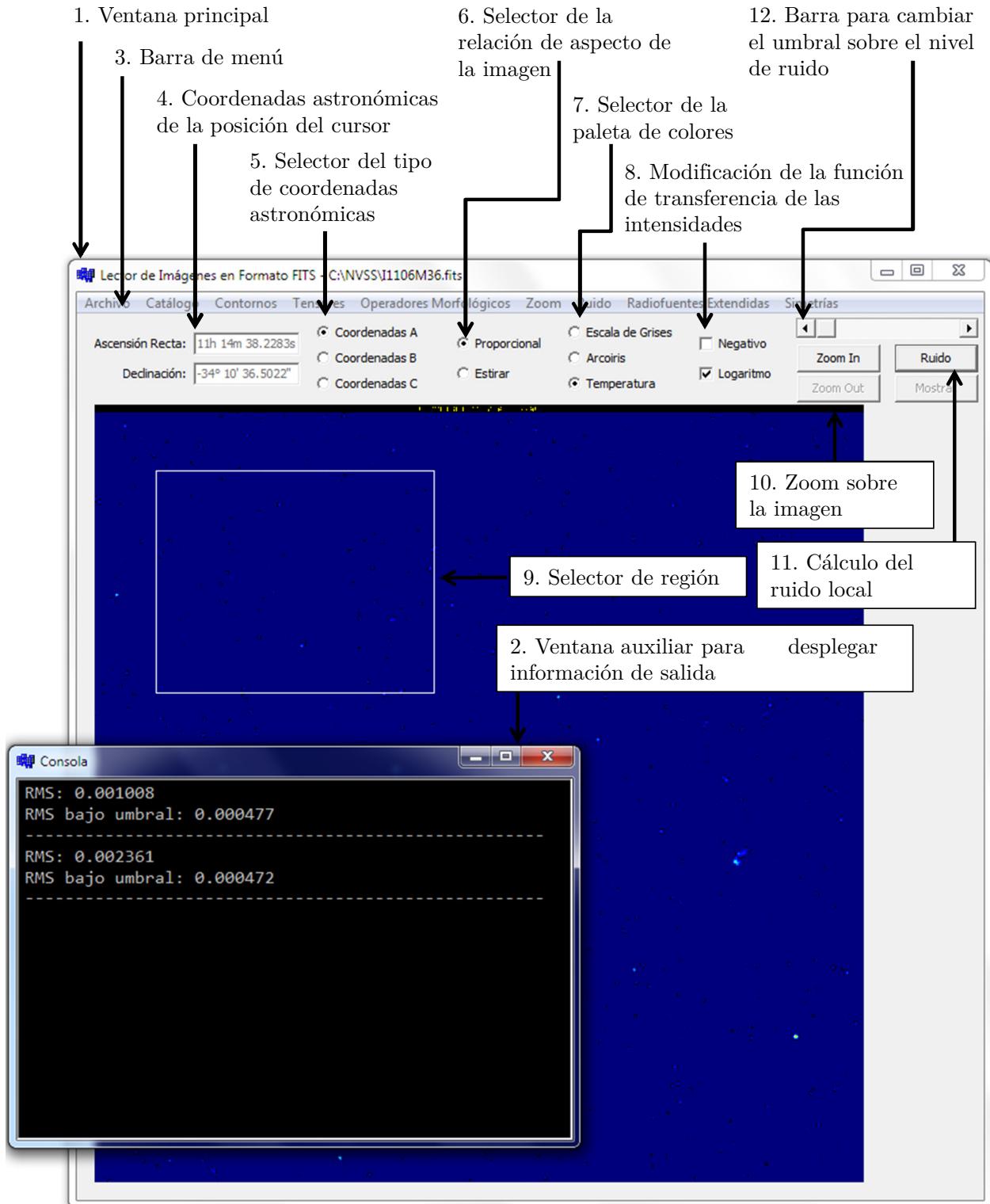
Apéndice B

Descripción del software

Se desarrolló una aplicación para trabajar con la información del NVSS. Se trata de un programa desarrollado en (Borland C++Builder 2005). Se utilizó la librería CFITSIO (Pence 2006) para la lectura de las imágenes en formato FITS. Por facilidad algunos cálculos se hicieron en R (R Core Team 2012) y otros en MATLAB.

En la página siguiente observamos la ventana principal de la aplicación. A continuación se dará una breve descripción de cada una de sus partes.

1. **Ventana principal.** En ella se manipulan las opciones y se despliega la imagen de interés del atlas.
2. **Ventana auxiliar.** Medio de salida de información tipo consola.
3. **Barra de menú.** Para manipular la mayoría de las opciones de la aplicación.
4. **Coordenadas astronómicas.** Para desplegar las coordenadas astronómicas del punto sobre el que está el cursor en la imagen.
5. **Selector del tipo de coordenadas.** Para cambiar entre coordenadas ecuatoriales con α en horas, minutos y segundos y δ en grados, minutos de arco y segundos de arco (A), coordenadas ecuatoriales con α y δ grados con decimales (B) y coordenadas galácticas en grados con decimales (C).



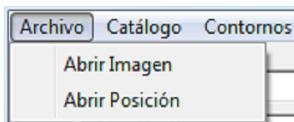
B.1. Barra de menú

6. **Selector de la relación de aspecto.** Para ajustar el tamaño de la imagen a la ventana o mantener la proporción original.
7. **Selector de la paleta de colores.** Para cambiar la escala de colores en la que se grafica la imagen.
8. **Modificación de la función de transferencia.** Para cambiar la paleta a negativo o usar escala logarítmica.
9. **Selector de región.** Para seleccionar una zona rectangular de la imagen con el ratón.
10. **Zoom.** Para hacer *zoom in* en la región seleccionada, o *zoom out* a la región previamente desplegada.
11. **Cálculo del ruido local.** Para estimar el ruido local de la región seleccionada.
12. **Barra para cambiar el umbral sobre el nivel de ruido local.** Barra para cambiar el nivel del umbral que se usa para descartar radiofuentes extendidas en zonas ruidosas.

B.1 Barra de menú

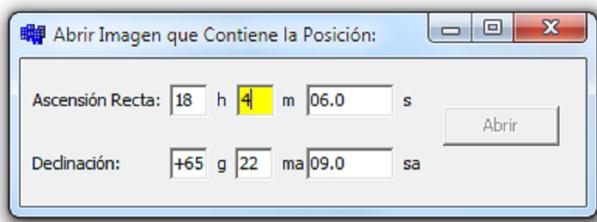
B.1.1 Archivo

Se tiene el siguiente submenú:



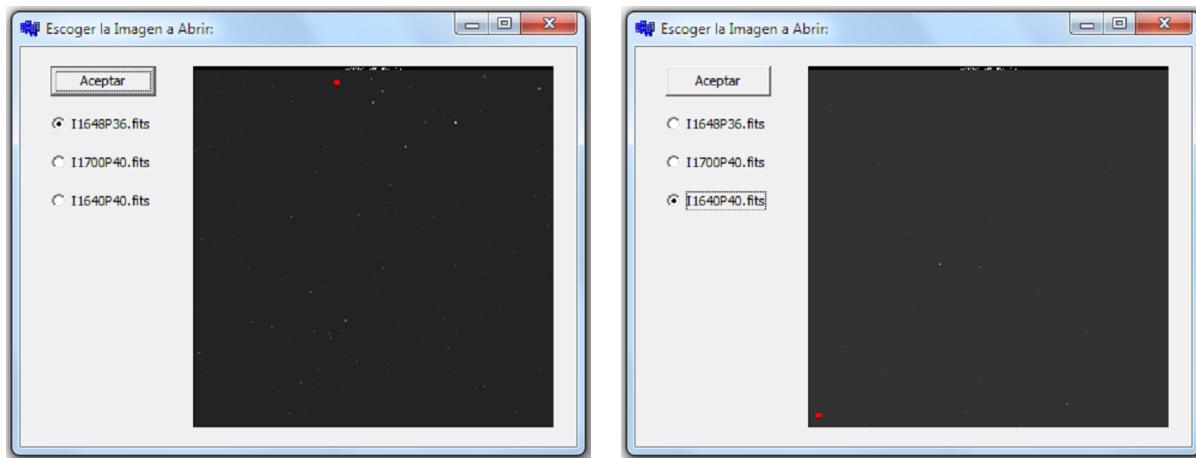
En “Abrir Imagen” seleccionamos el archivo FITS correspondiente a la imagen del NVSS que queremos abrir. En “Abrir Posición” abrimos la imagen que contiene las coordenadas

ecuatoriales (α, δ) dadas. Para capturar las coordenadas (α, δ) usamos la siguiente ventana:



En caso de que haya algún error en el formato de entrada, el recuadro cambia su color de fondo a amarillo.

Cuando la posición proporcionada se encuentre en más de una imagen del NVSS, se muestra una lista de entre la cual se elige la imagen a abrirse. En cada imagen se muestra con un punto rojo el pixel correspondiente a la posición en cuestión.



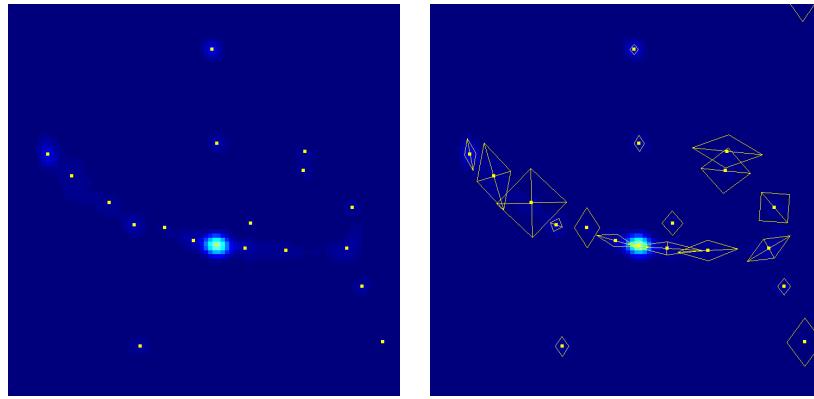
B.1.2 Catálogo

Se tiene el siguiente submenú:



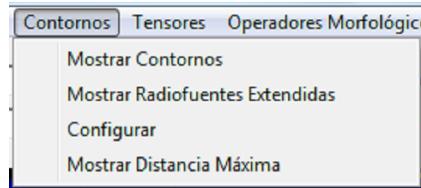
B.1. Barra de menú

Aquí activamos o desactivamos la opción de que aparezcan en la imagen las fuentes del catálogo. “Mostrar Fuentes” corresponde a la versión del catálogo adjunta a cada imagen, “Mostrar Fuentes V2” despliega la versión corregida del catálogo y “Mostrar Fuentes FITS” la versión cruda. Podemos también desplegar los ejes de las gaussianas ajustadas en las tres versiones del catálogo. En la siguiente imagen vemos un ejemplo⁴:



B.1.3 Contornos

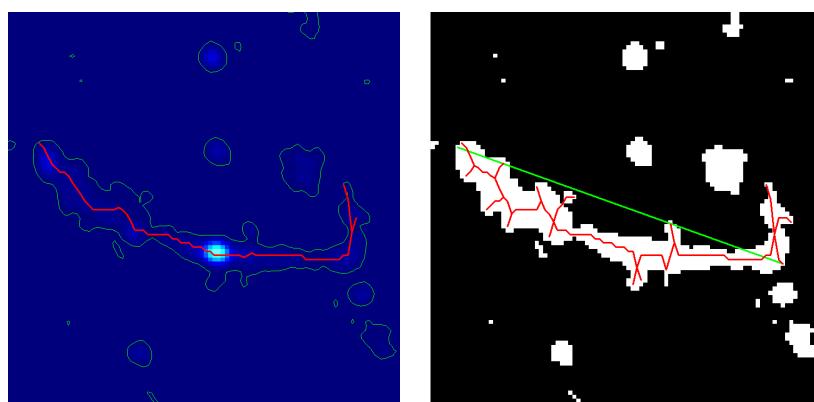
El submenú es el siguiente:



Aquí se permite desplegar en la imagen las curvas de contorno de la función de intensidad y calcular la imagen binaria que se obtiene al aplicar el umbral correspondiente. También se puede mostrar la linea que tiene sus extremos en los dos pixeles más alejados contenido

⁴En este y en los ejemplos posteriores se observa el objeto con centro en J110957-373231.

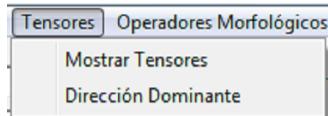
en la radiofuentre extendida seleccionada. En la siguiente imagen vemos ejemplos de esto.



Las curvas de contorno fueron precalculadas en R para cada imagen y exportadas para usarse en C++Builder en una base de datos.

B.1.4 Tensores

Este submenú se ve así:

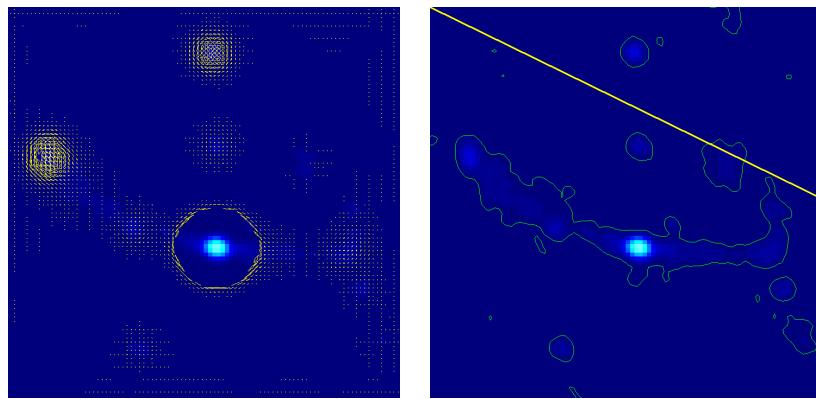


Podemos desplegar los tensores de la imagen⁵, los cuales calculamos guiándonos por (Aja-Fernández *et al.* 2009). También podemos graficar la recta que tiene la dirección dominante estimada (ver Subsección 3.2.2) del objeto seleccionado. En las siguientes imágenes tenemos

⁵Aunque los tensores no fueron utilizados para calcular características en el Capítulo 3, decidimos mencionarlos aquí ya que creemos que son un herramienta con mucho potencial.

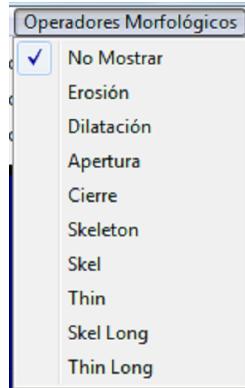
B.1. Barra de menú

un ejemplo de esto.



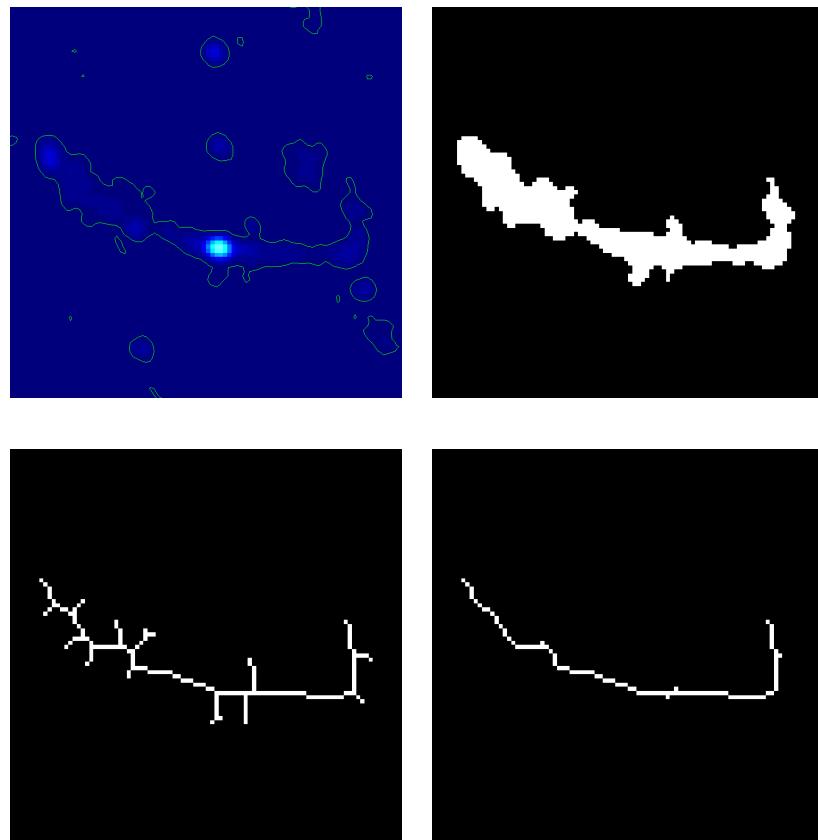
B.1.5 Operadores morfológicos

A través de este menú podemos ver el resultado de varias operaciones aplicadas a la binarización del objeto seleccionado. El submenu es el siguiente:



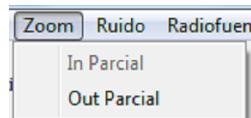
El cálculo del *skeleton* y de la operación *thin* (una variante del *skeleton*) fueron hechos previamente en MATLAB, e importados en una base de datos. La obtención de los caminos más cortos se hizo con un programa de MATLAB. El resto de las operaciones fueron imple-

mentadas directamente. En las siguientes imágenes vemos ejemplos de estas operaciones.

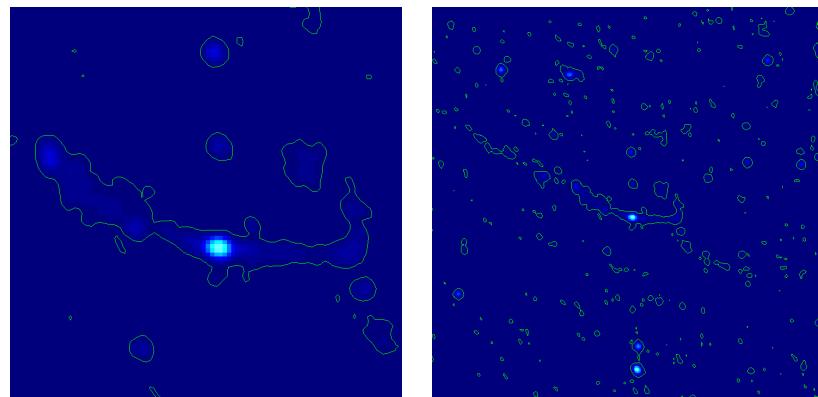


B.1.6 Zoom

Aquí podemos hacer un *zoom in* o *zoom out* parcial, es decir, ir cambiando el *zoom* poco a poco. El menú es el siguiente:



A continuación tenemos un ejemplo de *zoom out* parcial.



B.2 Ruido

En este submenú tenemos las opciones para descartar zonas con exceso de ruido local. El submenú es el siguiente:



Podemos capturar directamente el umbral a usarse, o especificarlo con una barra de desplazamiento (elemento 12 de la ventana principal del programa). También podemos observar la imagen del ruido binarizada de acuerdo al umbral o en escala de grises con extremos negro y blanco correspondientes al mínimo y máximo del ruido en la imagen. A

continuación unos ejemplos:

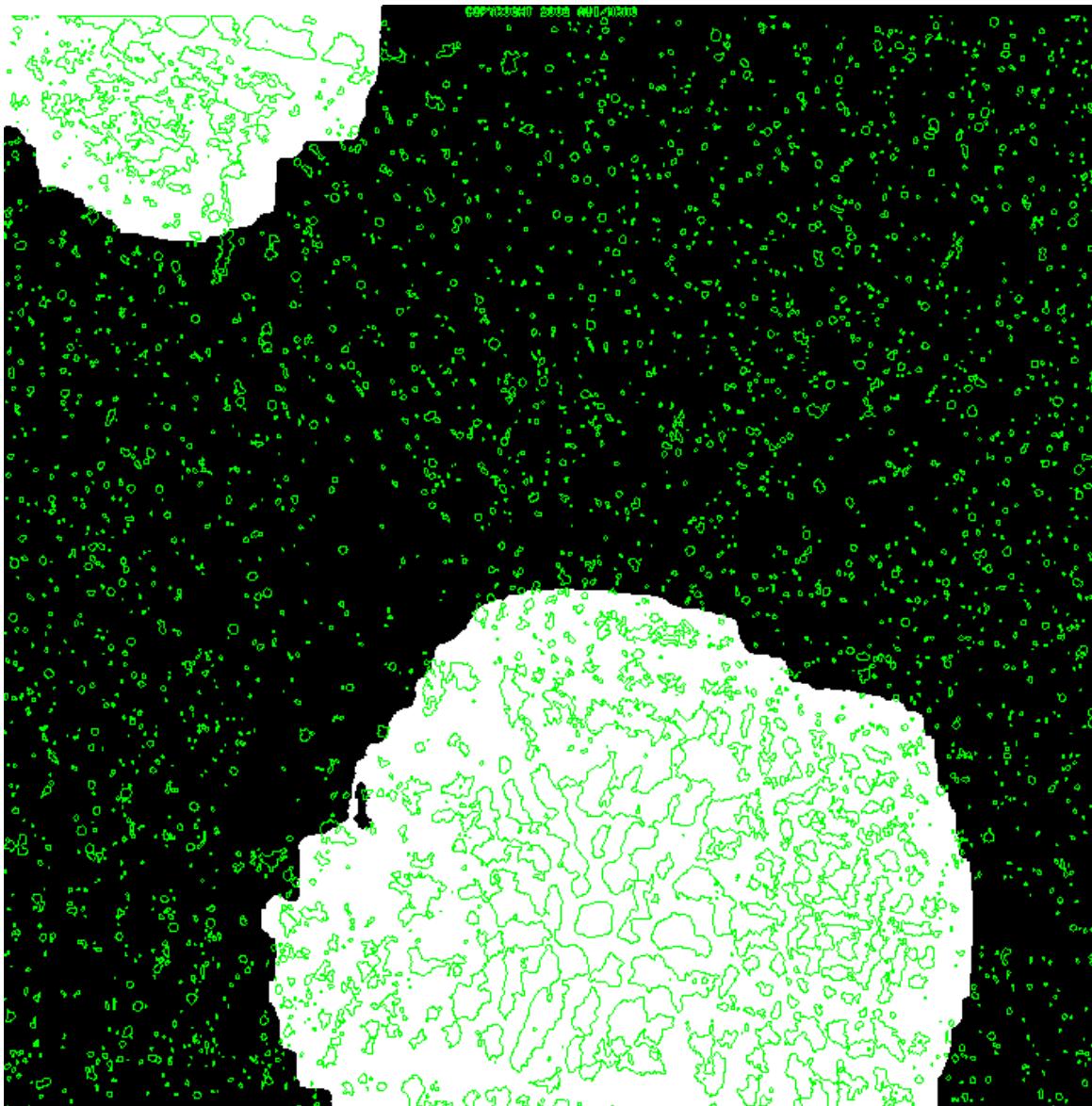


Imagen I0536M04 del NVSS.

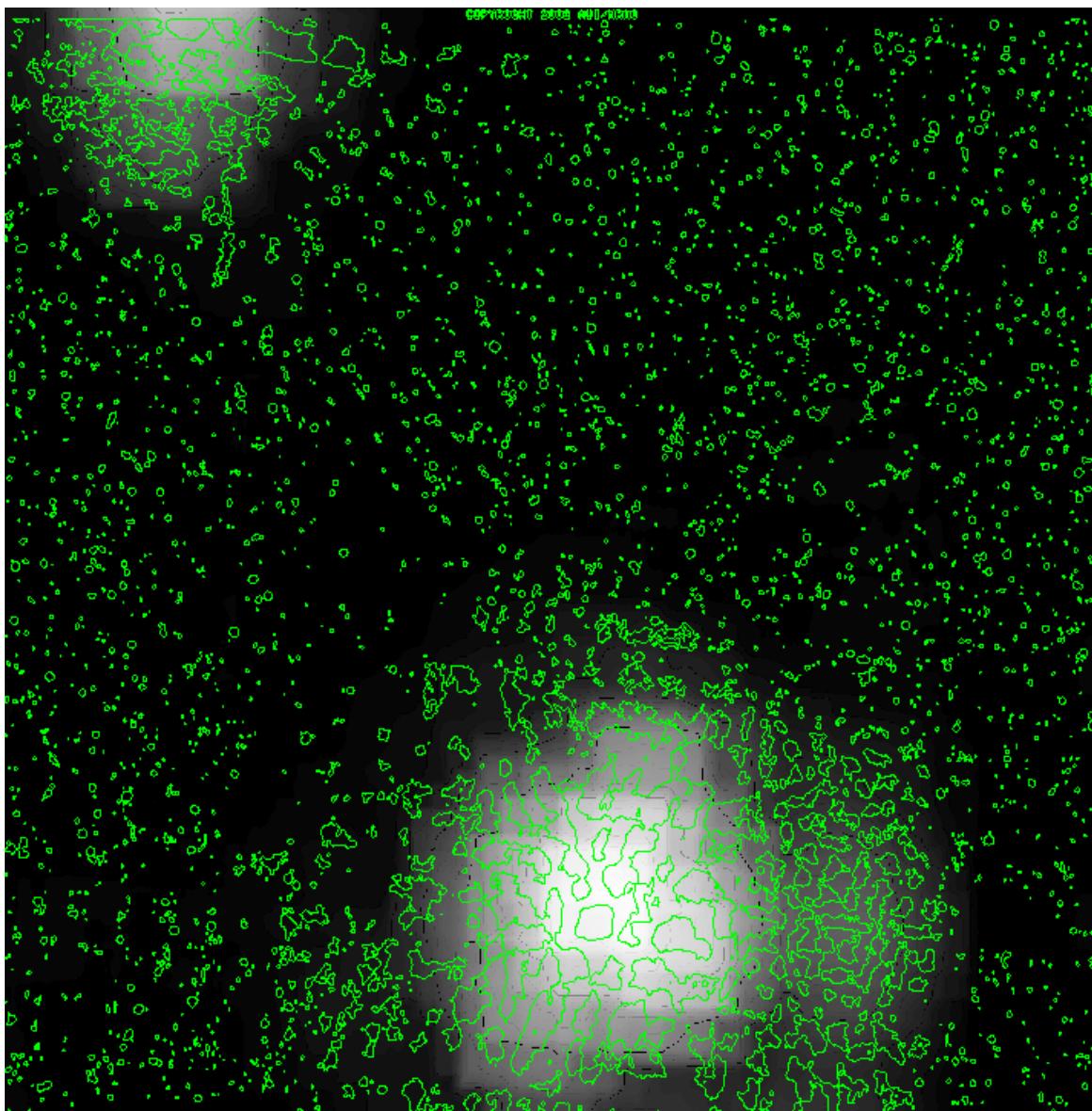
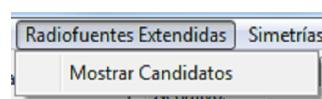


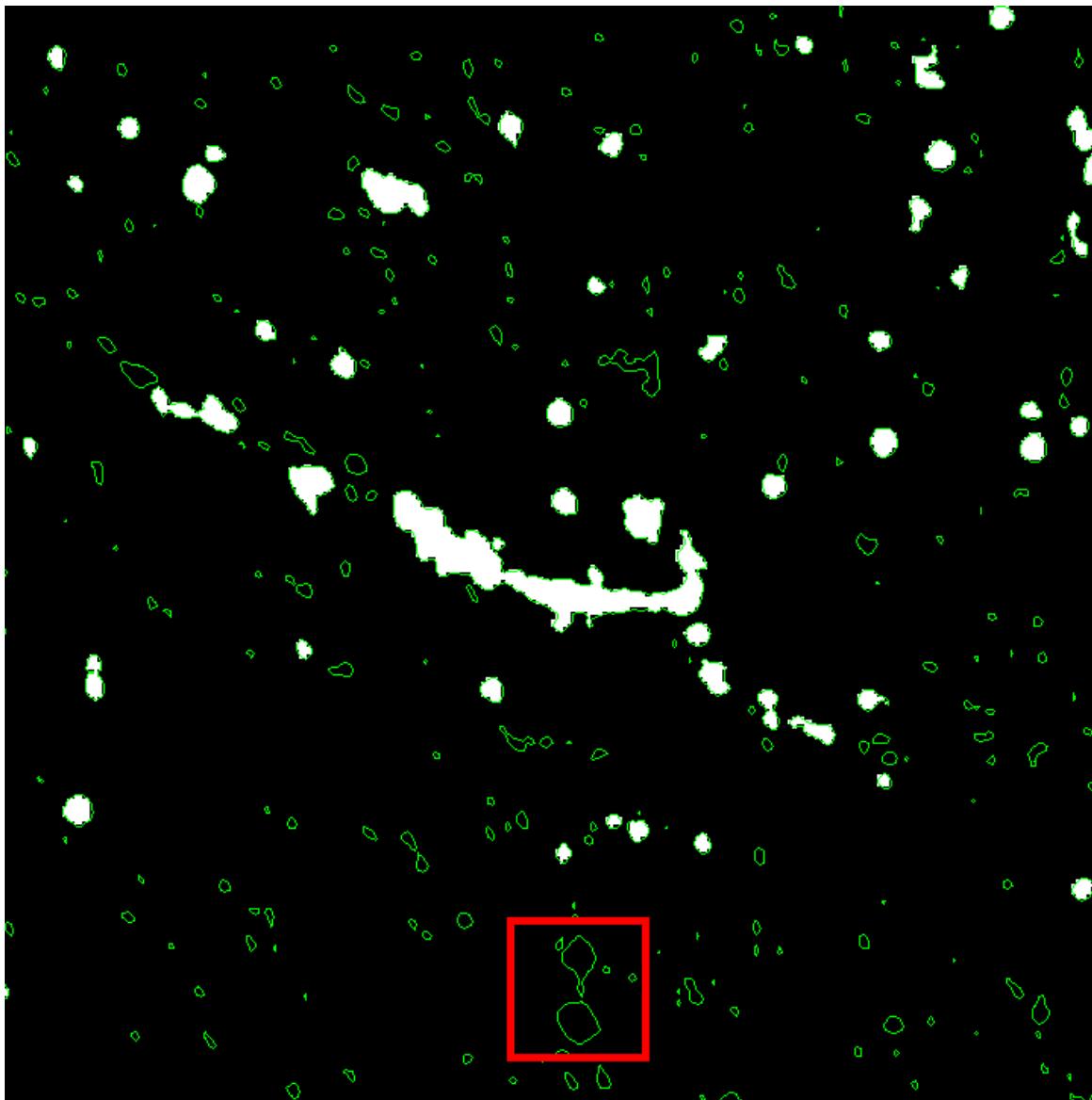
Imagen I0536M04 del NVSS.

B.2.1 Radiofuentes extendidas

En este submenú tenemos una única opción, la cual permite mostrar las radiofuentes extendidas extraídas de la imagen. El submenú se ve así:



Al dar clic en alguna radiofuentе extendida se cargan en memoria sus caracteristicas, que fueron calculadas previamente. Así es posible mostrar su *skeleton*, eje de simetría, dirección dominante, etc. A continuación tenemos una imagen de como se ven las radiofuentes extendidas de una subimagen:



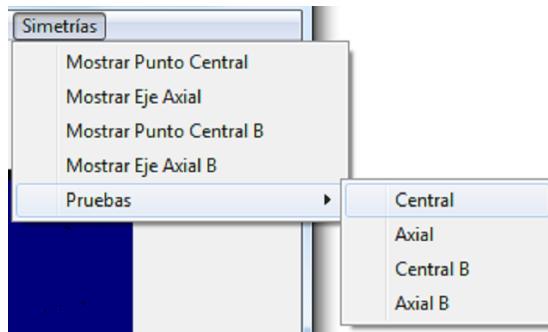
Subimagen de I1106M36 del NVSS.

Se descartan radiofuentes definidas por los contornos que no tienen asociada ninguna fuente del catálogo. Hay algunas radiofuentes extendidas que no se marcan debido a que están duplicadas y están asociadas a otra imagen, por ejemplo, el contorno encerrado en el

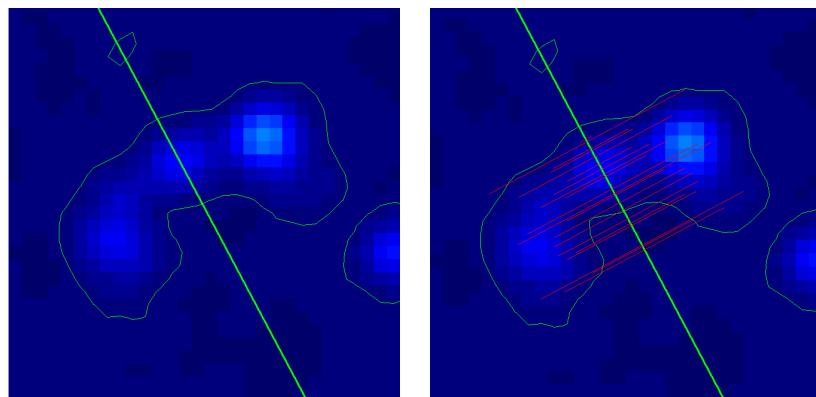
recuadro rojo la parte inferior de la imagen anterior.

B.2.2 Simetrías

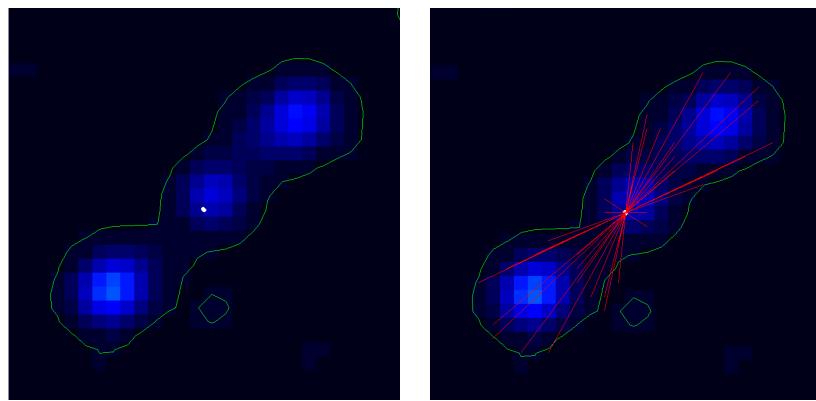
En este submenú se grafican los ejes y puntos de simetría calculados para la radiofuentre extendida seleccionada. El submenú es el siguiente:



Se calculan distintos tipos de simetría axial y central. La simetrías tipo B se calculan sobre la radiofuentes extendida, es decir, no se toman en cuenta las intensidades. La parte de pruebas sirve para graficar algunas de las correspondencias entre los pixeles para cada tipo de simetría. En las siguientes imágenes observamos ejemplos de pruebas axial y central, respectivamente. Las líneas rojas marcan la correspondencia entre los pixeles de acuerdo al tipo de simetría.



Objeto con coordenadas J051850+701320.



Objeto con coordenadas J224055+273421.

Bibliografía

- Aja-Fernández, S., García, R. de L., Tao, D. & Li, X. 2009. *Tensors in Image Processing and Computer Vision*. Advances in Pattern Recognition. Springer-Verlag.
- Becker, R. H., White, R. L. & Helfand, D. J. 1995. The FIRST Survey: Faint Images of the Radio Sky at Twenty Centimeters. *Astrophysical Journal*, **450**, 559–577.
- Bock, D. C.-J, Large, M. I. & Sadler, E. M. 1999. SUMSS: A Wide-Field Radio Imaging Survey of the Southern Sky. I. Science Goals, Survey Design, and Instrumentation. *The Astronomical Journal*, **117**(3), 1578–1593.
- Bonnarel, F., Fernique, P., Bienaymé, O., Egret, D., Genova, F., Louys, M., Ochsenbein, F., Wenger, M. & Bartlett, J. G. 2002. The ALADIN interactive sky atlas. A reference tool for identification of astronomical sources. *Astronomy and Astrophysics Supplement*, **143**, 33–40.
- Borland C++Builder. 2005. *Versión 10.0.2288.42451*. Borland Software Corporation.
- Carroll, B. W. & Ostlie, D. A. 2007. *An Introduction to Modern Astrophysics*. 2nd. edn. Pearson Addison-Wesley.
- Chaisson, E. & McMillan, S. 2001. *Astronomy: A Beginner's Guide to the Universe*. 3rd. edn. Prentice Hall.
- Condon, J. J., Cotton, W. D., Greisen, E. W., Yin, Q. F., Perley, R. A., Taylor, G. B. &

- Broderick, J. J. 1998. The NRAO VLA Sky Survey. *The Astronomical Journal*, **115**(5), 1693–1716.
- de Breucks, C. 2002. A Sample of Ultra Steep Spectrum Sources Selected from the Westerbork In the Southern Hemisphere (WISH) Survey. *Astronomy & Astrophysics*, **394**(1), 56–69.
- de Ruiter, H. R. 1998. Bright Galaxies from WENSS. I. The Minisurvey. *Astronomy and Astrophysics*, **339**, 34–40.
- Delmas, P. 2012. *Digital Image Processing: Selected Topics (Web-Based Lecture Notes)*. <http://www.cs.auckland.ac.nz/courses/compsci773s1c/lectures/ImageProcessing-html/topic4.htm>.
- Joye, W. A. & Mandel, E. 2003. New Features of SAOImage DS9. *Astronomical Data Analysis Software and Systems*, **295**, 489–492.
- Kirshner, S., Cadez, I. V., Smyth, P. & Kamath, Ch. 2002. Learning to Classify Galaxy Shapes Using the EM Algorithm. *Advances in Neural Information Processing Systems (NIPS)*, **15**, 1497–1504.
- Leahy, J. P., Bridle, A. H. & Strom, R. G. 2000. *An Atlas of DRAGNs*. <http://www.jb.man.ac.uk/atlas/dragns.html>.
- Machalski, J., Jamrozy, M. & Zola, G. 2001. The New Sample of Giant Radio Sources. *Astronomy & Astrophysics*, **371**, 445–469.
- Machalski, J., Koziel-Wierzbowska, D., Jamrozy, M. & Saikia, D. J. 2008. J1420-0545: The Radio Galaxy Larger Than 3C236. *The Astronomical Journal*, **679**, 149–155.
- MATLAB. 2011a. *MATLAB Documentation Center*. <http://www.mathworks.com/help/toolbox/images/ref/bwmorph.html>.

- MATLAB. 2011b. *MATLAB Documentation Center*. <http://www.mathworks.com/help/toolbox/images/ref/bwdistgeodesic.html>.
- MATLAB. 2012. *Versión 1.14.0.739 (R2012a)*. Natick, Massachusetts: The MathWorks Inc.
- Miley, G. K. 1980. The Structure of Extended Extragalactic Radio Sources. *Annual Review of Astronomy and Astrophysics*, **18**, 165–218.
- Pence, W. 2006. *CFITSIO - A FITS File Subroutine Library*. <http://heasarc.gsfc.nasa.gov/docs/software/fitsio/fitsio.html>. Versión 3.02.
- Pence, W. & Chai, P. 2009. *Fv: The Interactive FITS File Editor*. <http://heasarc.gsfc.nasa.gov/ftools/fv/>. Versión 5.3.
- Proctor, D. D. 2003. Low Resolution Pattern Recognition - Sorting Triples in the FIRST Database. *J. Electron. Imaging.*, **12**, 398–409.
- Proctor, D. D. 2006. Comparing Pattern Recognition Features Sets for Sorting Triples in the FIRST Database. *The Astrophysical Journal Supplement Series*, **165**, 95–107.
- Proctor, D. D. 2011. Morphological Annotations for Groups in the FIRST Database. *The Astrophysical Journal Supplement Series*, **194**(2), 31–63.
- R Core Team. 2012. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
- Soille, P. 1999. *Morphological Image Analysis: Principles and Applications*. Springer-Verlag.
- Taylor, G. 2006. *The Picture Gallery (Personal Homepage)*. <http://www.phys.unm.edu/~gbtaylor/gallery.html>.
- Wikipedia. 2012. *Active galactic nucleus — Wikipedia, The Free Encyclopedia*. http://en.wikipedia.org/wiki/Active_galactic_nucleus.

- Willis, A. G., Strom, R. G. & Wilson, A. S. 1974. 3C236, DA240: The Largest Radio Sources Known. *Nature*, **250**(5468), 625–630.
- York, D. G. 2000. The Sloan Digital Sky Survey: Technical Summary. *The Astronomical Journal*, **120**, 1579–1587.