

*Desarrollo de una plataforma astroinformática para la administración y análisis inteligente de datos a gran escala*

## **Generación de un índice de objetos astronómicos mediante la identificación y segmentación de imágenes FITS**

Mauricio Solar, Marcelo Mendoza, Jonathan Antognini, Walter Fariña, Gabriel Candia  
Jorge Ibsen, Lars Nyman, Eduardo Vera, Diego Mardones, Guillermo Cabrera,  
Paola Arellano, Karim Pichara, Nelson Padilla, Ricardo Contreras,  
Neil Nagar, Victor Parada.

Santiago, 18 de agosto de 2013

### **Resumen**

El presente informe busca mostrar los actuales avances en procesamiento de imágenes con el fin de poder utilizar los más aptos para identificar objetos en imágenes FITS, segmentarlos y poder guardarlos en una base de datos que posea un índice que permita búsquedas espaciales. Esto se quiere lograr con el fin de otorgar una herramienta a los astrónomos que permita un primer acercamiento a los objetos en estudio o a una cierta parte del cielo, debido a que mediante imágenes FITS será más rápido y menos tedioso que obtener los data sets completos.

**Palabras Claves:** FITS, índice, base de datos, procesamiento de imágenes.

## 1. Resumen Ejecutivo

La capacidad que posee la astronomía para procesar información ya se encuentra sobrepasada, con esto es necesario generar herramientas que permitan discriminar e identificar objetos a estudiar. Con esta premisa se busca generar un índice de objetos utilizando imágenes para así permitirle a los astrónomos tener un primer acercamiento frente a ellos además de encontrarlos con mayor facilidad.

Esto se pretende lograr mediante la segmentación de las distintas imágenes FITS con el fin de separar los distintos objetos que la imagen pueda contener identificando las coordenadas espaciales que estos ocupan. Luego de la identificación de los distintos objetos se guardarán en un índice que permita realizar búsquedas espaciales.

Con grandes cantidades de imágenes puede ayudar a verificar si los objetos son persistentes en el tiempo es decir son realmente objetos a estudiar o si son ruidos o explosiones temporales.

# Índice

|   |           |
|---|-----------|
| <b>1. Resumen Ejecutivo</b>   | <b>2</b>  |
| <b>2. Metodología de Trabajo</b>  | <b>7</b>  |
| <b>3. Definición del Problema</b>   | <b>8</b>  |
| 3.1. Problema . . . . .   | 8         |
| 3.2. Solución Propuesta . . . . .   | 8         |
| 3.3. Objetivos . . . . .  | 8         |
| 3.3.1. Objetivo General . . . . .   | 8         |
| 3.3.2. Objetivos específicos . . . . .                                      | 8         |
| <b>4. Estado del Arte</b>   | <b>9</b>  |
| 4.1. Detección de Bordes . . . . .  | 9         |
| 4.2. Segmentación . . . . .   | 10        |
| 4.3. Reconocimiento de Patrones . . . . .                                   | 10        |
| 4.4. De Imágenes a Catálogos . . . . .                                      | 11        |
| 4.4.1. Estimación del Fondo . . . . .                                       | 12        |
| 4.4.2. Convergencia . . . . .   | 13        |
| 4.4.3. Detección . . . . .  | 13        |
| 4.4.4. Combinación o Unión . . . . .  | 13        |
| 4.4.5. Fotometría y Clasificación . . . . .                                 | 13        |
| 4.5. Transformación de Wavelets . . . . .                                   | 14        |
| 4.6. Modelo de Visión Multiescala . . . . .                                 | 15        |
| 4.6.1. Definición del Modelo . . . . .                                      | 15        |
| 4.6.2. Identificación de objetos mediante Coeficientes de Wavelet . . . . . | 16        |
| 4.6.3. Reconstrucción Parcial . . . . .                                     | 18        |
| 4.7. Herramientas para procesar imágenes . . . . .                          | 19        |
| 4.8. Observatorios Astronómicos Virtuales . . . . .                         | 20        |
| <b>5. Planificación de Trabajo</b>  | <b>21</b> |
| 5.1. Alcance del proyecto . . . . .   | 21        |

|  |           |
|--|-----------|
| 5.1.1. Procesamiento de imágenes . . . . .   | 21        |
| 5.1.2. Base de datos . . . . .               | 21        |
| 5.1.3. Índice de objetos . . . . .           | 21        |
| 5.2. Carta Gantt . . . . .                   | 22        |
| 5.3. Explicación hitos importantes . . . . . | 22        |
| <b>Bibliografía</b>                          | <b>24</b> |

## Índice de figuras

|    |   |    |
|----|---|----|
| 1. | Bordes obtenidos de Saturno. . . . .                      | 9  |
| 2. | Ejemplo de dato astronómico. . . . .                      | 12 |
| 3. | Ejemplo de conectividad en el espacio de wavelet. . . . . | 16 |
| 4. | Ejemplo de extracción de estrella. . . . .                | 19 |
| 5. | Carta Gantt del proyecto. . . . .                         | 22 |

## Índice de cuadros

|                                      |    |
|--------------------------------------|----|
| 1. Actividades del proyecto. . . . . | 22 |
|--------------------------------------|----|

## 2. Metodología de Trabajo

La metodología elegida para el desarrollo del tema es partir por el ámbito teórico que consta de toda la investigación acerca de los distintos tópicos de interés como lo son el procesamiento de imágenes, los formatos de imágenes, las bases de datos utilizadas, los softwares involucrados y las técnicas utilizadas. Así como también investigar las distintas técnicas que están utilizando otros observatorios o sedes astronómicas.

Luego de la etapa de investigación se continuará con la elección e implementación de las distintas técnicas y herramientas elegidas.

Para finalmente poner a prueba la implementación y así poder validar la investigación realizada, para esto se utilizarán imágenes de prueba y se corroborarán los objetos obtenidos.

Cabe destacar que en primera instancia el procesamiento de imágenes será sobre imágenes ópticas para luego una vez que estén funcionando procesar imágenes de radio donde será necesario aplicar una transformación de Fourier antes de procesar la imagen.

## **3. Definición del Problema**

### **3.1. Problema**

Con el continuo aumento en las cantidades de imágenes y datos astronómicos obtenidos por los distintos observatorios, el manejo de estos se ha vuelto cada vez más complejo y a mayor escala. Es por esto que cada vez se hace más necesario el desarrollo de nuevas herramientas que permitan solucionar o aliviar estos problemas.

Dentro de este gran problema se identifico uno en particular a resolver. Cuando un astrónomo quiere investigar acerca de algún objeto o alguna parte del cielo en especial tiene que recurrir a solicitar los data sets mediante algunas queries en una base de datos. Por lo que su primer acercamiento frente a este no es del todo rápido y simple.

Dentro de los data sets solicitados puede encontrar tablas con grandes cantidades de información como también las imágenes FITS correspondientes, pero para poder ver estas imágenes necesita descargar los data sets completos.

### **3.2. Solución Propuesta**

La solución propuesta para generar un primer acercamiento frente a los objetos astronómicos es la generación de un índice de objetos astronómicos mediante imágenes FITS. Un índice que permita realizar búsquedas espaciales y poder desplegar las imágenes de interés.

### **3.3. Objetivos**

#### **3.3.1. Objetivo General**

Diseñar e implementar una herramienta de software que permita generar un índice de imágenes astronómicas, basado en sus coordenadas espaciales en una base de datos acorde a las necesidades.

#### **3.3.2. Objetivos específicos**

- Investigar sobre las imágenes FITS, con el fin de aprovechar toda la información contenida en la imagen incluyendo su metadata.
- Investigar y elegir una base de datos acorde para guardar los objetos astronómicos.
- Segmentar e identificar objetos que contengan las imágenes.
- Generar un índice de objetos astronómicos basado en sus coordenadas espaciales.



## 4. Estado del Arte

### 4.1. Detección de Bordes

Un borde es definido como una variación local de la intensidad de la imagen, algunos cambios de intensidad pueden ser muy grandes lo que facilita identificar de mejor manera los bordes, pero otros pueden ser muy suaves y ser difíciles de detectar.

El fin de esta detección es producir un dibujo de líneas desde una imagen más compleja, esto debido a que importantes conclusiones pueden obtenerse mediante solo los bordes de una imagen para luego dar paso a algunos algoritmos de computación que permitan la identificación de la imagen.

Algunos pasos a grandes rasgos que permiten la correcta detección de bordes se muestran a continuación [1]:

- *Suavizar*: Suprimir la mayor cantidad de distorsión posible, teniendo en cuenta la importancia de no suprimir los bordes.
- *Aumentar*: Aplicar un filtro que permita resaltar la calidad de los bordes de la imagen y agudizarlos.
- *Detectar*: Determinar que píxeles de borde deben ser descartados por ser ruido y cuáles deben ser mantenidos, puede variar mucho dependiendo del criterio de detección.
- *Localizar*: Determinar la posición exacta del borde, agudizar y conectar los bordes también forma parte de esta etapa.

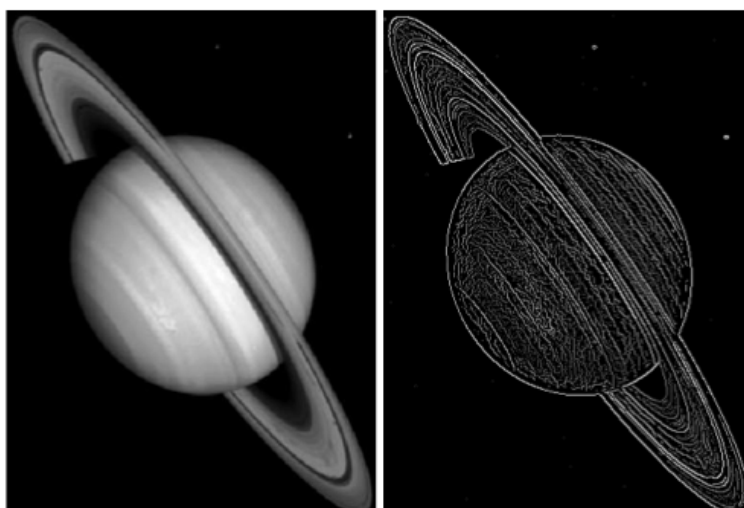


Figura 1: Bordes obtenidos de Saturno.

## 4.2. Segmentación

La segmentación de imágenes es un proceso en el cuál se particiona una imagen en regiones o segmentos basado en similitudes dentro de la región o diferencias entre regiones. Una imagen representa una escena en donde participan distintos objetos o regiones, el proceso de separación puede ser difícil de automatizar.

Existen tres acercamientos generales acerca de la segmentación de imágenes:

- *Clasificación por pixel*: Técnica que evalúa pixel por pixel la posición donde se encuentra y que nivel de gris o color (dependiendo el tipo de imagen) posee. Se usa el término clasificación por que se considera que las distintas regiones poseen pixeles de diferentes clases.
- *Método basado en límites*: Técnica que detecta los límites de las regiones, los píxeles encerrados en un margen pueden ser categorizados acorde a este.
- *Método de crecimiento de regiones*: Técnica basada en la identificación de grupos de píxeles de valores similares, este agrupamiento se realiza generalmente de forma iterativa y permite una relajación acorde al caso.

Existen muchos otros métodos para segmentar una imagen dado que no existe un consenso en esta área de procesamiento de imágenes.

## 4.3. Reconocimiento de Patrones

El reconocimiento de patrones es un área de estudio donde se busca tomar decisiones de manera automática, por lo general se toma una imagen en forma de matriz donde se revisa si los vectores o columnas cumplen con algún criterio de selección. Dado que es necesario conocer que número de coincidencias posee el vector en estudio con el criterio se dice que los patrones son una clase de clasificadores, en esta clasificación es donde el creador del reconocimiento de patrones especifica que porcentaje de similitud debe poseer la imagen o señal en estudio para clasificarla como tal.

En gran parte de los casos de estudio es necesario realizar un arreglo de la imagen o de parte de ella con el fin de reducir los ruidos o distorsiones que ésta pueda tener, dado que podría llevar a conclusiones erróneas si no se procesa la imagen de forma correcta. En general de la imagen se obtiene una matriz con información donde de ésta se deriva a otra del mismo estilo pero dejando de lado toda la información superflua o detalles irrelevantes. Este proceso es denominado “*feature extraction*” [9].

Los componentes de un vector que posee un patrón son comúnmente llamados *features*, características o atributos. Para una correcta extracción de estas características existen dos tipos de enfoque, extracción de bajo nivel y de alto nivel.

La extracción de bajo nivel corresponde por lo general a aplicar una transformación al vector o matriz en estudio para potenciar la separación de clases o reducir la dimensión de la información. Esto con el fin de hacer las tareas computacionales más manejables. Algunos esquemas de compresión también son denominados extracción de bajo nivel. Algunos ejemplos de transformaciones usadas son la transformada de *Fourier* y la transformada de *Karhunen-Loève*.

Por otra parte la extracción de alto nivel tiene como objetivo ayudar de una formas más evidente a sacar conclusiones, un ejemplo podría ser el reconocer formas de una imagen donde el vector podría contener información como largo, grosor, entre otros. La extracción de alto nivel es más tediosa que la de bajo nivel y utiliza herramientas más poderosas.

La extracción de características lo que busca es describir y discriminar los atributos del patrón como también hacer un clasificador más simple. Dentro de los clasificadores se pueden dividir en dos grupos, los supervisados y los no supervisados. El clasificador supervisado posee un mayor conocimiento sobre el objeto en estudio y utiliza un conjunto de parámetros de entrenamiento que definen el clasificador, por otra parte los clasificadores no supervisados se tiene un menor conocimiento del tema y un ejemplo podría ser la segmentación de imágenes.

Cuando nos referimos a un clasificador hay que tener en cuenta que maneja dos tipos distintos de datos, datos de entrenamiento y de prueba. Los datos de entrenamiento permiten calibrar el clasificador y fijar sus parámetros, por otra parte los datos de prueba son los que permiten revisar si es que el clasificador entrenado sirve y se puede utilizar con nuevos datos.

#### 4.4. De Imágenes a Catálogos

Extraer información de las imágenes es un paso fundamental para los astrónomos. Por ejemplo para construir catálogos, es necesario identificar correctamente estrellas y galaxias, tanto como realizar una estimación precisa sobre su posición y fotometría <sup>1</sup>. Varios métodos han sido propuestos para lograr esos resultados. Uno de los softwares más usados para lograr este cometido es SExtractor creado por Bertin and Arnouts en 1996 [5]. Posee la habilidad de manejar grandes imágenes (de hasta 60000 x 60000 píxeles) y lograr buenos resultados, por lo general los software incluido SExtractor realiza los siguientes pasos:

1. Estimación del Fondo
2. Convergencia
3. Detección
4. Combinación
5. Fotometría y Clasificación

---

<sup>1</sup>Rama de la astronomía que se dedica a medir el brillo de los distintos astros.

En la *Figura 2* se muestra un ejemplo de dato astronómico, donde se observa una fuente particular y otra fuente extensa, además de poseer ruido y un fondo. El objeto extendido que es detectado por el ojo humano podría no ser detectado por un mecanismo de detección estándar.

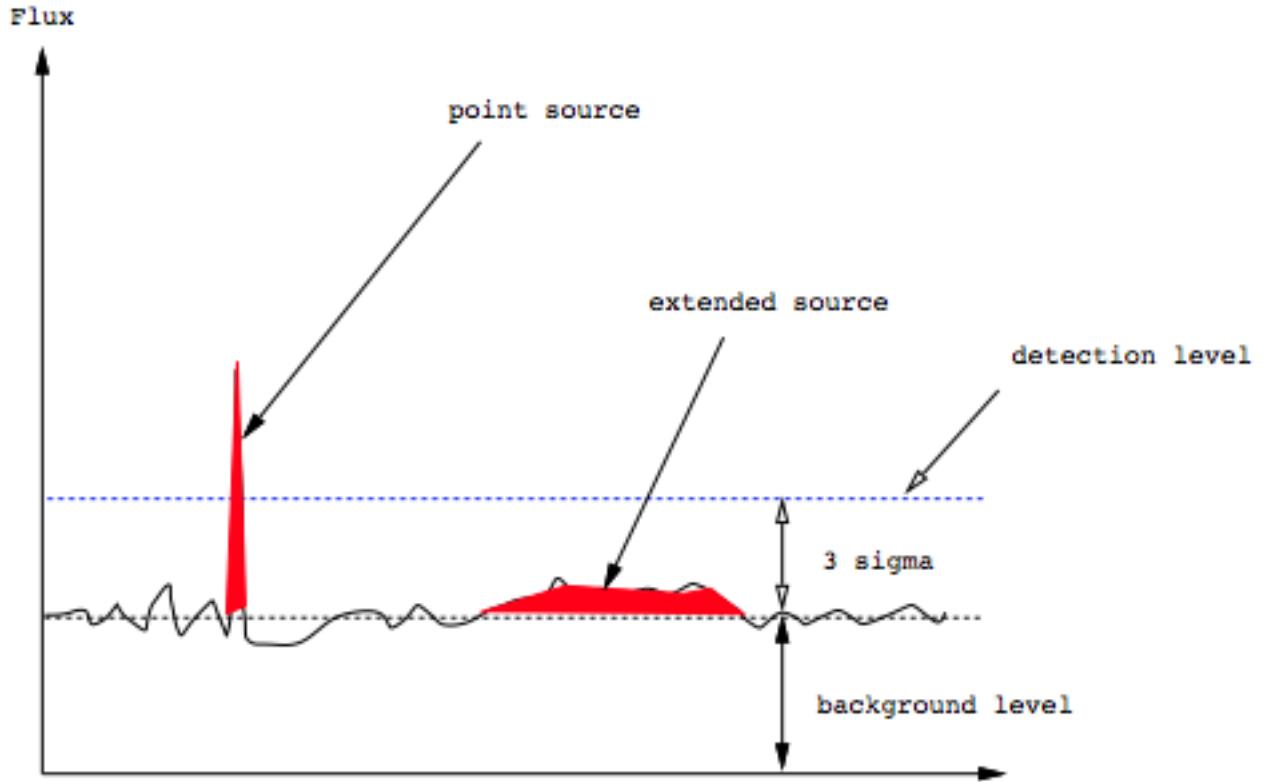


Figura 2: Ejemplo de dato astronómico.

#### 4.4.1. Estimación del Fondo

Por lo general los objetos en estudio pasan desapercibidos por pertenecer a una señal plana, a esto se le llama señal de fondo. En los estudios del universo pequeñas variaciones del fondo modifican de manera importante la intensidad y los diámetros de los diferentes astros, para obtener buenos resultados es fundamental obtener la magnitud correcta en cada punto del fondo del cielo. En algunos píxeles la medición de intensidad puede ser menor al fondo del cielo, esto es debido a la granularidad de las emulsiones provenientes del cielo. Bajo el caso anterior se hace casi imposible determinar directamente la magnitud que posee el fondo del cielo.

Muchos métodos tratan de solucionar el problema del fondo, algunos de estos tienen el defecto de usar información de los píxeles cercanos lo que lleva a una pérdida de resolución. En general el mayor problema es identificar las diferencias entre las medidas de intensidad y la verdadera intensidad, donde es vital tener en cuenta la granularidad [8].

#### 4.4.2. Convergencia

Para optimizar la detección es necesario que la imagen se haga convergir mediante algún filtro. La forma del filtro optimiza la detección de objetos de forma similar. Por esto mismo para detectar estrellas el filtro óptimo es PSF<sup>1</sup>. Para objetos extensos, se requieren filtros más largos. De esta forma para detectar objetos de diferentes formas, se debe probar la detección con varios tamaños de filtro, tendiendo a un acercamiento multiescala.

#### 4.4.3. Detección

Luego de que la imagen ha convergido, es decir los puntos significantes de un objeto han sido identificados, se segmenta la imagen mediante su *thresholded*<sup>2</sup> asignando a cada grupo de píxeles conectados una etiqueta, para luego separar los objetos conectados y que poseen la misma etiqueta.

Una alternativa a usar la técnica de *thresholded* es detectar los montes de la imagen, cabe destacar que esta técnica funciona bien para detectar estrellas pero no para objetos más extensos. Para el caso de objetos extensos el siguiente paso sería unir los píxeles que pertenecen al mismo objeto.

#### 4.4.4. Combinación o Unión

Este es el paso más delicado, los objetos extensos deben ser considerados como un solo objeto y los múltiples objetos deben ser bien separados. En el software SExtractor cada grupo de píxeles conectados es analizado a distintas intensidades, partiendo con una intensidad alta y luego disminuyendo. El grupo de píxeles puede ser visto como una superficie con valles y montañas. Al comienzo solo las grandes cúmbres son visibles y a medida que va disminuyendo otros objetos comienzan a ser visibles, definiendo varias estructuras. A un nivel determinado dos objetos pueden parecer conectados, ahí es donde se debe tomar una decisión si unirlos o dejarlos separados. La forma de evaluarlo es comparando las intensidades dentro de las cúmbres, si el radio entre ellas es muy bajo, se unen. De lo contrario se dejan como objetos independientes.

#### 4.4.5. Fotometría y Clasificación

Existen varios métodos que permiten calcular la fotometría de los objetos detectados. Un fotómetro con apertura adaptativa usa la primera imagen para determinar de que forma se obtiene mayor cantidad de flujo. *Kron* propone una apertura del doble de la primera imagen tomada, que lleva a una recuperación de flujo mayor al 90 % [9]. Dependiendo los objetos a clasificar que métodos se utilizarán, por ejemplo cuando una imagen contiene solo estrellas se especifica un método que puede contener la técnica PSF.

Cuando se quiere clasificar una estrella o galaxia hay una serie de parámetros que se han tenido en cuenta para hacerlo. Máxima intensidad, área radial, densidad de gradiente, momentos invariantes por

---

<sup>1</sup>Point Spread Function describe la respuesta de una imagen a un punto objetivo.

<sup>2</sup>Método de segmentación que utiliza escala de grises para identificar bordes.

mencionar algunas. Dada la gran cantidad de parámetros que hay para elegir existen diferentes puntos de vista o enfoques. Enfoques como la calidad de los datos, la facilidad que posee la computación para medir ciertos parámetros, la relevancia e importancia de los parámetros dependiendo el tipo de salida que se quiere lograr, van gatillando que parámetros usar.

Desde el punto de vista de algoritmos multivariados, se necesita una muestra homogénea de parámetros. Sabiendo esto y conociendo la gran cantidad de métodos para clasificar una galaxia, dos enfoques para seleccionar parámetros parecen ser fuertemente representados.

1. La luminosidad de los objetos a lo largo de los ejes mayores se determina en intervalos discretos. Esto puede realizarse rellenando los contornos elípticos y estudiando la luz de los anillos elípticos<sup>1</sup>. Para lograr ver el perfil de los radios presentes en la galaxia se puede asumir que la galaxia es simétrica con los ejes en estudio y que puede ser ajustada dependiendo los radios en estudio.
2. Otra opción a definir el perfil es definir la morfología de los parámetros. Puede ser interesante utilizar un modelo que logre definir una galaxia, medir el brillo de la superficie, estudiar el diámetro logarítmico que posee la galaxia, entre otros con el fin de definir futuros parámetros a ser utilizados.

El rango de parámetros a utilizar debe ser consecuente con la complejidad del objeto en estudio. Los parámetros pueden ser derivados de un perfil, pero los parámetros no pueden definir un perfil a estudiar.

#### 4.5. Transformación de Wavelets

La Transformación de Wavelets permite separar la imagen en varios componentes, los cuáles se denominan wavelets de la imagen. Una imagen de una galaxia puede estar formada por las wavelets más amplias, donde en gran parte puede estar compuesta por píxeles negros pertenecientes al fondo. Pero también hay wavelets con mayor detalle y nitidez que pueden pertenecer a una estrella de la galaxia. La imagen en estudio puede ser descompuesta en la cantidad de wavelets que se requiera dependiendo los detalles esperados. Cada wavelets no es una imagen en sí, pero varias wavelets pueden recrear una, de hecho se pueden identificar las wavelets que poseen mayor cantidad de información y utilizarlas para recrear una imagen, dejando de lado algunas wavelets que puedan tener información irrelevante, de esta forma la transformación de wavelets también sirve para comprimir imágenes [6].

Para aplicar la transformación de wavelets a una imagen, se requiere tener la imagen en su forma matricial. En la matriz de la imagen se encuentran los píxeles de ésta, donde sus valores representan la oscuridad de un pixel, siendo 0 el blanco y el valor más alto el negro. Los píxeles de valor central pertenecen a los grises.

Al aplicar la transformación de wavelets a la matriz genera tantos coeficientes como píxeles posee la imagen. Estos coeficientes son más fáciles de comprimir dado que la información esta concentrada en sólo

---

<sup>1</sup>Método utilizado por *Lefèvre en 1986*.

algunos de éstos. A su vez a estos coeficientes se les puede aplicar un *threshold* donde todos los coeficientes bajo éste se cambian por 0, dado que son coeficientes que no poseen información importante.

El uso de algunos de los métodos de la transformación de wavelets serán explicados cuando se muestre su uso.

## 4.6. Modelo de Visión Multiescala

La transformación multiescala de una imagen produce una matriz con tantos coeficientes como píxeles posea la imagen. La imagen original puede ser expresada por la suma de todas las escalas wavelets más la suma de un arreglo suavizado como se muestra en la *ecuación 1*:

$$I(k, l) = C_{J,k,l} + \sum_{j=1}^J W_{j,k,l} \quad (1)$$

Como la imagen es interpretada como una representación multiescala de píxeles se debe considerar poder representar los objetos presentes en ésta, es decir identificar la representación multiescala del objeto. Claramente la representación depende de la imagen a analizar. Un modelo para representar datos astronómicos es asumir que la imagen  $I$  puede ser descompuesta en varios componentes:

$$I(k, l) = \sum_{i=1}^{N_o} O_i(k, l) + B(k, l) + N(k, l) \quad (2)$$

En la *ecuación 2*  $N_o$  es el número de objetos,  $O_i$  son los objetos presentes en la información (estrellas, galaxias, etc),  $B$  es el fondo o background y  $N$  el ruido. Para lograr esta descomposición es necesario extraer de forma correcta toda la información de la imagen. Se analiza la representación multiescala y se va segmentando escala por escala. En general para poder identificar un objeto se analiza su conectividad, dado que ocupa una región física, y en esta región se pueden unir píxeles basados en la adyacencia de ellos. El espacio Wavelet de una imagen 2D queda representado en un espacio 3D, por lo que las conexiones en 2D deben ser transformadas a WTS<sup>1</sup>. De esta forma para identificar los objetos debemos analizar los píxeles presentes en el WTS.

### 4.6.1. Definición del Modelo

El Modelo de Visión Multiescala (MVM), describe un objeto mediante un set de estructuras jerárquicas y usa las siguientes definiciones:

- *Coficientes Wavelets Significantes*: Son significantes cuando sobrepasan un umbral dado dependiendo el ruido que posea la imagen.

---

<sup>1</sup>Wavelet Transform Space.

- *Estructura*: Una estructura  $S_{j,k}$  es un set de conexiones de coeficientes significantes en la misma escala  $j$ , donde  $p$  es el número de coeficientes significantes en la estructura y  $W_{j,x_i,y_i}$  es el coeficiente de wavelet en la escala  $i$  en la posición  $(X_i, Y_i)$ , como muestra la *ecuación 3*:

$$S_{j,k} = \{W_{j,x_1,y_1}, W_{j,x_2,y_2}, \dots, W_{j,x_p,y_p}\} \quad (3)$$

- *Objeto*: Un objeto es un set de estructuras, se define el operador  $L$  el cuál indica a que objeto pertenece la estructura, si pertenece igual a  $l$  de otra forma a 0.

$$O_l = \{S_{j_1,k_1}, \dots, S_{j_n,k_n}\} \quad (4)$$

- *Escala de Objeto*: La escala del objeto está dada por su máximo coeficiente de wavelet.
- *Relación interescala*: Criterio que permite conectar dos estructuras en un mismo objeto.
- *Sub-Objeto*: Aparece cuando un objeto posee un máximo de wavelet local. Por lo que puede ser compuesto por varios sub-objetos. Cada sub-objeto puede ser analizado.

#### 4.6.2. Identificación de objetos mediante Coeficientes de Wavelet

**Multiresolución soporta segmentación.** Al poseer la imagen en multiresolución se tiene a cada escala una imagen booleana, es decir cuando los coeficientes son significantes son igual a 1, de lo contrario a 0. La segmentación consiste en etiquetar las escalas booleanas. Cada grupo de píxeles conectados obtiene una etiqueta entre 1 y  $L_{max}$ , donde  $L_{max}$  es el número de grupos. Este proceso se repite en cada escala de multiresolución. Se define una estructura  $S_{j,i}$  como el grupo de conexión de los píxeles significantes con la etiqueta  $i$  en la escala  $j$ . En la *figura 3* se muestra la conectividad en el espacio de wavelet; coeficientes significantes de wavelet que aparecen contiguos forman una estructura y mediante una relación interescalar un set de estructuras forman un objeto.

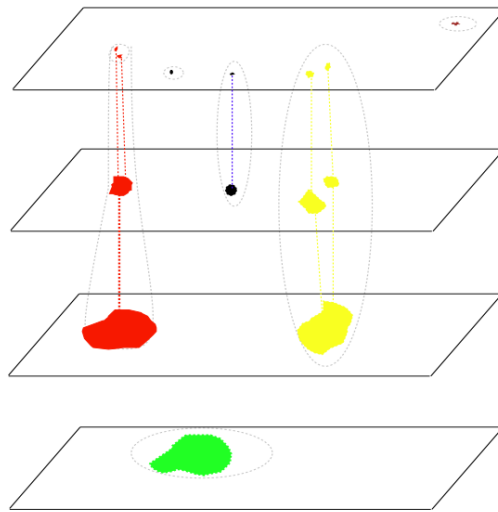


Figura 3: Ejemplo de conectividad en el espacio de wavelet.



**Grafo de conectividad interescala.i** Un objeto es definido como un set de estructuras jerárquicas. La regla que permite la conexión entre dos estructuras es llamada “*relación interescala*”. En la figura se muestra como estructuras de diferentes escalas son unidas y forman objetos. Consideremos dos estructuras sucesivas,  $S_{j,k}$  y  $S_{j+1,l}$ . Cada estructura está localizada en parte de la descomposición y corresponden a una región de la imagen donde los coeficientes son significantes. El pixel que posee el coeficiente de wavelet máximo debe estar presente en ambas estructuras para poder conectarlas. Las estructuras que aparecen en imágenes con coeficientes de wavelets sucesivos pueden ser conectados y se denominan objetos pertenecientes al grafo de conectividad interescala.

**Filtro.** Estadísticamente algunas estructuras pueden ser generadas debido al ruido. Por lo general contienen pocos píxeles y están aislados. Por lo que para evitar detecciones falsas los campos aislados pueden ser removidos del grafo de conexión interescalar. Las estructuras detectadas en los bordes de la imagen pueden ser generadas por problemas en el borde y también pueden ser removidas.

**Unión o Separación.** Objetos que se encuentran muy cerca pueden estar conectados o ser diferentes, en estos casos una decisión debe ser tomada. En el espacio de wavelet la decisión de unir o separar está basada en los máximos locales de las diferentes estructuras que pertenecen al objeto. Un objeto nuevo es derivado de una estructura si existe al menos una otra estructura en la misma escala perteneciente al mismo objeto y cumple con lo siguiente:

- El pixel anterior al máximo de wavelet de la estructura tiene valor 0 y no esta conectado a ninguna estructura de una escala menor.
- El máximo coeficiente de wavelet de la estructura  $S_{j-1,l}$  es el más cercano al máximo de la estructura  $S_{j,k}$ .
- Los coeficientes máximos de la estructura  $S_{j+1,k}$  pertenecen también a la estructura  $S_{j,k}$ .

Cuando se cumplen estas condiciones la estructura  $S_{j,k}$  y todas las estructuras en escalas menores que están directa o indirectamente conectadas con  $S_{j,k}$  definirán un nuevo objeto.

**Identificación de Objetos.** Resumiendo el método podemos identificar los objetos presentes en la imagen  $I$ .

1. Se aplica la transformación de wavelet  $W = WI = \{w_1, \dots, w_j, c_j\}$  que genera el espacio de wavelet del mismo tamaño al de la imagen original.
2. Se determina la desviación estándar del ruido en  $w_1$ .
3. Se deduce el *threshold* en cada escala desde el modelado de ruido.
4. Se determinan las relaciones interescales.
5. Se determinan los máximos coeficientes de wavelet en el WTS.
6. Se extraen todos los árboles resultantes conectados a los máximos de cada WTS.

#### 4.6.3. Reconstrucción Parcial

**Reconstrucción parcial como un problema inverso.** Un set de estructuras  $S_i$  donde  $S_i = \{S_{j,k}, \dots, S_{j',k'}\}$  definen un objeto  $O_i$  que puede ser reconstruido separadamente desde otros objetos. Se denota  $W_i$  como el set de coeficientes de wavelet pertenecientes al objeto  $O_i$ . Por lo tanto  $W_i$  es un subset de la transformación de wavelet de  $O_i$ ,  $\tilde{W}_i = WO_i$ . De hecho, la última escala de  $\tilde{W}_i$  es desconocida, así como también otros coeficientes de wavelet que no han sido detectados. El problema de reconstrucción consiste en buscar una imagen  $O_i$  tal que su transformación de wavelet reproduce los coeficientes  $W_i$  es decir pertenecen a las mismas estructuras  $S_i$ . Si  $W$  describe el operador de transformación de wavelet y  $P_w$  el operador de proyección en el subespacio de los coeficientes detectados, es decir los coeficientes donde no se detecto nada se toman como valor 0. La solución se busca minimizando la *ecuación 5*.

$$J(O_i) = ||W_i - A(O_i)|| \quad (5)$$

Se debe resolver el problema inverso que consiste en determinar  $O_i$  conociendo  $A$  y  $W_i$ . El tamaño de la imagen restaurada es arbitrario y puede ser mayor que los coeficientes conocidos. Por lo general la solución a la *ecuación 5* posee infinitas soluciones, depende de la naturaleza de la imagen.

**Algoritmos de reconstrucción.** El método de mínimos cuadrados puede ser utilizada para resolver la relación  $W_i = A(O_i)$  que lleva a buscar la imagen  $O_i$  que minimiza la distancia  $||W_i - A(O_i)||$ .  $||W_i - A(O_i)||$  es mínimo si y solo si  $O_i$  es una solución de la *ecuación 6*.

$$\tilde{A}(W_i) = (\tilde{A} \circ A)(O_i) \quad (6)$$

$W^{-1}$  es el operador de reconstrucción de wavelet. De una estructura de wavelete  $W_i$ , una imagen  $O_i$  es restaurada mediante la suma de las escalas de wavelet y la última imagen suavizada.  $W_i$  no es necesariamente la transformación de wavelet de una imagen, por lo que  $W^{-1}WO_i$  puede no representar la imagen.

Un algoritmo de reconstrucción puede descartar las capas que posean información irrelevante y aplicar la transformada inversa de wavelet para reconstruir la imagen con algunas capas.

**Ejemplo: Extracción de estrella de NGC2997.** En el ejemplo se aplica MVM a la galaxia NGC2997. Dos imágenes fueron creadas sumando objetos detectados de la primera y segunda escala, y desde la tercera a la sexta escala. Se muestran en la *figura 4* respectivamente arriba a la derecha y abajo a la izquierda. La imagen de arriba a la izquierda muestra la imagen original y la imagen ubicada a la derecha abajo muestra la diferencia entre la imagen original y la imagen que contiene los objetos de las escalas 1 y 2. Como se puede ver, al remover los pequeños objetos la galaxia puede ser mejor analizada.

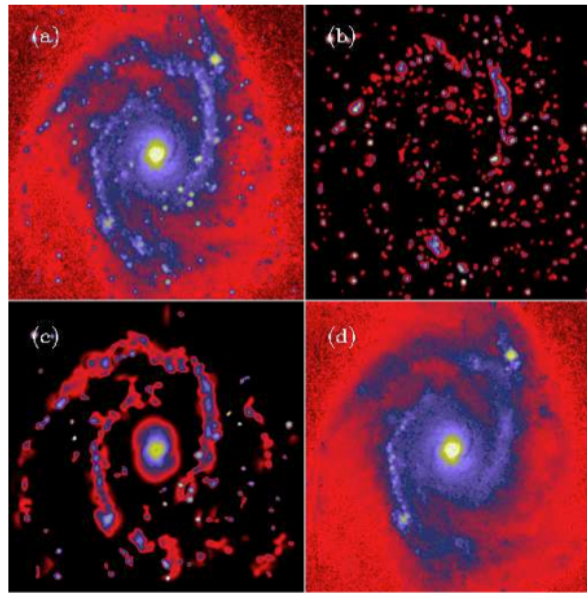


Figura 4: Ejemplo de extracción de estrella.

#### 4.7. Herramientas para procesar imágenes

**MATLAB** viene de *Matrix Laboratory* y es un software que posee un lenguaje de alto nivel en un ambiente interactivo que permite realizar intensas tareas de computación de mejor manera que los lenguajes de programación tradicionales. Su enfoque principal es facilitar el desarrollo de las operaciones matriciales.

MATLAB a su vez también posee una caja de herramientas destinadas al procesamiento de imágenes y tiene una gran variedad de operaciones que pueden ser utilizadas:

- Operaciones geométricas.
- Operaciones por bloque y por vecinos.
- Filtros lineales y filtros especiales.
- Transformaciones (Fourier, Wavelet, entre otras).
- Análisis de imágenes y aumento.
- Operación sobre imágenes binarias.

Además soporta varios formatos de imágenes por nombrar algunos: BMP, GIF, HDF, JPEG, PNG, entre otros [2].

Al igual que MATLAB, *Octave*[3] y *Scilab*[4] son programas utilizados para realizar cálculos numéricos, también poseen herramientas para el procesamiento de imágenes y poseen un lenguaje muy similar al de MATLAB, la gran diferencia que poseen es que son software libre a diferencia de MATLAB.

## 4.8. Observatorios Astronómicos Virtuales

Los Observatorios Astronómicos Virtuales hacen el esfuerzo de tener toda la información astronómica del mundo mediante un acceso fácil, utilizando una interfaz web. Los observatorios permiten que los usuarios puedan obtener información astronómica que se encuentre en algún lugar del mundo con el fin de poder analizarla [7].

Los astrónomos construyen nuevos telescopios cada día, la cantidad de información que estos colectan crece exponencialmente. Los astrónomos intentan estudiar detenidamente cada pieza de la información obtenida por los telescopios, para ayudar en esta tarea la información puede ser analizada por más personas a lo largo del mundo y a la vez permite relacionar la información con otros casos similares.

El desafío que tienen los observatorios virtuales es tremendo, dado que la cantidad de información que disponen crece cada día, otro desafío no menor es la conectividad que debe haber entre los distintos telescopios y los distintos observatorios virtuales, con el fin de que la información sea entregada de la manera más rápida posible y sin poseer errores que puedan entorpecer el análisis.

## 5. Planificación de Trabajo

### 5.1. Alcance del proyecto

Los alcances del proyecto se describirán en 3 áreas distintas que son procesamiento de imágenes, base de datos e índice de objetos.

#### 5.1.1. Procesamiento de imágenes

En el ámbito de procesamiento de imágenes los formatos a procesar serán FITS obtenidas de observatorios ópticos como también de radio. Para empezar solo se utilizarán FITS ópticos, una vez validado que los algoritmos funcionan se procederá a realizar transformaciones a las imágenes de radio para poder procesarlas de igual manera.

Los objetos encontrados en las imágenes FITS en un principio no se catalogarán es decir no se identificará que tipo de objeto es el que esta presente, sólo se identificará su posición como su radio mayor en ambas coordenadas.

El lenguaje a utilizar para el procesamiento será R con el uso de librerías como FITSio para lectura y escritura de FITS, las librerías EBImage y biOps se utilizarán para el procesamiento de imágenes.

#### 5.1.2. Base de datos

En el ámbito de base de datos los datos a archivar en esta por cada objetos serán sus coordenadas RA y DEC como también algún tipo de extracción de características que se logre deducir, como podría ser su excentricidad, su radio mayor, entre otras.

#### 5.1.3. Índice de objetos

El índice a utilizar para acceder a cada objeto será un R-Tree por lo que la base de datos utilizada debe permitir la implementación de este. Las búsquedas de objetos en un principio serán solo mediante coordenadas RA y DEC, obteniendo como resultados todos objetos presentes en una cierta área de interés como también poder acceder al FITS superior de donde se obtuvo el objeto.

## 5.2. Carta Gantt

| Actividades  | Fecha de Entrega | Estado  |
|--|------------------|---------|
| Documento de estado del arte                           | 01/06            | OK      |
| Documento extracción de características                | 28/06            | OK      |
| Hito Número 1 FONDEF                                   | 6/08             | Working |
| Documento obtención de coordenadas locales y absolutas | 20/08            | Pending |
| Documento sobre transformación de imágenes de radio    | 30/08            | Pending |
| Documento de elección y uso de base de datos           | 11/09            | Pending |
| Documento de implementación de índice R-Tree           | 20/09            | Pending |
| Documento sobre búsquedas espaciales                   | 05/10            | Pending |
| Documento de validación del procesamiento              | 15/10            | Pending |
| Documento final del proyecto                           | 22/10            | Pending |

Cuadro 1: Actividades del proyecto.

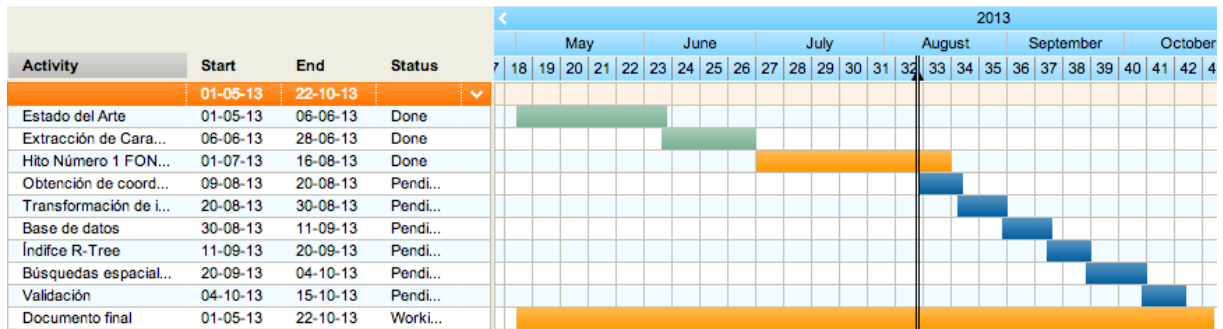


Figura 5: Carta Gantt del proyecto.

## 5.3. Explicación hitos importantes

- Documento de estado del arte: *Documento que posee la investigación sobre el estado del arte.*
- Documento extracción de características: *Documento que describe la extracción de características de los objetos presentes en la imagen FITS.*
- Hito Número 1 FONDEF: *Entrega de este presente informe.*
- Documento obtención de coordenadas locales y absolutas: *Descripción de obtención de coordenadas locales y absolutas de los objetos.*
- Documento sobre transformación de imágenes de radio: *Transformar imágenes obtenidas de un radio observatorio para poder procesarlas de igual manera.*
- Documento de elección y uso de base de datos: *Elección de base de datos a utilizar y su implementación.*

- Documento de implementación de índice R-Tree: *Descripción de implementación del índice R-Tree.*
- Documento sobre búsquedas espaciales: *Realizar búsquedas espaciales sobre el índice.*
- Documento de validación del procesamiento: *Validar lo realizado con imágenes de prueba.*
- Documento final del proyecto: *Documento final con todos los hallazgos y conclusiones.*

## Referencias

- [1] Edge detection. <http://www.cse.unr.edu/~bebis/CS791E/Notes/EdgeDetection.pdf>. Online; Accessed 27-10-2012.
- [2] Matlab. <http://www.mathworks.com/products/matlab/>. Online; Accessed 05-11-2012.
- [3] Octave. <http://www.gnu.org/software/octave/>. Online; Accessed 05-11-2012.
- [4] Scilab. <http://www.scilab.org>. Online; Accessed 05-11-2012.
- [5] SExtractor. <http://www.astromatic.net/software/sextractor>. Online; Accessed 27-10-2012.
- [6] Transformación de wavelets. <http://sabia.tic.udc.es/gc/Contenidosadicionales/trabajos/Imagenyvideo/compresion/3.2.3.2.htm>. Online; Accessed 29-10-2012.
- [7] Virtual observatory. <http://www.virtualobservatory.org/faq.aspx>. Online; Accessed 05-11-2012.
- [8] A. BIJAOU. Sky background estimation and application. *Astronomy and Astrophysics*. Page 81-84.
- [9] Fionn. STARCK, Jean-Luc y MURTAGH. *Handbook of Astronomical Data Analysis*. Springer-Verlag.