Instituto Politécnico Nacional
Centro de Investigación en Computación

Detección de objetos variables en imágenes astronómicas a través del tiempo

TESIS

Que para obtener el grado de
Maestra en Ciencias de la Computación

Presenta:
ISC. Ana Bertha Cruz Martínez

Director:
Dr. Adolfo Guzmán Arenas

México,D.F.
Enero 2016

SIP-14

INSTITUTO POLITÉCNICO NACIONAL

SECRETARÍA DE INVESTIGACIÓN Y POSGRADO

ACTA DE REVISIÓN DE TESIS

En la Ciudad de México, D.F. siendo las 12:00 horas del día 04 del mes de diciembre de 2015 se reunieron los miembros de la Comisión Revisora de la Tesis, designada por el Colegio de Profesores de Estudios de Posgrado e Investigación del:

Centro de Investigación en Computación

para examinar la tesis titulada:

"Detección de objetos variables en imágenes astronómicas a través del tiempo"

Presentada por el alumno:

CRUZ	**MARTÍNEZ**	**ANA BERTHA**
Apellido paterno	Apellido materno	Nombre(s)

Con registro:

B	1	3	0	0	7	3

aspirante de: **MAESTRÍA EN CIENCIAS DE LA COMPUTACIÓN**

Después de intercambiar opiniones los miembros de la Comisión manifestaron ***APROBAR LA TESIS***, en virtud de que satisface los requisitos señalados por las disposiciones reglamentarias vigentes.

LA COMISIÓN REVISORA

Director de Tesis

Dr. Adolfo Guzmán Arenas

Dr. Oleksiy Pogrebnyak

Dr. Juan Humberto Sossa Azuela

Dr. Salvador Godoy Calderón

Dr. Jesús Guillermo Figueroa Nazuno

Dr. Gilberto Lorenzo Martínez Luna

PRESIDENTE DEL COLEGIO DE PROFESORES

Dr. Luis Alfonso Villa Vargas

INSTITUTO POLITÉCNICO NACIONAL

SECRETARÍA DE INVESTIGACIÓN Y POSGRADO

CARTA CESIÓN DE DERECHOS

En la Ciudad de **México D.F.** el día **14** del mes **diciembre** del año **2015**, la que suscribe a **Ana Bertha Cruz Martínez** alumna del Programa de **Maestría en Ciencias de la Computación** con número de registro **B130073**, adscrito a **Centro de Investigación en Computación**, manifiesta que es autor (a) intelectual del presente trabajo de Tesis bajo la dirección de **Dr. Adolfo Guzmán Arenas** y cede los derechos del trabajo intitulado **Detección de objetos variables en imágenes astronómicas a través del tiempo**, al **Instituto Politécnico Nacional** para su difusión, con fines académicos y de investigación.

Los usuarios de la información no deben reproducir el contenido textual, gráficas o datos del trabajo sin el permiso expreso del autor y/o director del trabajo. Este puede ser obtenido escribiendo a la siguiente dirección **isc.anabcm@gmail.com** . Si el permiso se otorga, el usuario deberá dar el agradecimiento correspondiente y citar la fuente del mismo.

ISC. Ana Bertha Cruz Martínez

Tira los dados

Si vas a intentarlo,
ve hasta el final.
De lo contrario
no empieces siquiera.
Tal vez suponga perder novias,
esposas, familia, trabajos
y quizás hasta la cabeza.
Tal vez suponga no comer durante
tres o cuatro días,
tal vez suponga helarte
en el banco de un parque.
Tal vez suponga la cárcel, la humillación,
el desdén y el aislamiento.
Tu aislamiento.
Todo lo demás sólo sirve para poner
a prueba tu resistencia,
tus auténticas ganas de hacerlo.
Y lo harás.
A pesar del rechazo y
de las ínfimas probabilidades,
y será mejor que cualquier cosa
que pudieras imaginar.
Si vas a intentarlo,
ve hasta el final.
No existe una sensación igual.
Estarás sólo con los dioses
y las noches arderán en llamas.
Hazlo, hazlo, hazlo.
Hazlo.
Hasta el final.
Y llevarás las riendas de la vida
hasta la risa perfecta,
es por lo único que vale
la pena luchar.

Charles Bukowski

Dedicado a mi madre
Un ejemplo de lucha.

Agradecimientos

Al Dr. Adolfo Guzmán Arenas por su apoyo, disponibilidad y participación activa en el desarrollo mi trabajo de tesis. Su experiencia y consejos fueron fundamentales para la culminación de este proyecto.

Se agradece al Dr. Giuliano Pignata, investigador del departamento de astronomía de la Universidad Andrés Bello en Santiago de Chile, por recibirme como parte de su grupo de trabajo en el invierno(chileno) del 2015. Sus conocimientos en el área de astronomía y guía impulsaron mi trabajo.

Gracias al Dr. Omar López-Cruz, que por su innovación y emprendimiento incursioné en la astronomía y ha inspirado mis proyectos de investigación.

También agradezco a mi comité tutorial por sus comentarios y observaciones: Dr.Gilberto Martínez Luna, Dr. Salvador Godoy Calderón y Dr. Oleksiy Pogrebnyak.

Se agradece al Consejo Nacional de Ciencia y Tecnología(CONACyT) quien me brindo el apoyo de una beca nacional de posgrado durante dos años y el apoyo para realizar una estancia en extranjero. Al Instituto Politécnico Nacional por su apoyo en la realización de este trabajo de investigación, brindándome la beca BEIFI y beca de tesis de maestría. Para ambos institutos por brindar a los jóvenes mexicanos la oportunidad de contribuir a la ciencia que impulsa el desarrollo de nuestro país a través de estas becas y educación de alto nivel.

Un profundo agradecimiento al Centro de Investigación en Computación y al personal que ahí labora. Por todo el conocimiento y enseñanzas profesionales y personales que me brindaron durante mi estancia.

A mi familia que ha sido pilar de mi vida durante este tiempo. A mi madre ejemplo de fortaleza, a mi padre por su consejo y paciencia. A mi hermana Rocio por su esfuerzo diario. A mis hermanos Sara, Gloria y Nestor.

Agradezco a mis amigos en el mundo, por su invaluable apoyo en estos dos años y medio. Que han estado ahí, quizá no todos en el sentido físico de la palabra, pero han sabido hacerse presentes en los momentos importantes.

A cada mexicano que día a día trabaja por que nuestro país sea un lugar mejor para vivir.

Gracias a la vida que me ha dado tanto
Me dio el corazón que agita su marco
Cuando miro el fruto del cerebro humano,
Cuando miro al bueno tan lejos del malo.
Violeta Parra

Resumen

Uno fenómeno interesante y poco frecuente que se estudia en la astronomía es la explosión de una supernova, la cual tiene una duración corta, del orden de días. El estudio de las supernovas ha cobrado importancia, ya que son objetos que cuentan con una curva de luz conocida, en específico la supernova de tipo Ia. Estas pueden ser usadas como puntos de referencia para medir la aceleración de la expansión del universo. Al medir el corrimiento al rojo de una supernova, se sabe a qué velocidad se aleja de nosotros y por ende la velocidad a la que se expande el universo.

Para llevar a cabo esta tarea se han desarrollado diferentes tipos de herramientas y técnicas de observación automatizadas para la adquisición de datos que han producido grandes volúmenes de información disponibles.

Este trabajo propone una modelo que permite procesar imágenes del visible, implementando algunas técnicas de análisis de señales que son útiles para el tratamiento de series de tiempo, que los astrónomos denominan curvas de luz. También se implementó un módulo de clasificación supervisada en tres clases: estrellas, objetos variables y candidatos a supernovas, y adicionalmente una subclasificación de los candidatos en los diferentes tipos de supernova: Ia, Ibc y II.

El modelo fue construido experimentalmente, probando diferentes métodos para cada etapa del proceso, algunos de estos métodos se manejan en la astronomía de manera separada para procesar datos y hacer detecciones de supernovas y otros fenómenos, los módulos desarrollados son:

- Preprocesamiento: fotometría y etiquetado de imágenes: Se integró el software PPP(Picture Processing Package) del Dr. H. Yee que lleva a cabo estas tareas.
- Se automatizó la creación de la base de datos generando las diferentes sentencias SQL y usando MySQL como gestor de la base de datos, organizándola por regiones y catálogos.
- Construcción de series de tiempo de cada objeto que aparece en los catálogos e integración a la base de datos.
- Tratamiento de las series de tiempo: limpieza del ruido de fondo, interpolación, filtrado para suavizar la curva de luz y estandarización.
- Clasificación supervisada: objetos constantes, objetos variables y candidatos a supernovas. Subclasificación en los tipos de supernova usando ejemplos reales tomados de la literatura.

El software se desarrolló en Python, probando cada módulo con dos regiones de la base de datos del proyecto CHASE del observatorio Cerro Tololo, Chile. Posteriormente, se procesó la base de datos completa y se determinó su distribución de los objetos en tres clases. Éste trabajo presenta el modelo propuesto, el software desarrollado, los resultados finales de la clasificación y el conjunto de candidatos a supernovas, así como su clasificación en sus diferentes tipos.

Abstract

One interesting and unusual phenomenon that study the astronomy is the explosion of a supernovae, it has a short duration, on the order of days. The study of supernovae has become important because they are objects that have a known light curve, specifically the Ia supernovae type. These can be used as benchmarks to measure the acceleration of the universe's expansion. Measuring the redshift of a supernova we could known how fast away the universe expands.

To accomplish this task we have developed different types of tools and use techniques for automated observation data acquisition that have produced large volumes of information.

Astronomers have begun the process of extracting information with computational techniques as an alternative to the systematic analysis of the data. The main techniques used include the image and signal analysis.

This work proposes a model that can process images, implementing some signal analysis techniques that are useful for the treatment of time series, which astronomers call light curves. Supervised classification module was also implemented for three classes: stars, variable objects and supernovae candidates, and additionally a sub-classification of the supernovae candidates in their different types of supernova: I y II.

The model was built experimentally, trying different methods for each stage of the process, some of these methods are handled in astronomy separately to process data and make detection of supernovaes and other phenomena, the developed modules are:

• Preprocessing: photometry and labeling images with PPP software (Picture Processing Package) developed by Phd. H. Yee carrying out this task was integrated.

• The creation of the database was automated generating and using different SQL Statements and MySQL as manager of the database, organizing by regions and catalogs.

• Construction of time series of each object in the database like a tuple.

• Treatment of time series: clean background noise, interpolation supernovas filter to soften the light curve and standardization.

• Supervised classification: constant objects, variable objects and candidates. Subclassification in supernova types using real examples from the literature.

The software was developed with Python, testing each module with two regions of the database project CHASE Observatory from Cerro Tololo, Chile. Subsequently, the entire database is processed and its distribution of the objects was determined in three classes. This work presents the proposed model, the developed software, the final results of classification and the set of supernovae candidates, and their classification in their different types.

Índice general

Índice de figuras

Índice de cuadros

Algoritmos

Capítulo 1

Introducción

1.1. Antecedentes

Los fenómenos astronómicos han sido registrados a lo largo de la historia de la humanidad. Las estrellas fugaces y los cometas representaban eventos extraordinarios en el cielo, el cual generalmente se consideraba inmutable.

La primera supernova de la que se tiene registro se encuentra en la constelación de Tauro, descubierta por los astrónomos chinos en el año 1053 D.C. También Tycho Brahe observó en 1572, en la constelación de Casiopea una supernova[18], en la figura 1.1 se observa el remanente de dicho evento. Otro ejemplo de este tipo de observación fue hecho por Kepler en 1604, donde reporta la aparición de una nueva estrella, que en nuestros días conocemos como nova.

Hoy en día contamos con múltiples avistamientos de cometas, asteroides y estrellas fugaces que son de interés para la ciencia, además de ser una actividad cotidiana en la astronomía.

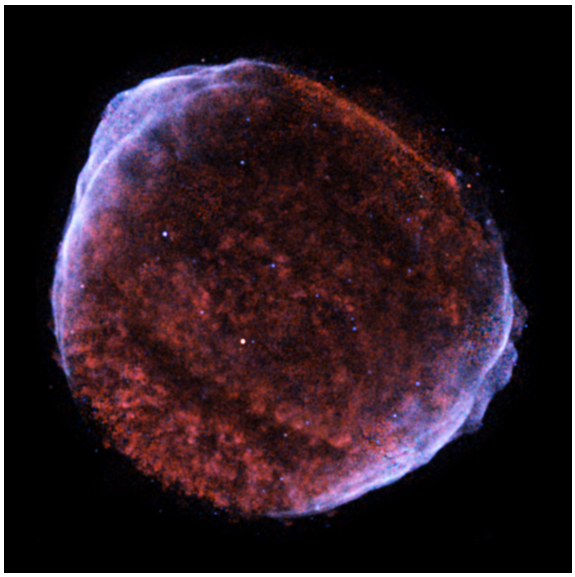


Figura 1.1: Remanente de Supernova de 1572, descubierta por Tycho Brahe

La astronomía como rama de la física ha evolucionado paralelamente al ritmo de desarrollo de la tecnología. Hoy en día es posible no solo detectar objetos cómo las supernovas, sino explorar toda la bóveda celeste en una sola noche. Esto ha traído como consecuencia la producción y almacenamiento de grandes volúmenes de datos del orden del pentabytes[8]. Estos datos no están siendo aprovechados en su totalidad por los expertos del área, ya que no se cuenta con las herramientas computacionales suficientes(hardware y software) para procesar estos datos y producir información de interés. Existen algunas áreas de la computación que han trabajado con esta problemática, entre ellas podemos mencionar el análisis de imágenes, la minería de datos y la inteligencia artificial.

Uno de los primeros trabajos de búsqueda de objetos variables, como fenómenos astronómicos más representativos fue hecho por Edwin Hubble [15], quien buscó estrellas variables periódicas, y al encontrar estrellas cefeidas pudo determinar que la nebulosa de Andromeda es un sistema externo a nuestra galaxia. Fritz Zwicky, sucesor de Hubble, es pionero en el trabajo de la búsqueda de supernovas. Perlmutter[21] uso la tecnología emergente de CCD, que esta disponible desde 1980, y el tratamiento de imágenes para construir el primer detector de supernovas eficiente. Los astrónomos han optado por tomar el trabajo de desarrollo de aplicaciones computacionales en sus manos, estas es una razón por la cual los científicos de la computación debemos incursionar en esta área y trabajar interdisciplinariamente.

1.2. Planteamiento del problema

Los seres humanos contamos con un sistema de visión que nos permite ver la luz en diferentes bandas del espectro electromagnético [24]. En este caso es apenas una pequeña porción de todo lo que abarca el espectro como se observa en la figura 1.2.

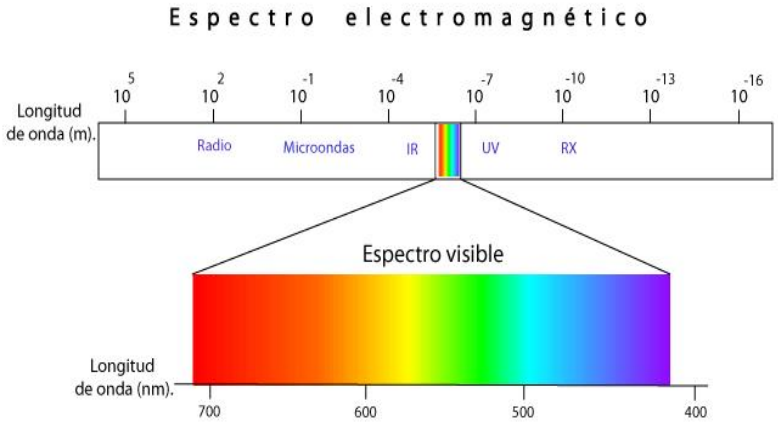


Figura 1.2: Espectro electromagnético

Los sistemas de observación astronómica cuentan con diversos detectores como fotómetros y espectrómetros para detectar fuentes en diferentes bandas que no son visibles para nosotros. Como se nota en la figura 1.2, existen diferentes frecuencias aparte de la visible donde se pueden realizar lecturas, ya sea en el infrarrojo, ultravioleta, rayos x o rayos gamma.

Y es en las frecuencias que se ilustran en la figura 1.2 que los astrónomos han comenzado a explorar el cielo en tiempo real. Es decir, exploran grandes extensiones del cielo y procesan los datos de manera inmediata. Esta nueva tendencia ha llevado a generar grandes volúmenes de datos, almacenados y disponibles. Tenemos por ejemplo los observatorios del proyecto Catalina Real-Time Transient Survey [7] del Caltech o el Sloan Digital Sky Survey (SDSS)[16], que están en constante observación del cielo, produciendo datos del orden de pentabytes. Los datos están disponibles para uso del público en general, por ejemplo los del SDSS que se encuentran en su sitio web, sin embargo no son recientes.

En ocasiones no es posible procesar todos estos datos y producir información útil para el análisis de fenómenos astronómicos, poblaciones estelares y algunas otras aplicaciones. Una de las problemáticas a la que se enfrentan es la detección de fuentes de luz que varían con respecto al tiempo. Para ello, primero, se debe contar con observaciones de una región en particular, donde los objetos observados se almacenan en una imagen como se muestra en la figura 1.3.

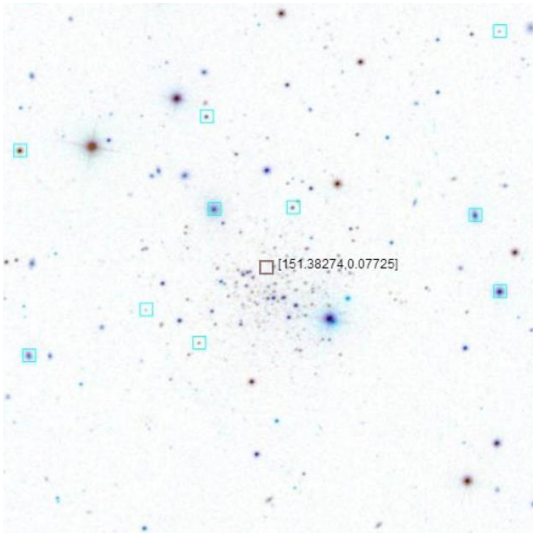


Figura 1.3: Detección de objetos en una imagen[16]

Como se observa, los objetos deben ser localizados y medidos para poder determinar si existe alguna supernovas, asteroides, blazares o evitar defectos en la detección por rayos cósmicos, ya que estos interactúan con el CCD y producen errores. Este proceso requiere la observación de por lo menos dos imágenes de la misma región, ya que en muchos de los proyectos la detección se realiza por medio de la diferencia de imágenes. El proceso de detección de un objeto variable se muestra en la figura 1.4, dónde se pueden ver la fecha de cada imagen y como se le dio seguimiento a un objeto que presenta una variabilidad.

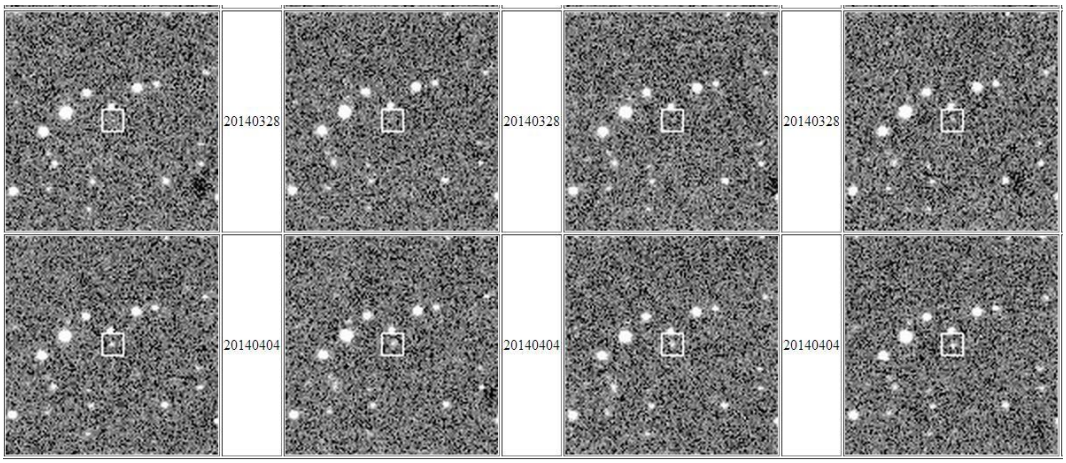


Figura 1.4: Variación de un objeto astronómico a través del tiempo [7]

Otro punto importante que se debe señalar, es que al realizar la detección de objetos con variabilidad se obtienen una gran cantidad de falsos positivos, en cuyo caso es difícil diferenciar entre un objeto de interés como una supernova y una estrella con un periodo largo de variación. Por ejemplo, en 2005 el proyecto SDSS-II, con una capacidad para medir objetos de hasta 0.6" de magnitud y con una medida de señal-ruido mayor a 3, generaba cerca de 4000 objetos candidatos a supernova por noche que requerían una clasificación visual. Otro ejemplo que podemos mencionar es el proyecto ESSCENCE[19], en el que se detectaban entre 200 y 400 candidatos por noche, esta búsqueda fue hecha con el telescopio de 4m del Cerro Tololo Inter-American Observatory (CTIO) en Chile.

1.3. Justificación

La astronomía moderna cuenta con detectores de gran formato que tienen la capacidad de explorar el cielo y cubrir una considerable región de la bóveda celeste en una noche. Este proceso genera grandes volúmenes de datos, que son almacenados en repositorio disponible para el área de astronomía. Además, se están poniendo en marcha nuevos proyectos como el Large Synoptic Survey Telescope(LSST)[17], que generará alrededor de 15 TB por noche, detectando 37 mil millones de estrellas y galaxias, estará en funcionamiento durante 10 años y requerirá un procesamiento rápido en la detección de eventos de interés.

El Panoramic Survey Telescope & Rapid Response System (PAN-STARRS)[6], que cuenta con una de las cámaras digitales más grande del mundo, con 1.4 mil millones de pixeles y que captura aproximadamente 500 imágenes por noche.

Por otra parte, algunos de los fenómenos astronómicos ocurren durante un período corto de tiempo, del orden de días, por ejemplo el paso de un cometa, en cuyo caso cumple con una trayectoria y un periodo de tiempo definido. El surgimiento de una supernova es registrado pocas veces, sobre todo el inicio de dicho evento, por lo regular su descubrimiento ocurre días después. Esto, debido a que no se observa la misma región todas las noches y probablemente tampoco en el momento en en que ocurre el fenómeno o no se han analizado los datos que se capturaron días antes. Es por esta razón que se pretende crear una herramienta computacional

que ayude al astrónomo a procesar imágenes, encontrar los objetos variables y proporcionar un listado de candidatos a supernova.

El descubrimiento y análisis de este tipo objetos han llevado a la conclusión de que el universo esta en expansión y se esta acelerando[13], y cuyo trabajo llevo a Saul Perlmutter, Brian P. Schmidt y Adam G. Riess a ganar el premio nobel de física 2011. Para ello algunos grupos de trabajo se dedicaron a la búsqueda de supernovas en particular de tipo Ia, ya que la forma de su curva de luz es bien conocida y por el efecto doppler y el corrimiento al rojo de los espectro es posible saber a que velocidad se alejan, usando la tierra como punto de referencia.

Las supernovas no son sólo objetos variables sino que están directamente relacionadas con los rayos cósmicos(RGB), también es internaste su estudio ya que indica el origen de los metales en un medio interestelar[12].

La aplicación de clasificación de objetos incursiona en muchas áreas de la astronomía, como la detección de asteroides, rayos gama, supernovas o algunos núcleos de galaxias que muestran variaciones, que en algunos casos requieren de una alerta rápida para dar paso al estudio de estos eventos. De la necesidad de procesar datos en tiempo real, clasificarlos y estudiarlos surgen los trabajos automatizados de dichos procesos. Los métodos de inteligencia artificial y machine learning se están integrando a los sistemas de detección astronómica, para analizar datos automáticamente, sin embargo, aun hay un largo trecho que recorrer y sobre todo explotar las técnicas computacionales para hacer eficientes estos procesos.

Debido a estas razones se considera importante la incursión de la computación a la astronomía para que ayude a procesar los grandes conjuntos de datos. Además que éste trabajo colaborativo ayude a realizar nuevos descubrimientos y mejorar el estudio del universo.

1.4. Hipótesis

Es posible aplicar métodos y técnicas computacionales a datos de astronomía para detectar objetos que varían de magnitud a través del tiempo, determinar un conjunto de objetos candidatos a supernova y realizar una clasificación en sus diferentes tipos.

1.5. Objetivos

1.5.1. Objetivo general

Desarrollar un software que aplique técnicas computacionales de manejo de base de datos, análisis de señales y clasificación para la detección de objetos variables a través de una serie de imágenes astronómicas y mejore la tasa de recuperación de los mismos.

1.5.2. Objetivos específicos

- Proponer el diseño de un sistema computacional y una base de datos que faciliten el almacenamiento de catálogos y procesamiento de las curvas de luz.

- Desarrollar un software que construya las series de tiempo de cada objeto localizado en los catálogos y almacenarlo en una base de datos.
- Proponer e implementar un conjunto de técnicas de estandarización, limpieza y suavizado que son útiles para el tratamiento de señales.
- Implementar y probar un clasificador en tres clases: objeto variable, objeto constante(estrella) y supernova.
- Implementar y probar un clasificador en tres clases de supernovas: Ia, Ibc y II.
- Probar el sistema con un conjunto de datos reducido y a gran escala con una base de datos de observación real.
- El sistema brindará como resultado la clasificación de la base de datos en tres tipos, localizando los candidatos a supernova.
- Se clasificarán los candidatos a supernova en sus tres tipos correspondientes.
- Proporcionar un reporte general de clasificación y las curvas de luz de los candidatos.

1.6. Alcances y límites

Los alcances del presente trabajo son los siguientes:

1. Diseñar, implementar y probar del modelo de detección de candidatos a supernova
2. Procesar imágenes y catálogos de astronomía en el visible con ayuda de PPP
3. Generar scripts para realizar el etiquetado y fotometría con PPP
4. Construir el conjunto de series de tiempo de cada objeto existente en la base de datos para su análisis
5. Filtrar el ruido de fondo para las curvas de luz con un algoritmo de estrellas de referencia
6. Clasificar las curvas de luz en tres clases: objetos variables, constantes y candidatos a supernova
7. Clasificar el conjunto de candidatos a supernova en tres clases: Ia, Ibc, II
8. Interfaz gráfica del software que integre los componentes propuestos
9. Proporcionar un reporte de clasificación para cada región
10. Proporcionar un reporte de clasificación del conjunto de candidatos a supernovas

Los límites del trabajo son:

1. Exploración limitada de los algoritmos de clasificación de series de tiempo, este trabajo es un acercamiento a la detección de supernovas.

2. No se exploran técnicas que pudieran llegar a ser útiles para este problema como el análisis de fourier.

3. En este trabajo no se realiza el proceso de registro de imágenes.

4. No se realiza la clasificación de los objetos variables en todas sus subclases existentes.

1.7. Especificidad del problema

El desarrollo de este trabajo se realizó tomando como base algunos métodos tratamiento de señales y clasificación, se trabajó en la búsquedas de candidatos a supernovas en la base de datos del proyecto CHASE (CHilean Automatic Supernovas sEarch), que tiene un tiempo de observación de dos años(2013-2014) aproximadamente, el lugar de observación fue Cerro del Tololo, Chile y cuya frecuencia de observación fue cada cuatro noches. Se generó una lista de posibles candidatos a supernova, y se agregaron 250 galaxias con una velocidad radial$<$8000km s^{-1}, estas galaxias tienden a generar nuevos objetos variables, las regiones donde se encuentran fueron observadas diariamente. Se usaron cuatro de los seis telescopios del PROMPT(Panchromatic Robotic Optical Monitoring and Polarimetry Telescopes), que tienen un diámetro de 40cm y son completamente robotizados. El tiempo de exposición fue de cerca de 10 segundos por imagen, con una magnitud límite de 18.0 y un tiempo de exposición de 40 segundos.

.

1.8. Organización del documento

Este trabajo esta organizado en siete capítulos, el resumen de cada uno de ellos se muestra a continuación:

En el capítulo 1, se presentan los primeros antecedentes de nuestro trabajo, el planteamiento del problema y los objetivos de desarrollo de software.

El capítulo 2, expone los trabajos previos más representativos que se han realizado para la búsqueda de supernovas y algunas de sus características más relevantes.

En el capítulo 3, se describen los conceptos más importantes respecto al área de astronomía para detectar objetos como supernovas, así como algunas técnicas computacionales desde el enfoque de bases de datos, tratamiento de señales y clasificación supervisada.

En el capítulo 4, se muestra el diseño general del modelo y se describen los procesos involucrados para obtener una serie de tiempo, el tratamiento de la serie como una señal y el proceso de clasificación para obtener un listado de candidatos a supernova.

El capítulo 5, expone la implementación del modelo, la forma en que trabajan los diferentes componentes para procesar los datos de catálogos y como se obtiene la lista de candidatos.

En el capítulo 6, se muestran las pruebas y validación realizadas al sistema con un conjunto de regiones reducida, así como el procesamiento de la base de datos completa. La selección de los parámetros de clasificación y las diferentes pruebas realizadas al sistema. También se muestran los resultados finales de la clasificación en tres clases de supernova.

Por último, el capítulo 7, presenta las conclusiones del presente trabajo, la reflexión de los resultados, las posibles aplicaciones y el trabajo a futuro.

Capítulo 2

Estado del arte

La tarea del cómputo en la astronomía va desde la captura correcta de los datos, pasando de lo analógico al almacenamiento digital, procesando los datos hasta llegar al análisis de la información. Es la tarea de la computación de nuestros días mejorar la calidad interpretativa para el ser humano del mundo que lo rodea, esto no deja fuera áreas de la ciencia como la astronomía.

Los procesos astronómicos conllevan, como todo proceso, la creación de nuevos conocimientos y la asignación de nuevos significados. Estos procesos trabajan particularmente con imágenes, que son generadas a través de detectores CCD, integrados a los telescopios, en donde, de manera automática se hace una lectura del cielo y se digitaliza produciendo imágenes de tipo FITS. En este caso la escala y resolución juegan un papel importante, ya que mucho del trabajo requiere una alta calidad de los datos. También se requieren algunos procesos de análisis de imágenes y tratamiento de señales, de los cuales se pueden mencionar los siguiente [28]:

- Visualización de señales
- Filtrado, que tiene la finalidad de eliminar el ruido
- Deconvolución
- Compresión
- Morfología matemática
- Detección de bordes
- Segmentación y reconocimiento de patrones
- Reconocimiento de patrones multidimensionales
- Transformaciones

2.1. Antecedentes

El trabajo del Catalina Real-Time Transient Survey(CRTS)[8], muestra el interés de los astrónomos por las aplicaciones de técnicas computacionales, dado a la problemática a la que se enfrentan para generar información. El proyecto está orientado a detectar los fenómenos astronómicos, identificar y caracterizar en tiempo real. Para la detección y clasificación de objetos se requieren datos limpios y completos, que también es un área que se ha estado mejorando a lo largo del tiempo.

En la siguiente sección se presentan algunos de los trabajos más representativos realizados en el área de astronomía, que están funcionado en observatorios virtuales o que han sido implementados para procesar grandes repositorios de datos y que intentan solucionar el problema de detección de objetos variables y clasificación.

2.2. Trabajos previos

El trabajo realizado en el proyecto **WOOTS**(Wise Observatory Optical Transient Search)[11], está dirigido a detectar objetos variables con seguimiento espectroscópico. Se usaron datos de cúmulos Abell, con un redshit, z= 0.06-0.2, el redshit es un indicador de la distancia a la que se encuentra el cúmulo de galaxias, con este trabajo se encontraron 12 supernovas. El conjunto de datos que se procesó consistió de 161 clusters de galaxias con características similares.

El proceso de detección inició con el registro de imágenes, usando un marco de referencia, definido por una imagen para cada campo y el algoritmo de triangulación propuesto por Giveon [14], se requirieron correcciones en la rotación de las imágenes, para esto se usó el software astronómico IRAF(Image Reduction and Analysis Facility).

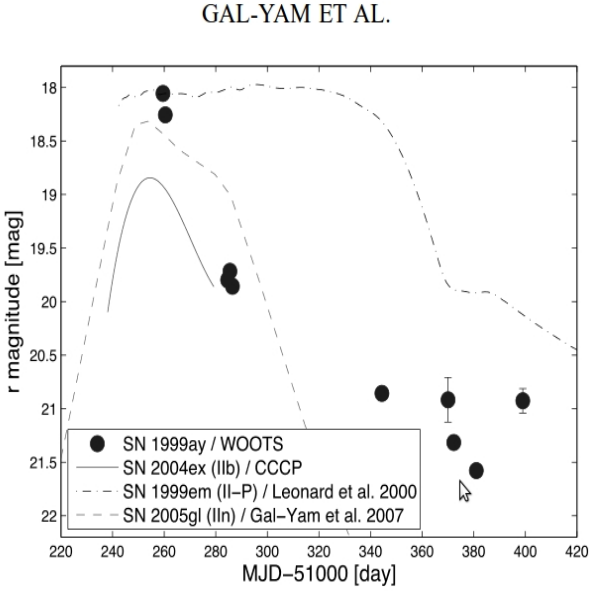


Figura 2.1: Curva de luz de una supernova descubierta por el proyecto WOOTS [11]

Se usó la resta de imágenes con el algoritmo ISIS de C. Alard[2] para la detección de los objetos variables, operando cada imagen con una imagen seleccionada como marco de referencia, obteniendo el residuo con una magnitud a lo más de 21.5. Después de realizar esta operación se generó una lista de objetos candidatos, donde se fue incrementando su prioridad conforme se analizaba cada imagen. De estas detecciones se construyeron las curvas de luz para verificar si correspondía a una SN(Supernova). Se usaron varios criterios para descartar objetos, principalmente su cercanía a las AGN(Active Galactic Nuclei). Otra aporte de este trabajo fue el descubrimiento de más de 50 asteroides.

Fang Yuan[33], desarrolló un trabajo de búsqueda de objetos variables con técnicas de coss-convolución, este método no requiere de imágenes de alta resolución. Usa dos procesos de convolución sobre la imagen de referencia y prueba, haciéndolas coincidir, ya que uno de los principales problemas en la detección de objetos es el registro de imágenes. Su trabajo consistió en reducir un par de imágenes para generar una imagen como marco de referencia, la resta de imágenes tiene algunos problemas de uso de PSF(point spread function) elípticas, que no toman en cuenta la rotación del telescopio, además no siempre se tiene una imagen con la suficiente luminosidad para tomarla como referencia.

El método matemático que uso Fang Yuag consistió en encontrar una PSF con un kernel k que sería operada con una imagen de referencia R y generaría una imagen R*. Entonces, la nueva imagen se compara con R* se llama T y es operada con R*, realizando una minimización pixel a pixel. La segunda convolución agrega grados de libertad para realizar la rotación. Para validar el proceso se usó una validación cruzada para determinar los parámetros fuertes, con el cálculo de la desviación estándar para cada punto. Con esto se resolvió el problema del emparejamiento de imágenes. Después sólo se realiza la resta de imágenes, para ello se utilizó la rutina POLYWARP de IDL, para detectar objetos superpuestos con respecto a una imagen de referencia. El trabajo se implementó para reducir 400 imágenes en un día, el principal problema al que se enfrentaron fue implementar la convolución eficientemente ya que con una matriz 9x9 kernels los parámetros de libertad crecían a 160. Entonces, se utilizó una función bicubic de ranura para reducir el número de operaciones. En la figura 2.2, se puede observar la distribución de los k-kernels sobre una imagen y los resultados en la imagen de la derecha del proceso de convolución y resta de imágenes.

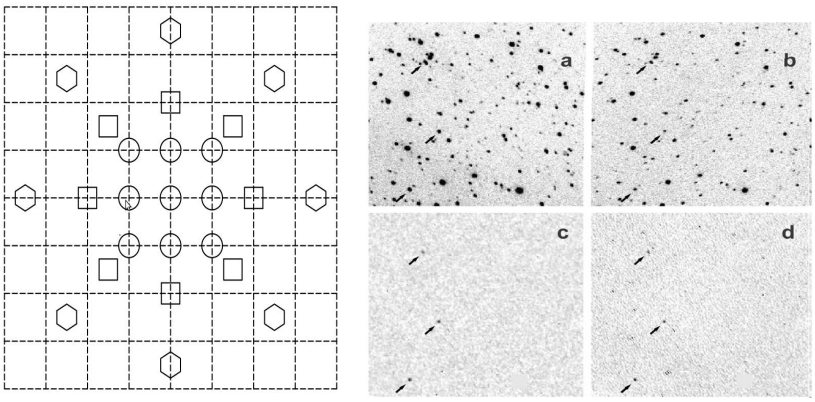


Figura 2.2: Diagrama que muestra la ubicación de los bicubic, para la construcción de kernels. Esta es una matriz de 9x9 da como resultado 160 parámetros. La segunda imagen muestra el resultado de la convolución (d), donde (a) es la entrada.

Se pudieron procesar imágenes de 2045x2045 en 4min aproximadamente con 2.0 GHz y 25 kernels. Como resultado se han detectado 7 novas en los campos de M31 y M33, este método trabaja con objetos de hasta 18.5 de magnitud.

MENcaCS[25], es un proyecto que tiene como objetivo buscar supernovas en clusters de galaxias, ha descubierto 23 supernovas de tipo Ia en el Abell Cluster 399 y Abell 85. Algunas de las características que se tomaron en cuenta están relacionadas con el ICL (intracluster light). Como primer paso aplicó un método de separación de objetos por elipticidad normalizada [29]. y la caracterización de cada objeto se realizó con SExtractor, donde se podían detectar objetos al borde de las galaxias del orden de R$\sim$3. Si R$<$5 del objeto, entonces no se tomada como candidato a supernova y si R$>$5, entonces se tomaba como parte del halo de la galaxia cercana. Para la selección de los candidatos se realizó el seguimiento de los objetos, observándolos por 30 días aproximadamente.

Para la observación se implementó un sistema de análisis en tiempo real para calibrar el flujo de luz de fuentes referentes a las supernovas. Después de la detección, se procedió a procesar los datos que se habían tomado por más de un año, procesando primero imágenes en donde se supiera que estaban libres de supernovas y donde se agregaron estrellas artificiales para mejorar la detección de objetos de baja magnitud. De esta manera, al agregar los objetos artificiales y aplicar algunas funciones de manipulación de imágenes y ampliar la magnitud en 1.5 con una PSF, se generaron nuevamente el catálogo. Usando SExtractor se pudieron encontrar los candidatos a supernova. En algunos casos se usó GALFIT, que es un modelador de galaxias, y con el componente de este modelador llamado Sércic se realizó la resta de la galaxia a la imagen con el fin de dejar más limpia el área donde se querían localizar las supernovas como se observa en la figura 2.3.

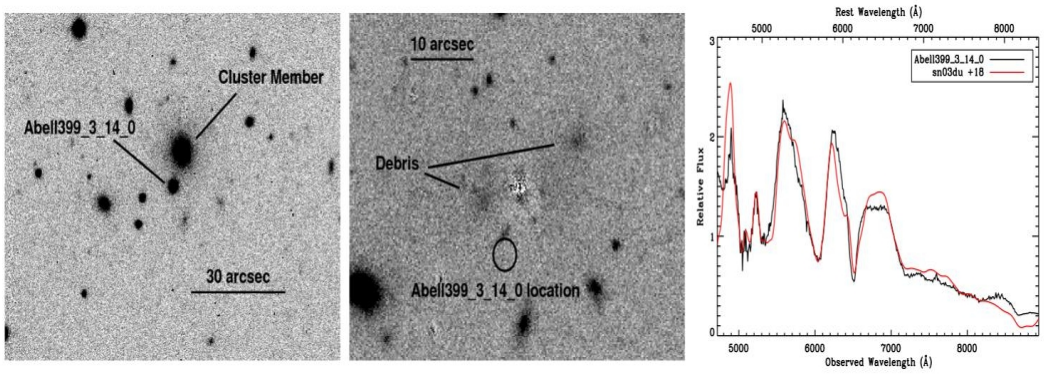


Figura 2.3: A la izquierda el objeto Abell-399_14_0 descubierto en un cluster, en medio se ve la imagen ampliada y sin la galaxia local, al que se le restó la componente Sércic de GALFIT. En la gráfica se puede observar el flujo de luz de los objetos.

En el caso de **CRTS de Caltech**[8], el equipo de trabajo desarrolló un conjunto de técnicas y metodologías para el análisis automático en tiempo real, y el manejo de grandes volúmenes de información, integrándolos al observatorio virtual. El sistema incorpora machine learning para la clasificación de eventos astronómicos fugaces, basados en las detecciones de sus telescopios automatizados. El proceso de clasificación es el más delicado de todos y se requirió de métodos robustos, para ello usaron máquinas de soporte vectorial. Una de las técnicas más adecuadas para resolver este problema es la probabilística, ya que ningún objeto se queda sin clasificar y se actualiza la clasificación conforme se reciben nuevos datos, esta técnica se implementó para descartar rápidamente objetos.

El problema de clasificación de objetos inicia con diferenciar entre un suceso astronómico y algún error de calibración del sistema de captura. Algunos de los métodos más usados para esta clasificación son las redes neuronales artificiales (ANN) y un método alternativo son las máquina de soporte vectorial (SVM), que trabajan sobre los parámetros morfológicos, lo cual elimina cerca del 95 % del ruido de entrada. La detección de los objetos variables sería posible usando estas técnicas, pero se carecen de los datos completos por errores de lectura. Para resolver el problema de clasificación con datos faltantes se experimentó con un modelo bayesiano, ya que con este método se pueden usar diferentes subconjuntos independientes de parámetros para clasificar. Cada parámetro necesita ser estimado para cada tipo de fenómeno en que se desea clasificar. Entonces, a cada evento se le puede estimar su clasificación de manera probabilística. Este método tiene la ventaja de aprender con cada evento que es clasificado ya que tiene un aprendizaje incremental. Se usó una BN (Bayes Network), con varias clases, parámetros y capas para este propósito.

El trabajo se centró en la curva de luz, en cuyo caso se representó como un histograma. Observado la distribución de los objetos acorde a su frecuencia y distribución de color. Cuando se intentaba clasificar un objeto se amulaba acorde a su curva de luz en el lugar del histograma que le correspondía. Como se van acumulando los datos se da la clasificación de los objetos, este proceso se muestra en la figura 2.4.

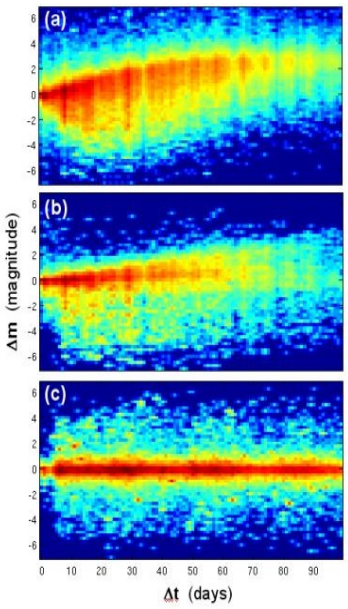


Figura 2.4: Función de distribución probabilística para la clasificación de una SN-Ia

Se seleccionó un conjunto de entrenamiento, el cual fue elegido acorde a los resultados de clasificación. Para verificar la clasificación el conjunto se dividió en 10 subconjuntos de manera aleatoria, donde cada subconjunto se tomó como el conjunto de entrenamiento y los demás como conjunto de control. Con esto se hizo la validación cruzada de la clasificación. Con esta metodología se lograron los resultados expuestos en la tabla del cuadro 2.1.

Clase	Completud	Contaminación
Blazar	83 %	13 %
CV	94 %	6 %
RR Lyrae	97 %	4 %

Cuadro 2.1: Resultados de pruebas de clasificación para CRTS

El proyecto **SuperMacho**[27], tuvo como meta detectar objetos masivos MACHO(**Massive Compact Halo Objects**), que están relacionados con la materia oscura, este proyecto fue pionero en el manejo de grandes bases de datos. Los datos fueron tomados del NOAO(National Optical Astronomy Observatory). Donde se hacen lecturas del orden de GB por imagen, para tener conjuntos de datos de hasta PBytes. Como se puede ver, algunos de los problemas a los que se enfrentan en los observatorios es la detección, procesamiento fotométrico y la clasificación de objetos. En este caso se intentó detectar supernovas, que se cree están relacionadas con la materia oscura. Para ello se crearon diversas rutinas en con el software astronómico IRAF y el lenguaje C. El proceso de detección de este proyecto se ilustra en la figura 2.5.

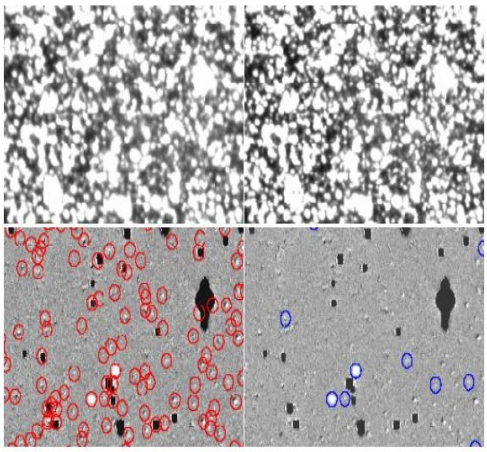


Figura 2.5: En la parte superior se pueden observar dos imágenes de diferentes periodos, entre las cuales se hará la detección, en las imágenes inferiores se puede observar la resta, y en azul se ven los objetos candidatos.

Para la detección de objetos variables uso del software DoPHOT, primero se hacían correcciones a la imagen y después se generaba un listado de objetos de interés que serian monitoreados. El siguiente paso fue la aplicación del método llamado análisis de imágenes por diferencia (DIA), de Phillips & Davis[22], donde se hizo el registro de imágenes con una PSF, y el emparejamiento del flujo de luz conforme se fue haciendo la detección de los objetos variables. Tomaney & Crotts[30], ampliaron el método calculando el núcleo de convolución en un espacio Fourier para hacer coincidir las PSF. El principal problema de éste método es que al hacer el registro de la imagen se puede generar mucho ruido en el resultado y dar lugar a un falso positivo. La solución fue elegir una buena PSF, el resultado fue la detección de objetos variables al procesar cada imagen, esta detección incluye la información producida por el análisis y la posición dentro de la imagen.

El trabajo de **Ciro Donaley**[9] muestra un enfoque de machine learning para la reducción de parámetros. El conjunto de datos está dado por catálogos de objetos que pueden ser representados por curvas de luz, estos datos provienen del CRTS, sin embargo no están completos. Este observatorio virtual ha detectado cerca de 1,800 supernovas y tiene la capacidad de procesar datos hasta una magnitud de 23. La selección de características es una de las tareas más importantes para el proceso de clasificación, ya que esta es la entrada que recibirá el clasificador para hacer una correcta separación. Los algoritmos de cubrimiento son los más usados para evaluar los subconjuntos de características, pero se vuelven filtros muy lentos para la clasificación.

El trabajo consistió en probar cinco algoritmos de cubrimiento y después usar sus salidas para el proceso de clasificación que se evaluaron en dos clasificadores diferentes, el KNN(K-nearest neighbor) y el DT(Desition Tree)). Los resultados de estas pruebas se muestran en la tabla de la figura 2.6.

SN vs "ALL THE OTHERS" (see Table IV for the complete parameters description)		
Feature Selection Strategy	***KNN Loss***	***DT Loss***
None (all parameters selected)	30%	18%
ReliefF (6 parameters selected: $x1, x2, x19, x17, x15, x7$)	22%	15%
CFS (3 parameters selected: $x2$, $x8, x13$)	24%	17%
FCBF (3 parameters selected: $x2, x8, x13$)	24%	17%
MCFS (4 parameters selected: $x9, x13, x14, x16$)	32%	19%
FDR (6 parameters selected: $x15, x5, x8, x14, x16, x17$)	22%	16%

Figura 2.6: Tabla de resultados para cada estrategia de selección de parámetros

En la tabla anterior se muestran los resultados de cada algoritmos de selección de parámetros, usando su salida para los algoritmos de clasificación. Se observan el nivel de eficacia de los conjuntos de parámetros seleccionados respecto a estos algoritmos de clasificación. Un ejemplo de separación de clases se observa en la figura 2.7, donde se tomaron dos parámetros, amplitud y periodo. Se puede ver que con una correcta selección de parámetros se obtiene una buena clasificación de objetos, de rojo se encuentran señaladas las estrellas W Ursae Majoris variable(W UMA), que son del tipo elípticas binarias que tienen contacto y transferencia de masa y energía entre ellas. De azul se ilustran las estrellas del tipo RR Lyrae, estrellas variables pulsantes, cambiantes en su radio. La figura 2.7 muestra claramente la separación de clases.

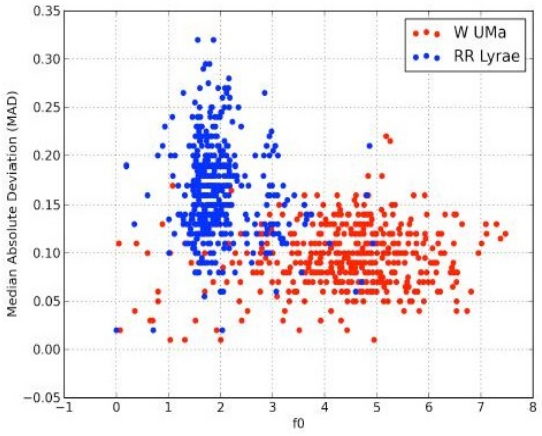


Figura 2.7: Es interesante observar que el periodo y la amplitud dan buenos resultados para separar las clases seleccionadas para el experimento, W UMa y RR Lyrae del CRTS.

En la literatura también podemos encontrar trabajos realizados con el telescopio espacial Hubble, mediante el proyecto **Cluster Supernovae Survey II**[5]. Para realizar la detección

de candidatos a supernovas adoptaron una estrategia de trabajo con imágenes. El primer paso fue seleccionar un conjunto de imágenes de la región a analizar, esta puede ser entre 50 y 80 días anteriores a la búsqueda, enmascarando los pixeles que no contribuyen a la detección. Se hace el registro de imágenes y se sustraen de una imagen de referencia previamente seleccionada. También se resta el ruido de un mapa de ruido del cielo previamente construido. Los candidatos se marcan en este conjunto de imágenes con una bandera. Después de esto se elimina los candidatos que fueron seleccionados de núcleos de galaxias(AGN), en este caso de cinco a diez candidatos pasan los requerimientos. Esto dio como resultado cerca de 1000 candidatos etiquetados para las 155 búsquedas que se realizaron. Cada candidato es analizado por un experto y de esta manera se eliminaron la mayor parte. Quedando unicamente 86 candidatos, de los que se eliminaron aquellos que tuvieran una duración mayor a 60 días, ya que las supernovas duran poco tiempo. Al realizar este proceso quedaron 60 objetos que se clasificaron en dos clases: AGN y supernovas. Al finalizar el proceso quedaron 29 candidatos bien identificados como posibles supernovas. En la siguiente imagen se muestra algunos de los candidatos a supernova de este trabajo.

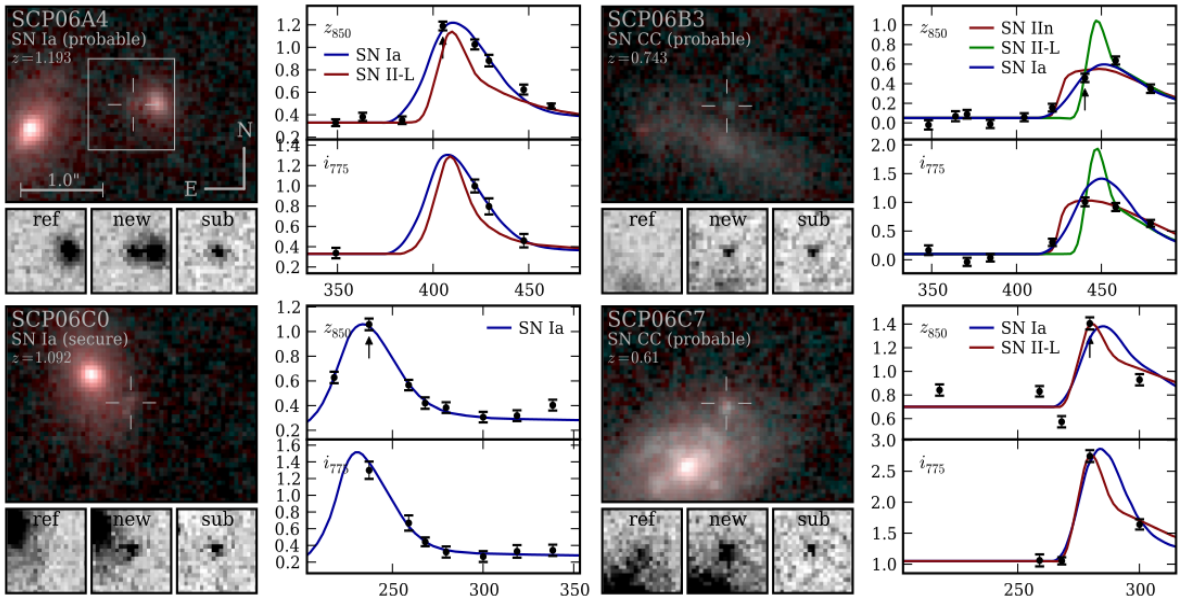


Figura 2.8: Se muestran los algunos ejemplos de candidatos a supernova y sus curvas de luz del proyecto Hubble. En la parte inferior se observa las imágenes de referencia, imagen a procesar y el residuo.

2.3. Tabla comparativa

El cuadro 2.2 resume los trabajos expuestos en este capítulo, sin embargo no son todos los proyectos de búsqueda de supernovas existentes en la astronomía.

Trabajo	**Métodos y técnicas**	**Análisis de cueva de luz**	**Gestor de Base de datos**	**Datos reales**	**Machine learning**	**Software externo**	**Resultados**
Modelo propuesto	**Análisis de señales Quita ruido de fondo con estrellas de referencia. Clasificación supervisada para detección de candidatos.**	**Si**	**Si**	**Si**	**Si**	**PPP**	**20,000 aprox candidatos que son el 10 % del total de la base de datos.**
WOOTS	Registro de imágenes con algoritmo de triangulación.	No	No	Si	Si	IRAF	12 supernovas y más de 50 asteroides.
Astronomical Image Subtraction by Cross-Convolution	Coss-convolution para mejor de resolución. Convolución sobre la imagen de referencia y prueba,haciéndolas coincidir.	No	No	Si	No	POLYWARP IDL	Se han detectado 7 novas en los campos de M31 y M33, este método trabaja con hasta 18.5 de magnitud.
MENcaCS	Se agregaron estrellas artificiales para mejorar la detección de objetos de baja magnitud.	No	No	No	No	SExtractor GALFIT	Ha descubierto 23 supernovas en el Abell Cluster 399
CRTS	Implementan una BN (Bayes Network), con varias clases, parámetros y capas.	Si	Si	Si	Si		Clasificación correcta de objetos variables de hasta un 93 %
SuperMacho	Se usó el método llamado análisis de imágenes por diferencia (DIA), de Phillips & Davis	No	Si	Si	No	IRAF DoPHOT	Detección de objetos variables
Ciro Donale	Algoritmos de cubrimiento.	Si	Si	Si	Si		Más de 1,800 candidatos a supernova
Cluster Supernovae Survey II	Se resta el ruido de un mapa de ruido del cielo previamente construido.	Si	No	Si	No		29 candidatos a supernova.

Cuadro 2.2: Resumen de los proyectos de estado del arte y características reelevantes.

Capítulo 3

Marco teórico

El manejo de datos de origen astronómico conlleva el conocimiento de algunos conceptos que estan íntimamente relacionados con el objeto de estudio de esta área: el universo y su naturaleza. En este caso, nos enfocaremos en los objetos variables, que tienden a cambiar de brillo en periodos regulares. También es importante señalar que estos objetos cuentan con una clasificación, de los que destacan las supernovas, estrellas que cambian de magnitud aparente en un periodo muy corto de tiempo, treinta días en promedio. El concepto de supernova y su clasificación también se revisan en éste capítulo.

3.1. Conceptos de astronomía

El seguimiento de la variación de las fuentes de luz se puede hacer a través de técnicas fotométricas como **curvas de luz**, estas representan la variación de la luz con respecto al tiempo. Un ejemplo de esta herramienta fotométrica se muestra en la figura 3.1, donde se observa cómo varia la magnitud del objeto en cada fase.

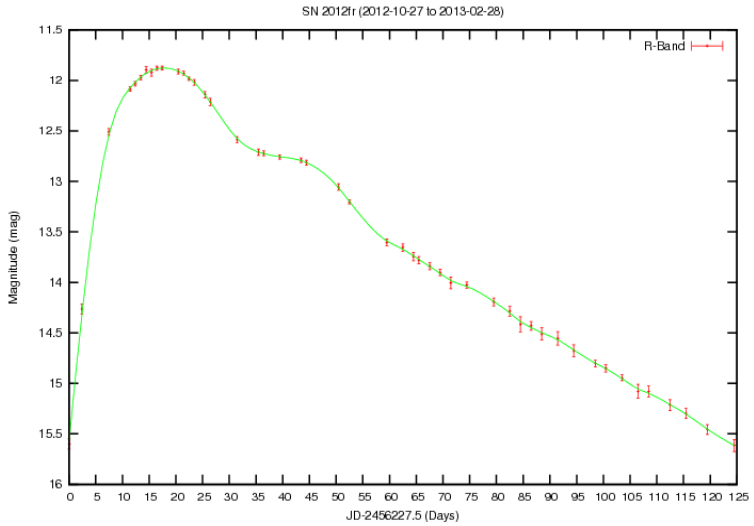


Figura 3.1: Gráfica de una curva de luz de una supernova[7]

La gráfica anterior muestra la variación de la magnitud de una supernova en un periodo de tiempo, es posible construir esta curva a través del análisis de imágenes.

3.1.1. Objetos y estructuras en la astronomía

El universo contiene billones de objetos astronómicos en constante evolución. Con la finalidad de entender el cosmos, los astrónomos se han dedicado a estudiar estos objetos a través de la obtención de miles de imágenes. En estas imágenes se puede analizar la composición de los objetos y entender el comportamiento del universo que nos rodea y del que somos parte. Los estudios del universo indican que algunas de las estructuras que lo componen son [**?**]:

- Sistemas planetarios formados por planetas, satélites y asteroides.
- Estrellas y sus sistemas plenarios que forman parte de una galaxia.
- Cuásares, novas, supernovas, agujeros negros y cometas.
- Galaxias y cúmulos de galaxias
- Gas intergaláctico
- Materia y energía oscura.

Las principales estructuras se muestran en la figura 3.2, también se ilustra la escala de los diferentes objetos:

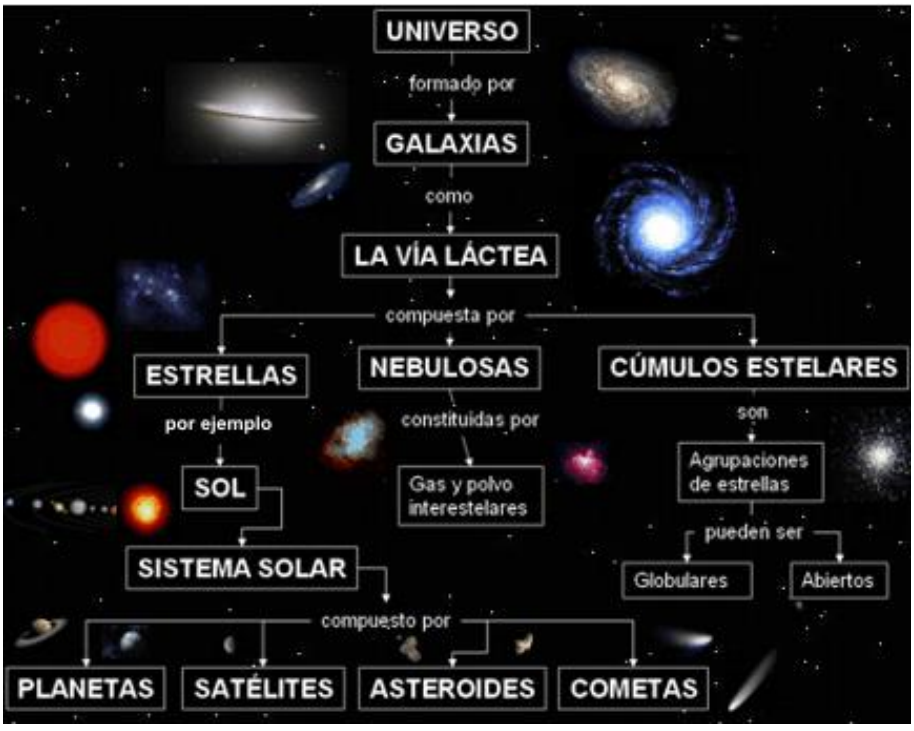


Figura 3.2: Principales objetos y estructuras en el universo.

3.1.2. Objetos astronómicos variables

La observación de los objetos variables es una de las actividades astronómicas con mayor valor científico, ya que describe el comportamiento de las estrellas y en algunos casos la periodicidad de la misma. Otro punto importante es que es difícil observar fenómenos astronómicos trascendentes, ya que estos tardan en ocurrir largos periodos de tiempo. Esta actividad se lleva a cabo a través de la estimación del brillo de la estrella[**?**] y de sus componentes espectrales.

Una estrella variable es aquella en que el comportamiento de su brillo respecto al tiempo muestra variaciones. Estas estrellas tienen una periodicidad regular en la mayoría de los casos, estos periodos se pueden observar en la gráfica de la figura 3.3 que es llamada curva de luz.

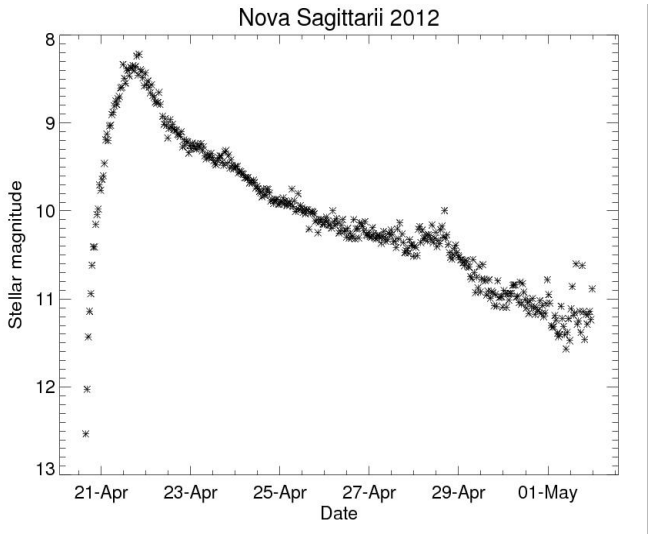


Figura 3.3: Curva de luz de la nova de Sagitario. Se gráfica la magnitud del objeto respecto al tiempo.

En el caso de las estrellas existen dos tipos de variaciones:

- **Intrínseca:** tiene que ver con los cambios físicos en la estrella o sistema estelar.
- **Extrínseca:** la variabilidad depende de otra estrella o de su efecto de rotación.

Algunos de los objetos variables pueden ser estrellas, cometas, asteroides u otros eventos astronómicos. Estas estrellas variables se dividen con frecuencia en cinco clases: variables pulsantes intrínsecas, cataclísmicas, eruptivas, estrellas binarias y giratorias.

- **Variables pulsantes:** son estrellas que muestran expansión y contracción periódica de sus capas superficiales. Las pulsaciones pueden ser radiales o no radial . Una estrella radialmente pulsante sigue siendo de forma esférica , mientras que una estrella no radiales experimenta pulsaciones que pueden desviarse de una esfera periódicamente. Dentro de sus subclases podemos encontrar las Cepheids,estrellas RR Lyrae, estrellas RV Tauri, estrellas de periodo de variación prolongada, estrellas de variación irregular.

- **Cataclísmicas:** como su nombre indica, son estrellas que se comportan violentamente causando procesos termonucleares ya sea en sus capas superficiales o profundas. La mayoría de estas estrellas se encuentran en sistemas binarios cercanos, sus componentes tienen una fuerte influencia mutua en su evolución. Algunas de estas estrellas son las supernovas, novas, novas recurrentes, novas enanas,estrellas simbióticas e hypernovas.
- **Binarias eclipsantes:** son sistemas binarios de estrellas con un plano orbital situada cerca de la línea de visión del observador. El eclipse periódico causa una disminución en el brillo aparente del sistema para observador. El período del eclipse, que coincide con el período orbital del sistema, puede variar desde minutos hasta años.
- **Estrellas variables eruptivas:** que varían en brillo debido a procesos violentos y llamaradas que ocurren en sus cromosferas y coronas.
- **Giratorias:** muestran pequeños cambios en la luz que puede ser debido a las manchas oscuras o brillantes, o manchas en sus superficies. Estas son a menudo sistemas binarios.

Estas estrellas se pueden clasificar de la siguiente manera, figura 3.4:

Figura 3.4: Clasificación de objetos variables

También existen otro tipo de objetos variables, por ejemplo, un pulsar que puede tener un periodo definido. La siguiente gráfica muestra el periodo de la estrella R And, en el periodo de 1899 al 2000. Ésta es una estrella típica con variaciones de 2.5 en magnitud y periodos entre 80 a 1000 días, la curva de luz se muestra en la figura 3.5 .

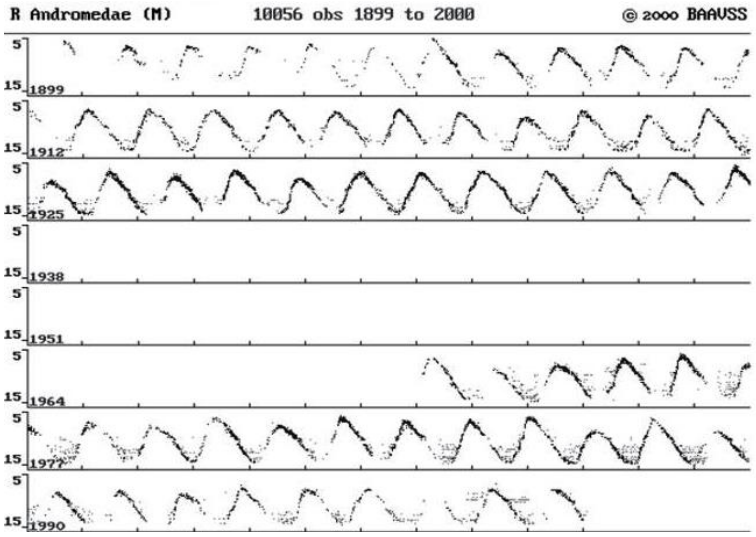


Figura 3.5: Curva de luz de un pulsar

Por otra parte se tienen las estrellas eclipsantes, que son sistemas binarios, donde una estrella eclipsa la luz que se recibe de la otra, ya que giran en un sistema que tiene un ángulo de inclinación adecuado para que podamos observar este fenómeno, la curva de luz de un fenómeno de este tipo se observa en la figura 3.5.

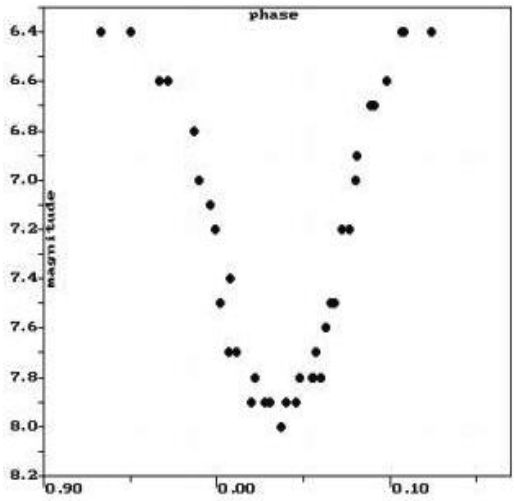


Figura 3.6: Curva de luz de la estrella eclipsante RZ Cas. Estimación del periodo del eclipse con base en varias observaciones.

El grupo de estrellas eruptivas, donde se encuentran agrupadas las estrellas T-Tauri, estrellas R-Corono Borialis, novas, supernovas, rayos gama, pulsares, entre otras. Éstas son de especial interés ya que existen cerca de los agujeros negros y están relacionadas con el origen de sistemas proto-planetarios[?]. La curva de luz de RS Op que ha mostrado un comportamiento recurrente de nova en 1933,1858, 1967 y 1985 se muestra en la figura 3.7.

Su magnitud incrementa entre 2 y 6 unidades, mientras que una nova suele hacerlo entre 6 y 15 unidades.

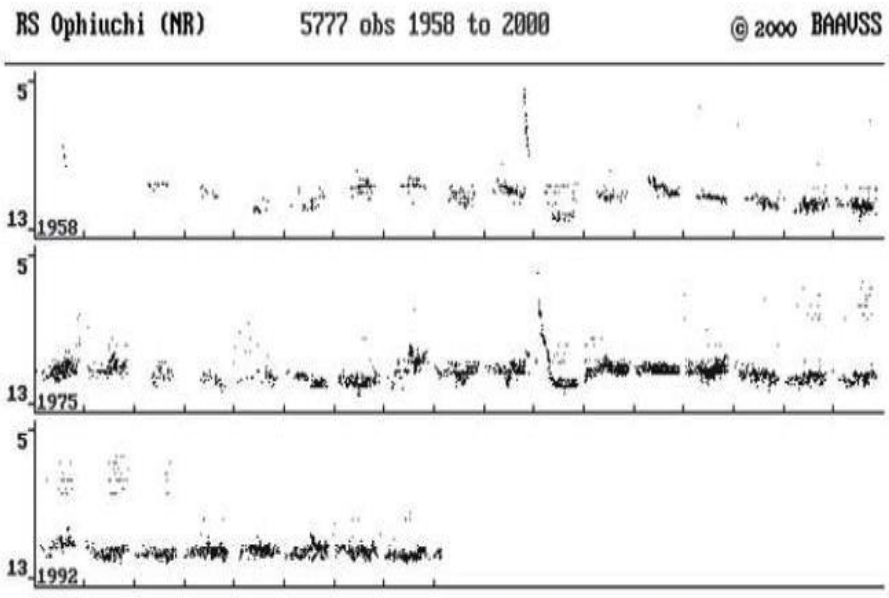


Figura 3.7: Curva de luz de RS Op, que ha mostrado un comportamiento recurrente de nova en 1933,1858, 1967 y 1985. Su magnitud incrementa entre 2 y 6 unidades, mientras que una nova suele hacerlo entre 6 y 15 unidades.

También existen otros objetos astronómicos variables, como los asteroides, que son una serie de objetos rocosos o metálicos que orbitan alrededor del Sol. La mayoría se encuentran en el cinturón principal, entre Marte y Júpiter, la curva de luz del asteroide Penélope se muestra en la figura 3.8. Y los cometas que son cuerpos celestes constituidos por hielo, polvo y rocas que orbitan alrededor del Sol, siguiendo diferentes trayectorias elípticas, parabólicas o hiperbólicas.

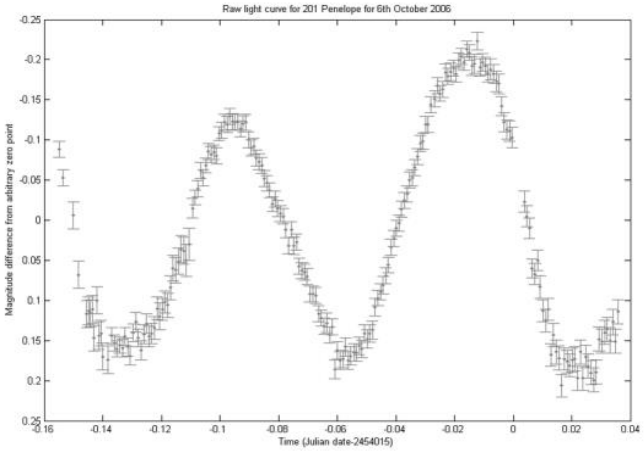


Figura 3.8: Curva de luz del asteroide Penélope en octubre del 2006

3.1.3. Supernovas

Existe diferentes tipos de supernova acorde a su proceso de formación,éstos proceso son: la detonación de una estrella masiva cuando esta ha agotado el hidrógeno en su núcleo y la segunda ocurre cuando un par de estrellas binarias interactúan, dónde un de las compañeras es una enana blanca que acreta materia de su compañera que puede ser una estrella super-gigante. Si el material acretado llega a alcanzar el límite crítico, llamado el límite de Chandrasekhar, la estrella puede tener una detonación termonuclear.

Con esta explosión la estrella lanza a su alrededor la mayor parte de su masa a altas velocidades. Después de este fenómeno explosivo se pueden producir dos casos: la estrella es completamente destruida o deja un remanente. Por lo regular las explosiones que dan lugar a las supernovas se producen en otras galaxias, es difícil observar este fenómeno en la vía láctea. La última supernova que se registró en nuestra galaxia ocurrió hace más de cien años.

El proceso de creación de una supernova tiene lugar en el núcleo de la estrella, esta estrella debe ser por lo menos de siete masas solares, donde el helio que contiene el núcleo de la estrella es transformado en carbono y oxígeno en su totalidad. Entonces el núcleo se vuelve incapaz de contenerse lo que hace que se incremente la temperatura. En algunos casos se producen elementos más pesados y el núcleo continua incrementado su temperatura en un proceso de fusión e implosión, figura 3.9,[18]. Hasta que ocurre un desequilibrio y el núcleo no resiste la fuerza de gravedad. Entonces ocurre la explosión por proceso termonuclear, en la que en muchas ocasiones el brillo de la nueva supernova iguala al de la galaxia en la que reside.

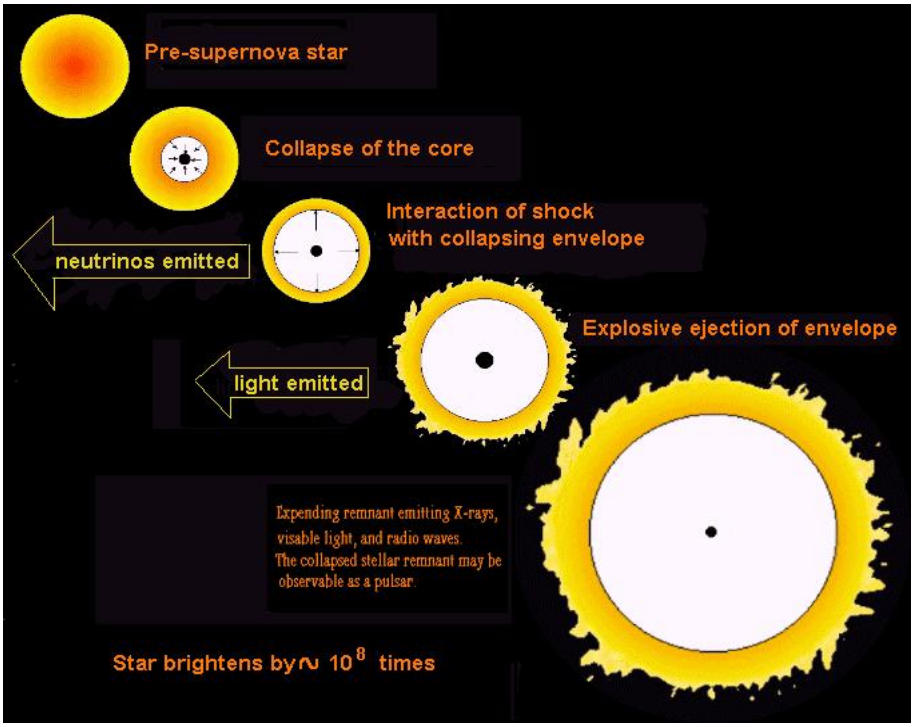


Figura 3.9: Proceso de creación de una supernova

Existen otras formas en las que puede surgir una supernovas, a través de un proceso gravitacional, por ejemplo, en sistemas binarios donde dos estrellas giran en torno a un mismo punto. Una de ellas acreta materia de la otra estrella, de tal manera que llega a un punto donde acumula demasiada materia y sufre un desequilibrio que produce una supernova [18].

3.1.3.1. Tipos de supernova

Las supernovas se clasifican como de Tipo I o Tipo II dependiendo de la forma de sus curvas de luz y la naturaleza de sus espectros. El proceso de nacimiento de una supernova se puede observar en la figura 3.10.

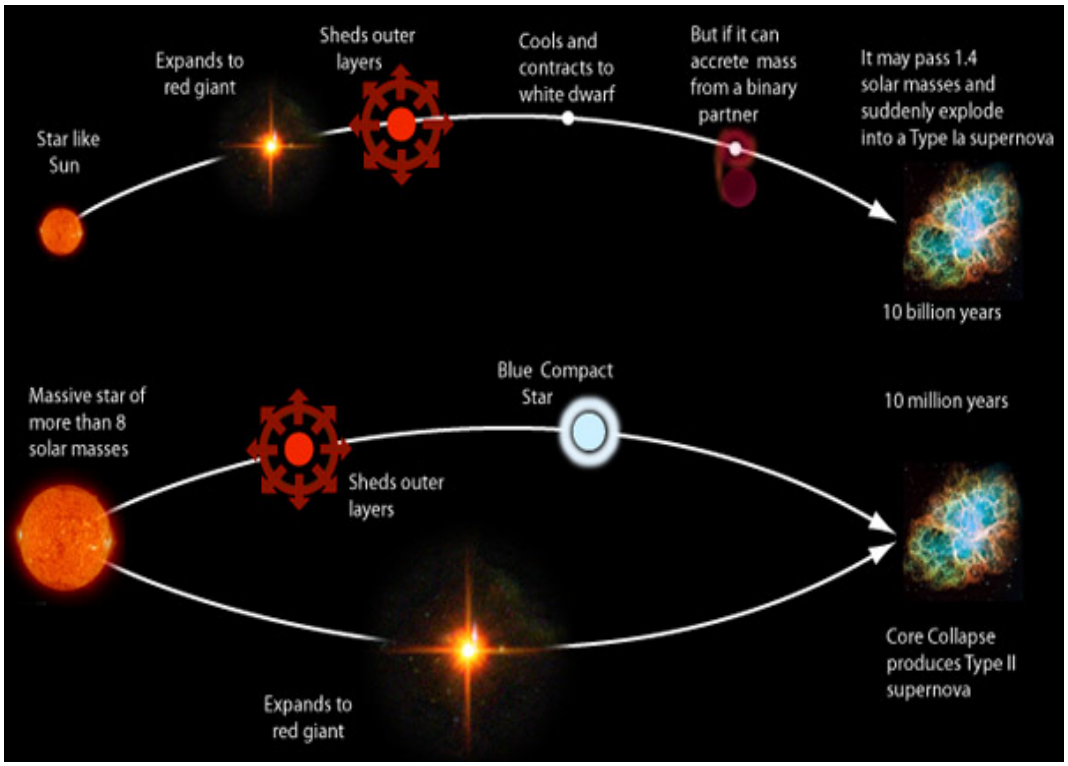


Figura 3.10: Síntesis de los elementos pesados, se cree que ocurre en las supernova y dan origen a su tipo.

Las supernovas se clasifican como Tipo I si sus curvas de luz presentan máximos agudos y luego se desvanecen gradualmente. Los máximos puede ser de aproximadamente 10 mil millones de luminosidades solares. Las supernovas de tipo II tienen un menor pico agudo en su máximo de aproximadamente mil millones de luminosidades solares. Mueren más pronunciadamente que las de tipo I. Las supernovas de tipo II no se observan que se produzcan en las galaxias elípticas, y se cree que se producen en estrellas de tipo de Población I en los brazos espirales de las galaxias. Las supernovas de tipo I ocurren típicamente en las galaxias elípticas, por lo que son, probablemente, estrellas de Población II.

El modelo para la iniciación de una supernova de Tipo I es la detonación de una enana blanca de carbono, que colapsa bajo la presión de degeneración de los electrones. El hecho de que los espectros de las supernovas de tipo I sean pobres en hidrógeno es consistente con este modelo, ya que la enana blanca no tiene casi nada de hidrógeno. El decaimiento suave de la luz es también consistente con este modelo ya que la mayor parte de la producción de energía, sería de la desintegración radiactiva de los elementos pesados inestables producidos en la explosión.

Las supernovas de tipo II se modelan como eventos de implosión-explosión de una estrella masiva. Muestran una planicie característica en su curva de luz unos meses después de la iniciación. Esta planicie es reproducida por los modelos de ordenador, que suponen que la energía proviene de la expansión y enfriamiento de la envolvente exterior de la estrella cuando es arrojada al espacio. Este modelo está corroborado por la observación de hidrógeno fuerte y espectros de helio de las supernovas de Tipo II, en contraste con las de Tipo I.

Como el proceso de explosión de una supernova no siempre es el mismo, tampoco lo es su curva de luz. En la figura 3.11 se observa la curva de luz para cada tipo de supernova.

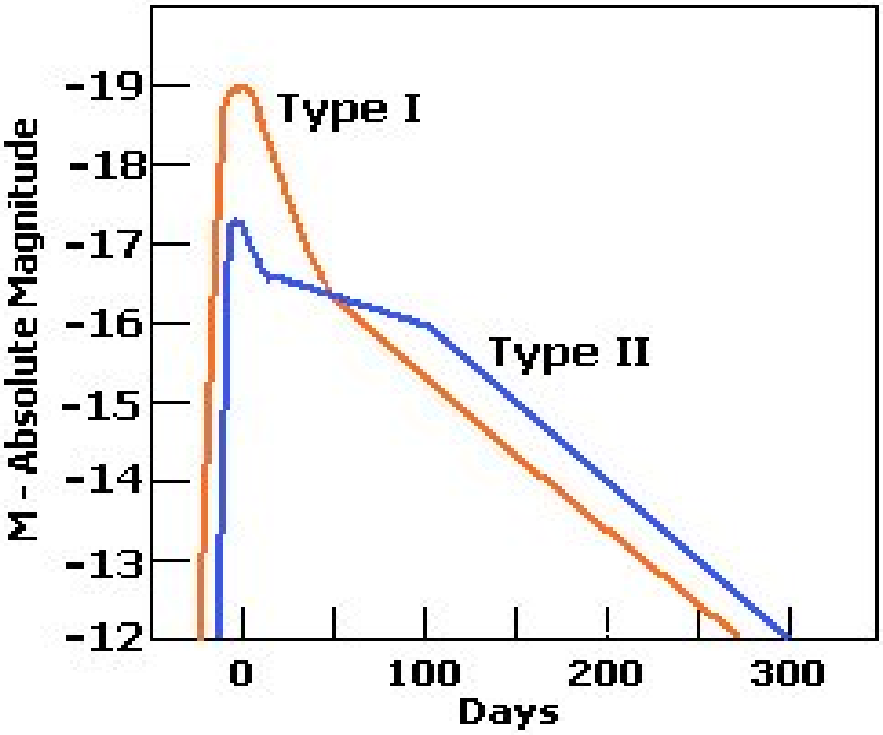


Figura 3.11: Curva de luz de los modelos de supernova.

3.1.4. Imágenes en la astronomía

A pesar de que las imágenes astronómicas son en escalas de grises con un rango dinámico grande, existen diferentes tipos de imágenes usadas para diferentes propósitos y con diferentes características.

3.1.4.1. Tipos de imágenes

Los objetos astronómicos que contienen gas, polvo y otros elementos pueden ser vistos en frecuencias específicas, algo común en la astronomía es tomar imágenes multi-espectral para estudiar los fenómenos con diferentes fuentes de información, la figura 3.12 muestra una serie de imágenes vistas desde diferentes longitudes de onda.

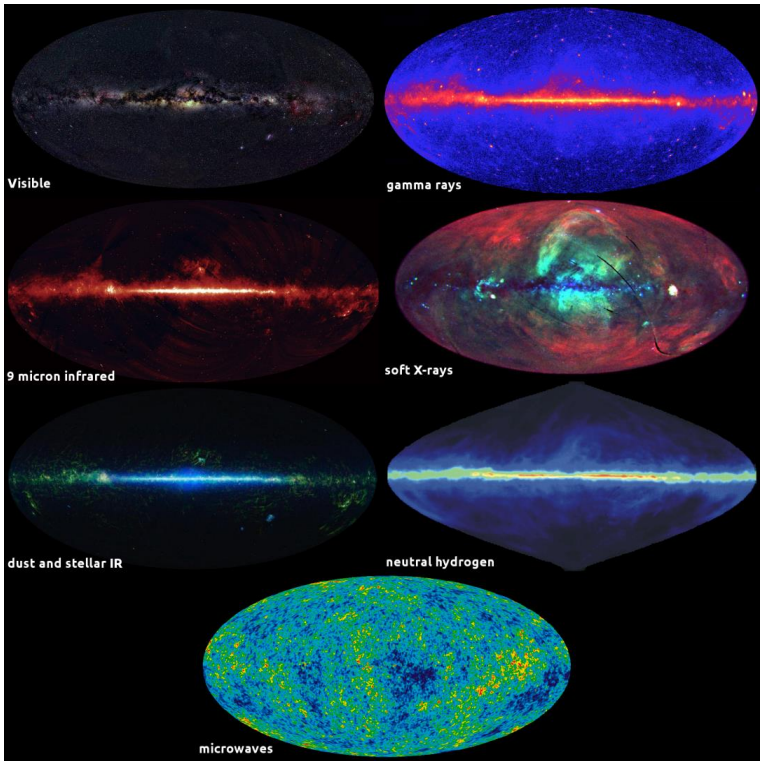


Figura 3.12: Objeto astronómico visto en diferentes frecuencias

Se han construido telescopios terrestres en las diferentes bandas según el área de interés, de los cuales se han creado diferentes proyectos de observación, así como telescopios espaciales. En México se cuenta con observatorios en la banda de lo visible, radio astronomía y también la banda milimétrica con el GTM (Gran telescopio milimétrico) .

Otros ejemplos de telescopios que trabajan en otras bandas son el Spitzer Space Telescope, que trabaja en la banda infrarroja o el Chandra X-ray Observatory [20].

3.1.5. Formato FITS

FITS es el acrónimo de Flexible Image Transport System y es el formato de datos estándar que los astrónomos usan, ya que a diferencia de otros formatos de imagen este permite almacenar datos adicionales de fotometría y calibración [31]. Las imágenes consisten de arreglos multidimensionales y dos tablas bidimensionales que contienen los datos en específico. También puede contener información como espectro electromagnético, listas de fotones entre otras. Puede contener exposiciones de la misma región del cielo en diferentes bandas.

El formato es usado como un arreglo de dimensiones arbitrarias y algunos encabezados. Consiste de un HDU (header and data units), el principal contiene la información de los pixeles de la imagen. Las extensiones de la imagen serian HDU extendidas. El encabezado contiene información que es posible analizar para entender el contenido de la imagen; algunos de los datos que contiene son: tamaño, fecha y hora, origen, coordenadas, formato de datos, comentarios, historial de los datos y algunos otros detalles.

3.1.6. Sistema de coordenadas celestes

La práctica más común en la astronomía es medir la posición de los objetos en el sistema de coordenadas celestes en un sistema esférico, el más común es el sistema de coordenadas ecuatorial, que consiste en la proyección de la latitud y longitud de la tierra en la esfera celeste. Las coordenadas son expresadas en como un par latitud-longitud. La latitud es llamada declinación y se mide de -90º a 90º. La longitud es llamada ascensión recta y va de 0º a 360º, la figura 3.13 muestra la proyección de las coordenadas a la esfera celeste.

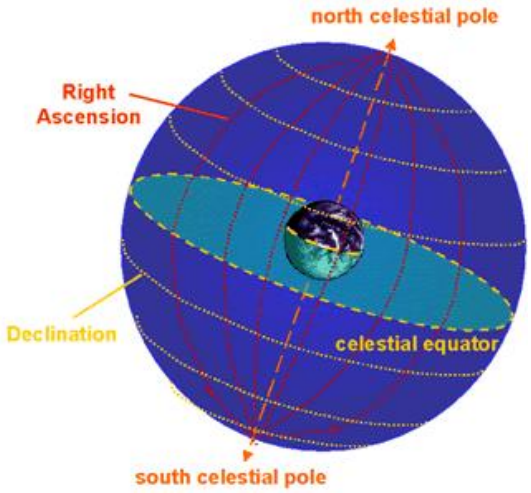


Figura 3.13: Sistema de coordenadas astronómicas

El sistema de coordenadas galácticas es otro sistema también muy usado, donde se pone al sol como origen en el plano galáctico. En este sistema se mide la distancia debajo del plano galáctico y es llamada latitud, la longitud es la distancia angular al plano galáctico.

3.1.7. Ruido de fondo

En la astronomía se habla del cielo o del fondo sin distinción, es la región donde no hay objetos presentes o la luminosidad de los objetos es poca. Las imágenes astronómicas están influidas por la atmósfera, a veces contaminada por las luces, por ejemplo, de las ciudades. En otras ocasiones el ruido no puede ser detectado por el ojo humano, ya que esta en otra frecuencia, para ello se usan algunas herramientas computacionales. De cualquier manera, el fondo de la imagen es difícil de determinar y por ende no se pude determinar con exactitud cada una de las fuentes, en algunos casos inclusive el fondo no es homogéneo y requiere largos

tiempos de exposición, ya que las condiciones atmosféricas cambian y las imágenes se toman en forma de mosaicos que lucen como la figura 3.14.

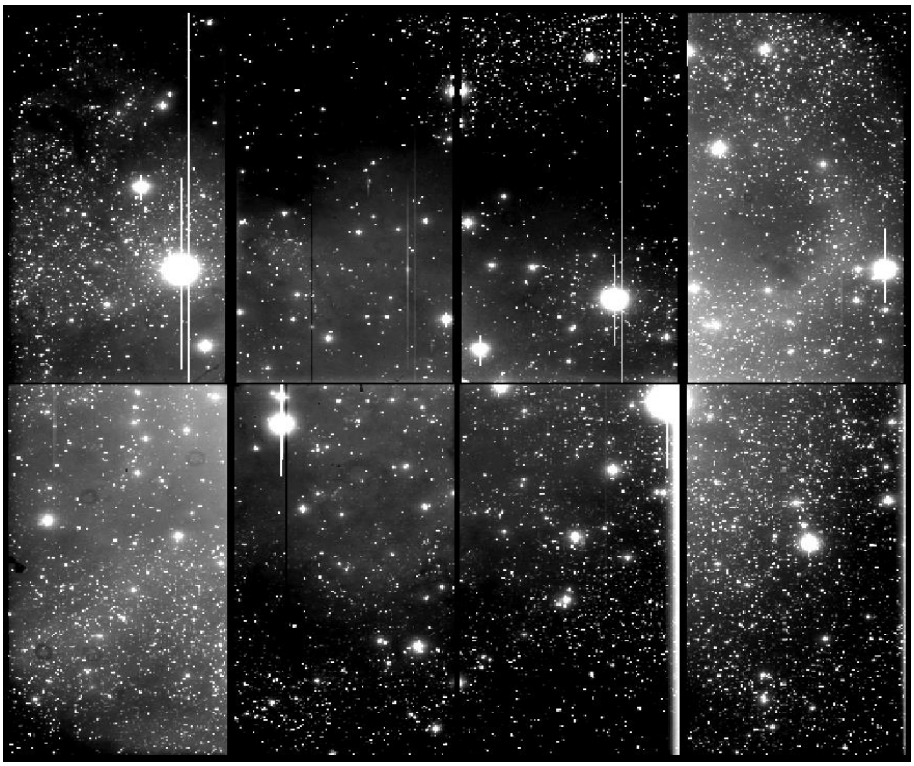


Figura 3.14: Imagen astronómica en forma de mosaico con ruido de fondo.

Como se observa, las imágenes astronómicas tienen ruido de diferentes fuentes, desde las variaciones aleatorias en el número de fotones de luz, variaciones térmicas en el momento de la adquisición, hasta en la conversión analógica digital. Por lo general, el ruido que se tiene es tomado como Gaussiano aditivo o Poisson, en general se asume que es de tipo gaussiano. Una forma de medir este ruido es SNR (Signal-to-Noise) que cuantifica cuanto se corrompió la señal dada, se calcula dividiendo la cuantificación de señal sobre la cuantificación del ruido. Se toman en cuanta todas las posibles fuentes de ruido, inclusive la del objeto observado.

Algunos otros rasgos importantes a tomar en cuenta son el ruido intrínseco en los CCD's así como la luz de la luna.

3.1.8. Proceso de observación y extracción de fuentes

El pre-procesamiento de las imágenes astronómicas tiene como subprocesos la reducción de los datos, a través de varios pasos, para generar un catálogo con los diferentes objetos contenidos en una región del cielo. Un catálogo es una lista de objetos que contiene algunas de las principales características como es el tipo, morfología y ubicación, por lo regular es la salida de la lectura de alguna región o del procesamiento de una imagen.

3.1.8.1. Adquisición

Las observaciones se llevan a cabo en los observatorios por telescopios construidos para este propósito. Los telescopios más usados son los que trabajan en la banda del visible, ultravioleta e infrarrojo. Estos estan constituidos por dos o tres espejos o lentes para conseguir un lectura lo más exacta posible. En algunos casos se usan filtros para trabajar sobre zonas específicas del espectro electromagnético. Los observatorios se encuentran en sitios específicos donde las condiciones son las más adecuadas posibles para evitar el ruido. En otros casos, ya que la atmósfera altera las lecturas, se instalan telescopios en aeronaves o satélites. Otro de los instrumentos frecuentemente usados es el CCD, que es sumamente sensible para realizar lecturas sobre todo en el rango de lo visible. Las cámaras CCD tienen un arreglo de sensores, cada sensor corresponde a un pixel de la imagen, de donde se hace la conversión analógica-digital acorde a la intensidad leída por el sensor, la figura 3.15 muestra el proceso de conversión análogica-digital de una lectura con CCD.

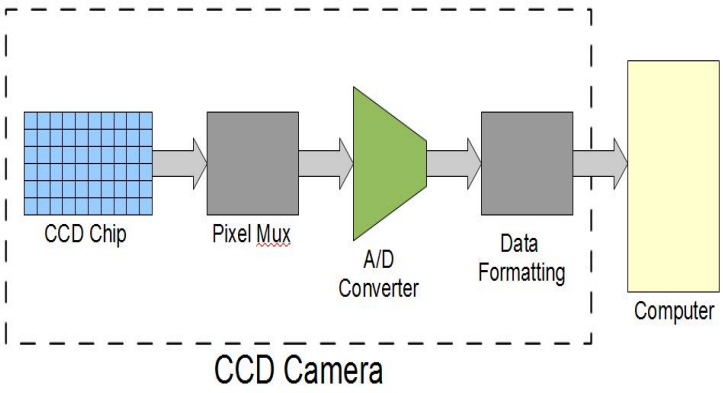


Figura 3.15: Proceso de adquisición de una imagen FITS

Por otra parte las ondas de radio son capturadas por antenas de radio direccionales. Estas también pueden usarse en forma de arreglo.

3.1.8.2. Pre-procesamiento

El pre-procesamiento de las imágenes puede hacerse en varios pasos. El primer paso es la calibración de CCD con una imagen sin exposición a la luz, que por consecuencia debe ser cero en todos los sensores. Después se realizar la lectura de una imagen con una exposición a la luz homogénea. Algunas veces los pixeles tienen errores o se corrompen en la lectura, por algún error o rayo cósmico. Para corregir estos datos atípicos se usan filtros o se reconstruye la imagen. Adicionalmente, para corregir las variaciones de fondo se usan la media o mediana de la lectura para cada pixel.

3.1.8.3. Extracción

La detección de las fuentes es el primer paso de la extracción, que en otras palabras es la adquisición de las propiedades y características de los objetos presentes en una imagen

astronómica. Una vez que las fuentes son localizadas, se mide la intensidad conocida como flux, este proceso se conoce como fotometría y provee información sobre las estructuras de los objetos, su temperatura, distancia y edad. Las técnicas de fotometría se usan para medir la luz emitida por un objeto, existen diferentes formas para realizar este proceso, por ejemplo por apertura o diferencial. La técnica de apertura mide el nivel del cielo promediando, la intensidad de los pixeles del centroide y se realizan aperturas circulares hasta cubrir todo el objeto, figura 3.16. El método diferencial, en cambio, usa como referencia fuentes con brillo constante y calcula la diferencia relativa respecto a las fuentes a extraer. Esta técnica se usa para determinar la evolución de estrellas variables.

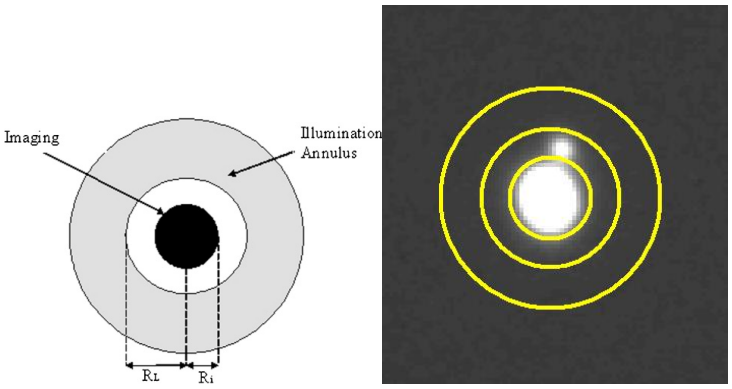


Figura 3.16: Extracción de objetos por aperturas circulares

La magnitud del objeto es la característica más importante, ésta se calcula con base en el brillo medido. La determinación de la posición de los objetos se hace a través de las coordenadas celestes, este proceso es llamado astrometría, donde se determina la posición de los objetos en las imágenes. Algunos de los proceso de extracción también realizan la clasificación de objetos estrella/galaxia o la determinación de la morfología de las galaxias

3.1.9. Software de detección

La comparación cuantitativa de los objetos ya se encuentra implementada en el software de uso en la astronomía, la figura 3.17 muestra los programas más populares, los métodos de transformación de imágenes implementados, el criterio de detección de objetos y el tipo de datos con el que trabajo.

Name	Reference	Image transformation	Detection criterion	Aim
		Basic		
SExtractor[c*]	Bertin & Arnouts (1996) [6]	σ-clipping	Global thresholding	Multiband
SAD (AIPS)[f*]	Greisen (2003) [29]	-	Global thresholding	Radio
Distilled Sensing[m]	Haupt et al. (2009) [35]	Distilled Sensing	Global thresholding	Multiband
Astrometry.net[c*]	Lang et al. (2010)[50]	Median filter	Global thresholding	Multiband
Perret[N/A]	Perret et al. (2010) [77]	-	Connected components trees	Multiband
		Matched filtering		
Mopex[p*]	Makovoz & Marleau (2005) [57]	Median filter + Matched filter	Global thresholding	Infrared
		Bayesian		
SourceMiner[s]	Savage & Oliver (2007) [81]	Bayesian filter	Local peak search	Infrared
		Multi-scale		
González-Nuevo[m]	González-Nuevo et al. (2006) [26]	Mexican hat wavelet family	Local peak search	Cosmic microwave background

Figura 3.17: Software en la astronomía

3.1.9.1. Picture Processing Package

Picture Processing Package(PPP) es un software que identifica objetos astronómicos automáticamente, calcula la magnitud total y clasifica objetos en dos clases (estrella-galaxia) [32].

El programa en general usa la curva de crecimiento con aperturas circulares concéntricas y la propiedad Kron para realizar la clasificación de objetos (estrella-galaxia), también hacer fotometría integrada. Algunas de sus funciones son: la carga de datos, manipulación, análisis en imágenes directamente con énfasis en el trabajo con galaxias. Funciona en sistemas Unix y está escrito en Fortran 77. El sistema trabaja sobre línea de comandos y modo batch. Sus principales funciones se agrupan en cuatro áreas:

- Comandos de entrada y salida de datos, trabaja con formato FITS de imágenes.
- Manipulación de imágenes, cuenta como comando básicos de manipulación y filtrado.
- Despliegue de imágenes.
- Análisis de imágenes, es la parte más importante del software, cuenta con comandos de fotometría, estadística, localización de objetos y clasificación.

3.1.9.2. Descripción

Se basa en un filtrado (pasa bajas) y determina pixel a pixel los picos de intensidad, considerando un umbral mínimo que depende del valor del cielo, de esta manera se calcula el centroide del objeto, para calcular el ruido de fondo se calcula la media y mediana de fondo de la imagen y se estima el valor de fondo, figura 3.18 [32].

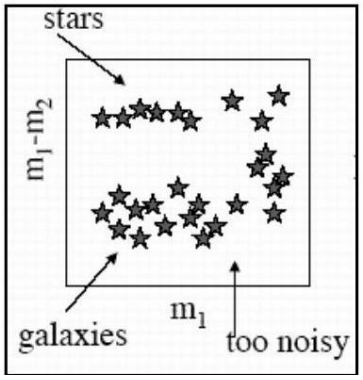


Figura 3.18: Detección de objetos por PPP

El proceso de clasificación se lleva a cabo comparando la forma que tiene la curva de crecimiento de un objeto que se va a clasificar con la de una estrella de referencia (brillante y aislada), figura 3.19. Para calcular esta curva de crecimiento se realizan aperturas en radios circulares concéntricos al centroide del objeto, de esta manera se va calculado el flujo y por ende la curva de crecimiento, este proceso se repite hasta alcanzar el fondo de la imagen. Se calcula el parámetro C_2 ,que calcula la diferencia promedio por apertura entre dos curvas de crecimiento, luego de que éstas son escaladas entre sí.

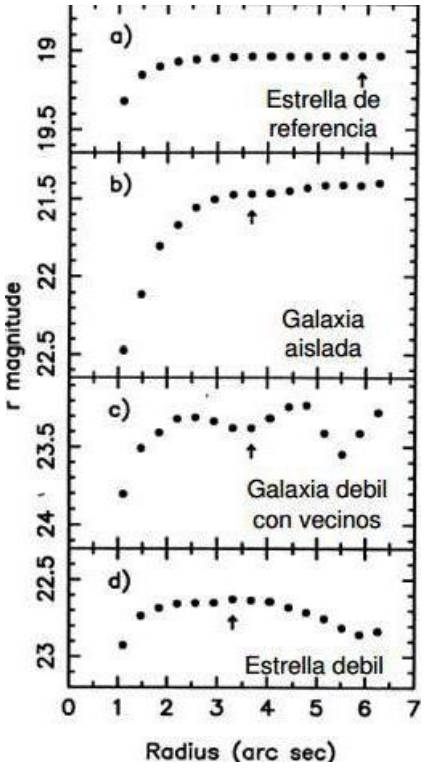


Figura 3.19: La imagen ilustra las curvas de crecimiento de diferentes objetos astronómicos.

La clasificación de los objetos depende del ruido del cielo y también de la proximidad de objetos brillantes. Éste es factor que puede influir es la falta de resolución de la imagen.

La clasificación del objeto está dada por la comparación de su curva de crecimiento con la de un objeto conocido, calculado C_2,figura 3.20,como se observa en la siguiente imagen, el parámetro calculado nos da la clasificación del objeto.

$$C_2 = \frac{1}{N_A - 2}\sum_{i=3}^{N_A}(m_i^* - m_i) - C_0$$

N_A = *Apertura óptima*
m_i = *magnitud instrumental*
m_i^* = *magnitud instrumental de referencia*
c_0 = *constante de normalización (tiene en cuenta la diferencia de magnitud entre el objeto y la estrella de referencia)*

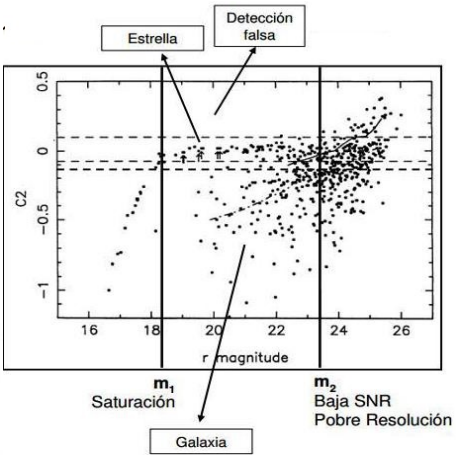


Figura 3.20: Ecuación para calcular C_2 y gráfica de clasificación por PPP

El paquete está desarrollado en FORTRAN 77 para Unix. Trabaja sobre línea de comando, recibe como entrada las imágenes en formato FITS y un script de comandos. La información que entrega puede ser en formato gráfico en pantalla o un catálogo con la caracterización de los objetos extraídos.

3.1.10. Conjunto de datos

El conjunto de datos para realizar los experimentos provienen del proyecto CHASE-CALAN (CHilean Automatic Supernovae Search)[23], con el objetivo científico de estudiar el origen de la aceleración del Universo a partir de las distancias medidas con supernovas de galaxias cercanas. Consiste en observar muestras de 10.000 galaxias escogidas a través de la colección PROMPT de telescopios robóticos instalados en Cerro Tololo. Tiene una magnitud limitada de 18 unidades con 40 segundos de exposición. Se cuenta con imágenes de aproximadamente 2 años de seguimiento de observación.

3.2. Conceptos computacionales

En esta sección se expondrán los conceptos y técnicas computacionales que implican el desarrollo de este trabajo. Iniciaremos con el ámbito de señales; pensando en las curvas de luz como señales a ser analizadas y en cuyo caso requiere procesos de filtrado y normalización. Para terminar, se explorará el área de clasificación y sus principales componentes.

3.2.1. Análisis de señales en la astronomía

El procesamiento digital de señales se distingue de cualquier área de la ciencia ya que en algunos casos los datos se originan de algún sensor del mundo real. A este tipo de señales se

les conoce como analógicas, después vienen un conjunto de algoritmos y técnicas matemáticas para convertir estas señales a digitales. El uso de señales tiene un amplia gama de aplicaciones, algunos ejemplos de las áreas de aplicación son: ciencias aeroespaciales, medicina, telefonía, aplicaciones militares, industrial y en la ciencias exactas. A continuación se definen los conceptos y técnicas del procesamiento digital de señales que se usan en le presente trabajo.

3.2.1.1. Señales

Una señal es la descripción de la relación entre dos parámetros. Un ejemplo típico es la variación del voltaje respecto al tiempo. Una señal que tiene sus parámetros cuantizados es llamada señal digital. En nuestro caso, curva lumínica de cada objeto puede ser vista como una señal, ya que el parámetro de medición es el flujo del objeto en una imagen y que se puede explorar a través de una serie de imágenes, es decir a través del tiempo.

En la astronomía este tipo de análisis de señales es conocido como timing analysis, donde se intenta hacer un análisis de las propiedades dinámicas de los objetos. En la imagen 3.21 podemos observar una curva de luz de un pulsar. Generalmente este tipo de señales se analiza con transformada de fourier para visualizar la curva como un gran conjunto de componentes. Muchas de las propiedades de las señales que se manejan estan dadas por el instrumento con que el fueron adquiridos los datos, depende la banda del espectro con que se esta trabajando, además de los periodos de casa señalas estarán dadas por el tiempo de observación de cada región, con esto se pude intuir que existirán perdidas de datos.

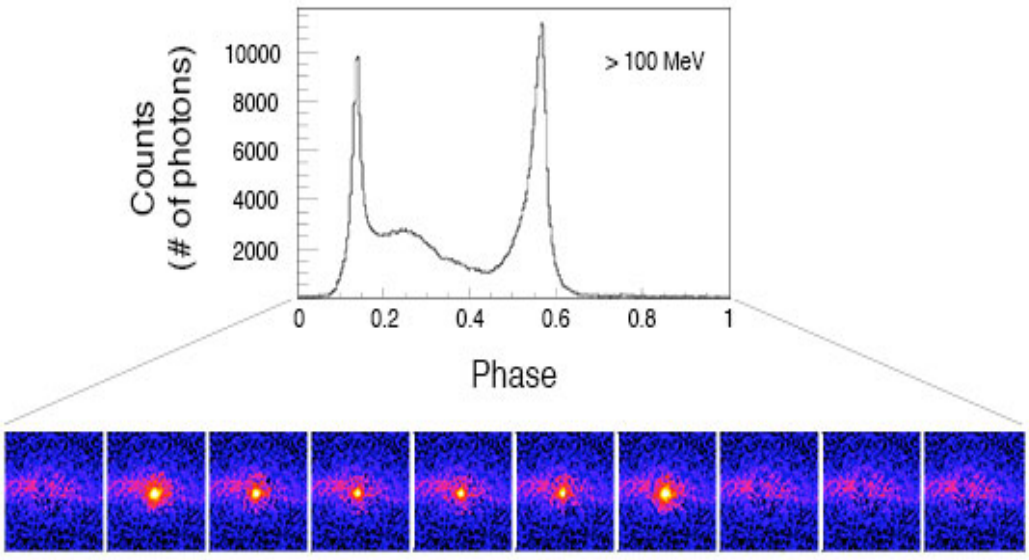


Figura 3.21: Esta imagen muestra el lapso de tiempo de pulsar Vela y como se transforma a una curva de luz. Grodin et al. 2013 ApJ NASA/DOE/Fermi LAT

El análisis de señales a menudo no conoce que es lo que esta buscando, inclusive si lo que se esta analizando es realmente una señal o solamente ruido. Para trabajar con las señales se

utiliza una una serie de herramientas, que en su mayoría estan basadas en la transformada de Fourier, estas son realmente útiles para analizar señales periódicas.

En astronomía, la mayor parte de lo que conocemos del universo proviene de fuera de la atmósfera terrestre, este es el estudio de la emisión electromagnética de diversos objetos. Dependiendo de su temperatura estos objetos emiten en una amplia gama de frecuencias(ondas de radio, radiación infrarroja, luz visible, ultravioleta, rayos X, rayos gamma). Muchas veces la emisión tiene variaciones interesantes que pueden ayudar a entender los procesos físicos del objeto que la emite.

3.2.1.2. Series de tiempo

Un señal física usualmente se puede pensar como una función continua en el tiempo, x(t) donde x representa una cantidad física como la intensidad de luz o la amplitud de una onda sonora en un punto t determinado. El proceso de observación o medición convierte esto en una serie de valores discretos $x_0, x_1, x_2...$ de una señal física que por lo general es continua en el tiempo. y que es llamada serie de tiempo. El proceso de convertir una función continua a una secuencia discreta es llamado muestreo.

Un serie de tiempo por lo general es una sucesión de puntos en el tiempo, separados por una intervalo constante de tiempo $\triangle t$. Por lo tanto $t_k = t_0 + k\triangle t$ y $x_k = x(t_k) = x(t_0 + k\triangle t)$. $\triangle t$ es llamado intervalo de muestreo. El número de muestras por unidad de tiempo esta dado por $f_s = 1/\triangle t$ que es llamado frecuencia de muestreo.

Por ejemplo, si la intensidad de una estrella esta medida una vez cada minuto, obtenemos una serie de tiempo con un intervalo de muestreo de $\triangle t = 1min = 60s$, y una frecuencia de muestreo de $f_s = 0.016667Hz = 16.667mHz$.

Algunas de las características sobresalientes en el análisis de las series de tiempo son:

- Dimensionalidad: son los grados de libertad de las series de tiempo, es decir en el espacio en que trabajan(binario, real u otro).
- Tamaño: la cantidad de datos que la conforman.
- Representación: Puede ser representada en un plano cartesiano, donde Y representa la magnitud y X el índice consecutivo que corresponde a cada valor que puede ser el tiempo.
- Estructura: Las series de tiempo contienen picos los cuales no son derivables ni integrables.

Matemáticamente una serie de tiempo puede ser representada por la suma de senos y cosenos, sin embargo las series de tiempo de fenómenos naturales raramente tienen oscilaciones tan simples. En general suelen ser más complicadas, ya que contienen picos estrechos o alargados y generalmente está compuesta de diferentes oscilaciones de onda.

Algunos de los principales problemas que tiene las series de tiempo al ser comparadas son: el ruido, desfase y escala. Para solucionar algunos de estos problemas es común aplicar algunas técnicas del área de análisis de señales como son el filtrado, interpolación y normalización de las series, estos temas se exponen a continuación.

3.2.1.3. Filtrado

Un filtro es un sistema continuo o discreto que sirve para procesar señales.El filtrado modifica el espectro de las señales de entrada de acuerdo a ciertas especificaciones. Un filtro digital consiste en un proceso computacional implementado con circuitos y/o programación, en el cual una secuencia numérica de entrada se transforma en otra secuencia numérica de salida con características predeterminadas. Matemáticamente, un filtro digital se representa por una ecuación diferencial. Sus características los hacen apropiados para un amplio campo de aplicaciones, entre las que se encuentra compresión de datos, procesamiento de señales biomédicas, procesamiento digital de audio, procesamiento de voz o procesamiento de imágenes.

Algunas de las ventajas de un filtro digital son la respuesta a la frecuencia más cercana a la ideal, no requieren sintonización, sus componentes son independientes de la frecuencia de operación del filtro. La máxima frecuencia de operación de un filtro digital es igual a la mitad de la frecuencia de muestreo(Teorema de Shannon).

Consisten fundamentalmente en un algoritmo mediante el cual una señal digital o secuencia numérica denominada "entrada" se transforma en una segunda secuencia de números denominada señal digital de salida.

Una amplia variedad de filtros digitales es descrita por medio de una ecuación de diferencias lineal de coeficientes constantes, que relaciona la secuencia de entrada del filtro x(n) y la secuencia de salida del mismo y(n) de la ecuación (1):

$$y(n) = \sum_{k=0}^{N} a_k y(n-k) + \sum_{k=0}^{M} b_k x(n-k) \tag{3.1}$$

Estos sistemas pueden ser representados por una secuencia de respuestas al impulso h(k) donde k=0,1,2..., y la señal de salida se obtiene a partir de operaciones de suma y convolución de dicha secuencia con la señal digital de entrada. En términos de su respuesta al impulso los filtros digitales se clasifican de dos formas: FIR(finite impulse response) o filtros de respuesta finita al impuso; e IIR (infinite impulse response) o filtros de respuesta infinita al impulso, que deben su comportamiento a la existencia de lazos de realimentación en su estructura.

Algunos ejemplos de filtrado de señales se muestran en la siguiente figura 3.22.

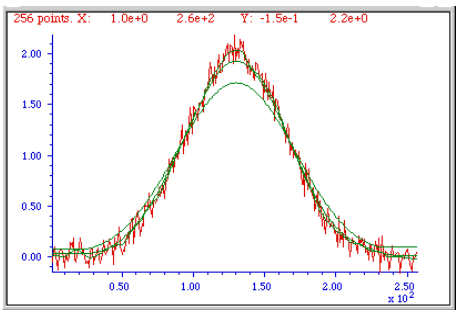


Figura 3.22: Ejemplo de filtrado de una señal

3.2.1.4. Interpolación

La interpolación, es decir, el ajuste de una señal que sirve para reconstruir, en forma aproximada una función a partir de sus muestras. Una forma común de hacer interpolación es la lineal, en cuyo caso dos puntos adyacentes son conectados mediante una línea recta. En casos más complejos los puntos pueden ser unidos por medio de polinomios de mayor orden o mediante otras funciones matemáticas.

Para una señal de banda limitada, si los instantes de muestreo son lo suficientemente cercanos, entonces la señal se puede reconstruir exactamente: en otras palabras, mediante el uso de un filtro pasa bajas se puede efectuar la interpolación exacta entre los puntos de la muestra. La interpretación de la reconstrucción de x(t) como un proceso de interpolación se hace evidente cuando se considera el efecto en el dominio del tiempo del filtro pasa bajas, en particular la salida es:

$$x_r(t) = x_p(t) * h(t) \tag{3.2}$$

o, con $x_p(t)$ dada la ecuación de muestreo

$$x_p(t) = \sum_{n=-\infty}^{+\infty} x(nT)\delta(t - nT) \tag{3.3}$$

donde x(t) es el dominio del tiempo
la frecuencia fundamental esta dada por $p(t) = \sum_{n=-\infty}^{+\infty} \delta(t - nT)$
Tenemos que la ecuación que describe el ajuste de una curva centre los puntos de la muestra es la siguiente:

$$x_p(t) = \sum_{n=-\infty}^{+\infty} x(nT)h(t - nT) \tag{3.4}$$

Donde la ecuación 4 describe el proceso de interpolación de una curva.

3.2.2. Reconocimiento de Patrones

El reconocimiento de patrones[10] es un subproceso dentro de un sistema de visión artificial, donde la entrada es un patrón natural y el resultado es una etiqueta. Las etapas funcionales del reconocimiento de patrones son:

- **Adquisición de datos:** La entrada de un sistema de reconocimiento de patrones es un vector numérico que contiene valores muestrados cuantificados de un sistema de señales naturales.
- **Selección y extracción de características:** Sustracción de información relevante para la clasificación.
- Clasificación El objetivo de un sistema de reconocimiento de patrones es el de etiquetar de forma automática patrones de los cuales desconocemos su clase. La construcción de un clasificador no es una tarea trivial ni directa e involucra una serie de etapas: elección del tipo de clasificador, aprendizaje y validación de los resultados.

3.2.3. Tipos de clasificación

El proceso de clasificar consiste en organizar y categorizar los objetos en clases o grupos diferentes. Entre las formas de clasificar podemos encontrar:

- **Clasificación supervisada:** conocida también como clasificación con aprendizaje. En este tipo de problemas ya se encuentran definidas las clases y se cuentan con objetos previamente clasificados, el objetivo es clasificar nuevos objetos.
- **Clasificación parcialmente supervisadas:** es la familia de problemas menos estudiada en reconocimiento de patrones; se conoce como aprendizaje parcial, en estos problemas se tienen clases y existen objetos solo en algunas de las clases.
- **Clasificación no supervisada:** también conocida como clasificación sin aprendizaje, en este tipo de problemas no existe ninguna clasificación previa de objetos y en algunos casos tampoco esta definido el número de grupos a formar. Consiste en encontrar la agrupación de los datos dado que puede ser restringida o libre.

3.2.3.1. Clasificación supervisada

La clasificación supervisada se caracteriza por incluir como conjunto de entrada muestras de conjuntos previamente clasificados. Podemos enunciar de manera general el concepto general de cubrimiento de clasificación:

Sea un cubrimiento del universo Ω la tupla (5)

$$(\Omega, \Re, \delta, Q, \pi, f) \tag{3.5}$$

Donde Ω es el universo no vacío de objetos conocidos, en el caso de clasificación supervisada se subdivide en dos conjuntos(supervisión y control).

$\Re$ es el conjunto de rasgos descriptivos de los patrones

δ es la relación funcional o relación de descripción en términos de R

Q Será el conjunto de clases o categorías

π Será la relación de pertenencia a cada clase

f Función de analogía

Entonces el objetivo es clasificar los objetos que pertenecen a la muestra de control. Un ejemplo común es calcular la distancia entre los objetos y clasificar acorde a la distancia existente entre el patrón y los ejemplos del conjunto de muestra. Existen dos tipos de funciones de distancia: de semejanza y diferencia. Dependerá de la aplicación la selección de esta función.

Los algoritmos de clasificación tienen tres etapas:

1. **Aprendizaje:** extraer información de los datos de entrada para realizar el proceso de clasificación.
2. **Síntesis:** resumir la información de aprendizaje.
3. **Regla de solución:** tomar la decisión de pertenencia.

En general un algoritmo de clasificación supervisada se puede plantear de la siguiente manera:

Algoritmo 3.1 Algoritmo general de clasificación supervisada

Entrada: Muestra de supervisión y muestra de control.
Salida: Clasificación de los patrones

1. Selección de una función de comparación.
2. Calcular la semejanza/diferencia entre cada patrón por clasificar y los patrones que pertenecen al conjunto de muestra.
3. Para cada patrón calcular la distancia promedio a cada clase.
4. Aplicar la regla de solución de clasificación para cada patrón.

3.2.3.2. Comparación entre series de tiempo

La función de distancia entre series de tiempo ayuda a determinar la semejanza entre los diferentes patrones. Este proceso requiere un análisis entre dos series de tiempo, además se requiere que la asignación se semejanza sea cuantificable. De ahí que podamos definir el proceso de medir como la asignación de un número a un fenómeno. La similitud entre dos objetos en nuestro caso series de tiempo, estará dada por la relación existente de su magnitud. Entonces la distancia que exista entre estas series será la medida de discrepancia entre ambas, basándonos en algunas características en concreto.

Entonces la distancia entre una colección de objetos se puede definir como la definición 3.6:

$$d = S \times S \rightarrow \mathbb{R} \tag{3.6}$$

Donde $S \times S$ es el producto cartesiano entre los objetos de la colección y $\mathbb{R}$ es el contradominio que pertenece a los reales. Entonces si consideramos $a, b, c \epsilon S$, entonces podemos decir que para cualquier función de distancia se debe cumplir que:

1. No negatividad$d(a, b) \geq 0$
2. Identidad $d(a, b) = 0 \Longleftrightarrow a = b$
3. Simetria $d(a, b) = d(b, a)$

Existen diferentes tipos de funciones de distancia que se pueden utilizar en el proceso de comparar series de tiempo, la más común es la distancia euclidiana o la Minkowski. Pero también existen procesos estadísticos como el eficientemente de correlación y algunos algoritmos que optimizan el cálculo como el Dynamic Time Warping(DTW).

Capítulo 4

Construcción del modelo de detección y clasificación

En este capítulo se expondrá el modelado del software para la clasificación de objetos, con este fin se propone la implementación de tres módulos a PPP como parte de su estructura.

- Interfaz gráfica del usuario
- Módulo de creación de bases de datos
- Módulo de selección de estrellas de referencia y normalización
- Módulo de detección de objectos variables(candidatos)
- Clasificación supernova/no-supernova

El software se integró en una aplicación a la que denominaremos SDS(Supernovae Detection Software).

4.1. Modelo conceptual del software

La figura 4.1 muestra la forma general de trabajar del software,antes de pasar a SDS.

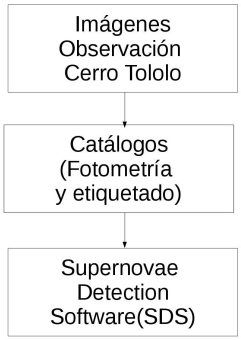


Figura 4.1: Los datos llegan del telescopio y son preprocesados por un programa externo antes de pasar a ser analizados por nuestro software.

Se implementó una interfaz gráfica para el usuario para el software PPP, ya que actualmente trabaja sobre línea de comandos o a través de scripts. PPP sirve para realizar el trabajo de preprocesamiento de las imágenes y produce catálogos que son usados por nuestro modelo para construir y analizar series de tiempo, la figura 4.1 muestra este concepto.

Se integraron los módulos anteriormente mencionados, la figura 4.2 muestra el modelo conceptual de la propuesta de software que lleva acabo la detección de objetos variables. Se muestran los procesos que se deben realizar para detectar las variaciones desde la fuente(telescopio) hasta el etiquetado de candidatos.

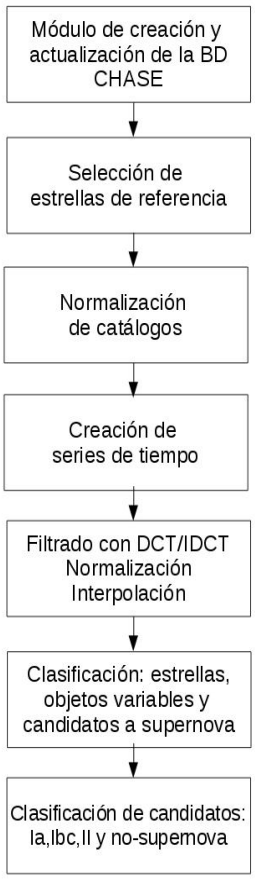


Figura 4.2: Procesos principales del software SDS

4.2. Extracción de las fuentes

Los catálogos de este trabajo fueron preprocesados por el grupo de astronomía del Dr. Giuliano Pignata en la Universidad Andrés Bello. Las imágenes en formato FITS tuvieron un proceso de registro a una imagen de referencia. Posteriormente se localizaron los centroides y se realizó la fotometría de cada imagen. La base de datos CHASE ya esta preprocesada y contiene los catálogos registrados.

Las funciones principales que se integraran al software son el etiquetado(Object Finding) y fotometría(photometry) de PPP. Con la intención de proporcionar una herramienta integral para la selección de candidatos a supernovas, que se pueda realizar desde los datos duros. El software PPP proporciona un catálogo similar al que se procesará con SDS.

Puesto que PPP ya puede realizar el trabajo de detección de fondo y caracterización de objetos. Algunas de las características que entrega PPP para cada objeto son:

- Etiqueta o número de objeto
- Ascensión recta
- Declinación
- Posición de centroide en X
- Posición de centroide en Y
- Magnitud aparente del objeto
- Radio de apertura Kron
- Elipticidad
- Posición angular
- Clasificación

Para PPP se implementó una interfaz gráfica del usuario desarrollada en TKinter para python, integrándolo a nuestro software de detección.

4.3. Creación de la base de datos

La base de datos se creo a partir de los catálogos de las imágenes. El software es capaz de crear la base de datos completa únicamente indicando el directorio donde se encuentran los catálogos, para la creación de la base de datos se utiliza el manejador MySQL, el programa genera automáticamente las sentencias SQL para crear tablas, almacenar datos y verificar la relación entre los diferentes catálogos, organizando el contenido de la base de datos como se muestra en la figura 4.3.

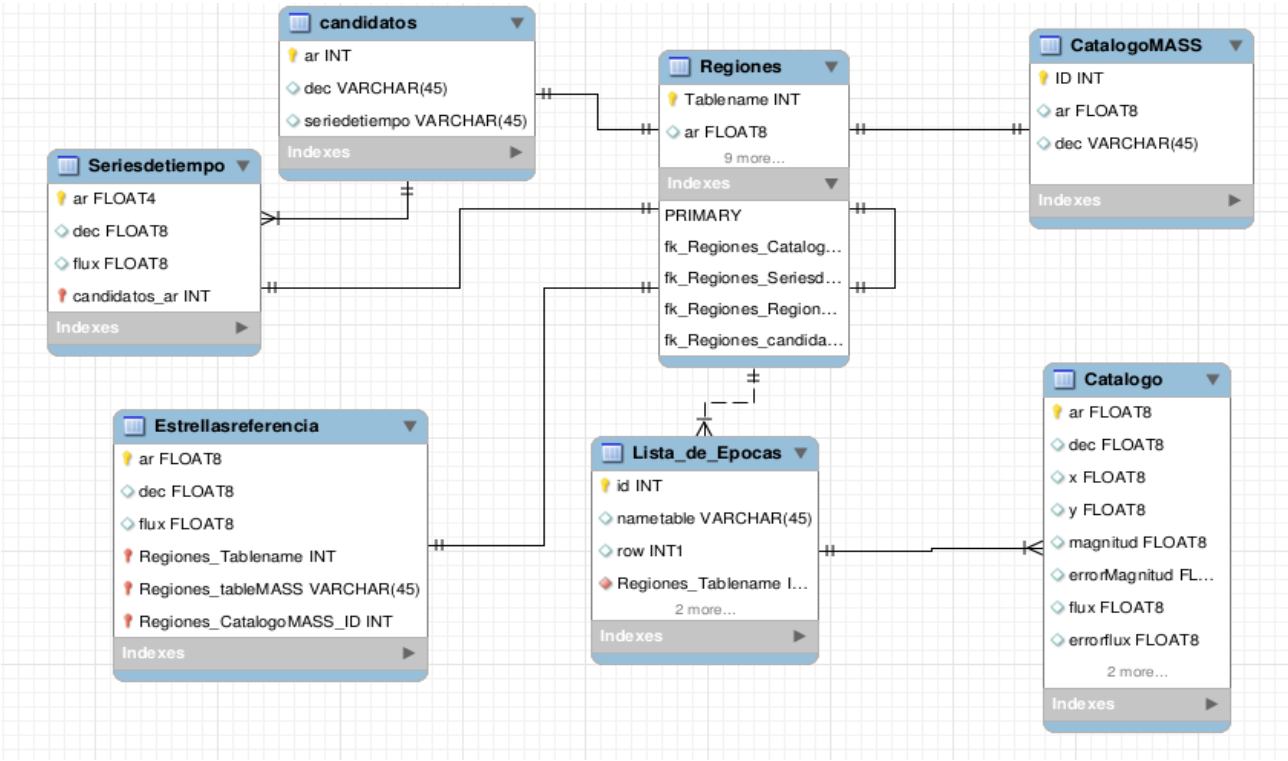


Figura 4.3: Modelo relacional de la base de datos

La base de datos CHASE se organiza de la siguiente manera:

- **Regiones:** Para cada día de observación, el área total observada se subdividió en regiones.
- **Épocas de cada región:** Es un catálogo adquirido de la imagen de una región dada.
- **Catálogos 2MASS:** Contiene objetos conocidos, en su mayoría estrellas, cada región tiene un 2MASS propio.
- **Estrellas de referencia:** Serán los objetos que ayudaran a normalizar el flujo a lo largo de las diferentes épocas.
- **Series de tiempo:** Al procesar la base de datos y mapear cada objeto genera un tupla con las coordenadas de ascensión recta(AR) y declinación(DEC) que es llamada la llave primaria y sirve para identificarlo. Se almacena para cada región una tabla con sus respectivas tuplas(series de tiempo).
- **Candidatos:** Al ejecutar los algoritmos de clasificación se genera una tabla con los objetos que tengan probabilidad de ser supernovas, para cada región.

4.4. Selección de estrellas de referencia

Las estrellas de referencia para cada región tienen el propósito de eliminar el ruido producido por los cambios ambientales en la atmósfera. Esto debido a que el instrumento de medición CCD puede tener errores, así mismo las noches de observación no son completamente nítidas y la nubosidad afecta la cantidad de fotones que se reciben como medida de flujo de cada objecto. Para poder normalizar los catálogos se selecciona un conjunto de N de estrellas de cada región. Estas estrellas se consultan en la base de datos 2MASS[26], se construye su serie de tiempo y se calcula el valor promedio del flujo, esté se toma como el valor verdadero de la estrella. La razón por la que se seleccionan estrellas es que se supone que una estrella debe tener un brillo contante, aunque en la práctica se tienen estrellas variables que por ende no cuentan con un flujo constante. El algoritmo 4.1 muestra como se realiza el proceso de selección de las estrellas de referencia.

Algoritmo 4.1 Selección de estrellas de referencia y cálculo del valor medio

---

Entrada: **R número de estrellas de referencia deseadas, N regiones y M épocas por región, catálogos 2MASS**
Salida: **Conjunto de estrellas de referencia**

- Para cada región nϵ N

  1. Calcular el número de objetos en cada época M
  2. Seleccionar época m donde contenga el **max número de objetos**
  3. Sea s=**2MASS** que corresponde a esa región $\cap$**Época con max número** de objetos, s será el conjunto de estrellas de referencia
  4. Para cada época m en M, Sea F el arreglo que almacena el flujo para cada estrella
     a) Para cada estrella en s
     Si existe el elemento s en m y es diferente de 0
     Suma flujo de la estrella S=S+flujo de las estrella
  5. Seleccionar las R estrellas con menor número de datos diferentes de 0.
  6. **Calcular la media del flujo** para cada estrella de referencia $\overline{x} = \frac{S}{N}$

1.

4.5. Construcción de las series de tiempo

Para la construcción de las series de tiempo se utilizaron los recursos que brinda el manejador de la base de datos, sin embargo algorítmicamente podemos ilustrarlo de la siguiente manera,algoritmo 4.2, para cada región existente en la base de datos :

Algoritmo 4.2 Construcción de series de tiempo

Entrada:$Region_i$
Salida: Series de tiempo de la $Region_i$,una serie por cada objetos

1. Sea U un conjunto vacío que sera el conjunto objetos que pertenecen a la $Region_i$
2. Para cada época $\epsilon Region_i$ seleccionar las coordenadas de todos los objetos **o** que aparecen en la época y hacer la unión con **U**
3. Para cada **o**ϵ**U**, generar la serie de tiempo F
   a) Para cada época, Si existe el objeto **o** en la época,agregar su flujo a F
   b) Almacenar la serie de tiempo F

4.6. Ruido de fondo

Un paso fundamental para poder trabajar con los datos de los catálogos es el compensar, eliminar o minimizar el ruido de fondo. La idea básica es encontrar un k factor de normalización para cada estrella de referencia. En la figura 4.4 se puede observar el clásico ejemplo de clasificación por el algoritmo de vecino más cercano. Usaremos esa idea. En la sección anterior se explico que para cada catálogo se tiene un conjunto de estrellas de referencia y su valor de flujo medio. Ahora para cada posición i en la serie de tiempo es necesario multiplicar el valor de esa posición por un k factor de normalización. En nuestro caso se usó el estrella de referencia más cercana a dicho objeto en la época en que se este normalizando. Sea el círculo verde el objeto a normalizar, los triángulos rojos podrían representar las estrellas de referencia y los cuadros azules los objetos que ya fueron normalizados. Se puede observar que existe una triangulo que es su vecino más cercano, entonces para este triangulo se calcula un factor que después multiplicará al calor del círculo.

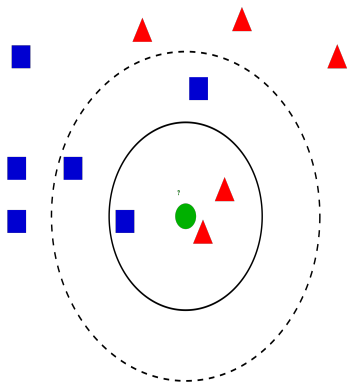


Figura 4.4: Cálculo de normalización de un objeto, con la idea del vecino más cercano.

A éste proceso lo llamaremos normalización, que se hará con base en el flujo de las de las estrellas de referencia, el algoritmo 4.3 muestra paso a paso la formalización de esta idea.

Algoritmo 4.3 Normalización del flujo para cada época para eliminar ruido de fondo

Entrada: Conjunto de estrellas de referencia S, conjunto de series de tiempo
Salida: Conjunto de series de tiempo de objetos normalizadas

1. **Para cada estrella de referencia i**

   a) En cada época de la región

      1) Buscar el flujo de la estrella de referencia i
      2) Calcular el factor de normalización k_i para esa estrella k_i =**valor_actual/valor_medio**
      3) Almacenar el factor k_i y las coordenadas de la estrella a la que pertenece

2. Para **cada curva de luz** de la región

   a) Para cada **posición i** en la serie de tiempo del $objeto_i$

      1) Calcular la estrella de referencia sϵS más cercana al $objeto_i$con distancia euclidiana y consultar su factor de normalización k.
      2) **Multiplicar el valor de flujo actual del objeto por el factor** $k_i$de tal estrella más cercana.

Por supuesto, se podría experimentar con el número de vecinos cercanos, o agrupar las estrellas de referencia por regiones para disminuir el número de factores de normalización y después para cada objeto calcular cual es el cluster más cercano. En nuestro caso experimentamos con 3 casos de factor de normalización:

- Factor k global: es el promedio de los factores k, cada objeto se multiplico por este factor.
- Cuatro factores k: se divide la imagen en cuatro cuadrantes y se calcula un factor k para cada cuadrante, que es el k promedio de las estrellas de referencia que se encuentra en ese cuadrante, luego se multiplica el flujo cada objeto acorde a sus coordenadas por su factor correspondiente.
- N factores de normalización: el algoritmo 4.3 muestra el calculo de estos factores. Este fue el mejor caso, ya que cada objeto se normalizó respecto su estrella más cercana.

4.7. Filtrado

Para seleccionar el filtro a aplicar a las series de tiempo se experimentó con Gauss y DCT. La conclusión a la que se llegó es que el segundo algoritmo representa mejor la señal ya que esta basado en la transformada rápida de Fourier, que se expone en la sección 4.7.1. El algoritmo Gauss demostró ser un buen filtro, ser un buen filtro, sin embargo no conservó los cambios significativos en las señales.

4.7.1. DCT y IDCT

La transformada de coseno discreta (DCT) fue desarrollada por Ahmed[1], y ha sido estudiada y utilizada ampliamente desde entonces. La Transformada Discreta del Coseno (DCT) expresa una señal cualquiera, como la suma de señales sinusoidales con distintas frecuencias y amplitudes. Es una de las transformadas más ampliamente utilizadas en compresión de imágenes, ya que permite expresar la información de la imagen como una combinación de unas pocas frecuencias. La DCT está bastante relacionada con la Transformada Discreta de Fourier (DFT), de hecho es equivalente a la parte real de esta transformada, razón por la cual se compone exclusivamente de funciones coseno.

Además la DCT minimiza algunos de los problemas que surgen con la aplicación de la DFT a series de datos. La transformada discreta de coseno es la más ampliamente utilizada en compresión de imágenes y videos. Esta transformada cuenta con una buena propiedad de compactación de energía, donde los vectores base de la DCT dependen sólo del orden seleccionado de la transformada y no de las propiedades estadísticas de los datos de entrada.

Otro aspecto importante de la DCT es la capacidad de cuantificar los coeficientes utilizando valores de cuantificación que se eligen de forma visual. Esta transformada ha tenido una gran aceptación dentro el tratamiento digital de imágenes, debido al hecho de que los datos de una imagen convencional tienen una alta correlación entre sus elementos.

Podemos definir la DCT unidimensional para N número de datos como:

$$C(u) = \alpha(u) \sum_{x=0}^{N-1} f(x) cos(\frac{(2x+1)u\pi}{2N})$$

Y su transformada inversa como:

$$f(x) = \sum_{u=0}^{N-1} \alpha(u) C(u) cos(\frac{(2x+1)u\pi}{2N})$$

donde

$$\alpha(u) = \begin{cases} \sqrt{\frac{1}{N}} & u = 0 \\ \sqrt{\frac{2}{N}} & u = 1, 2, ..., N-1 \end{cases}$$

Para realizar el filtrado de las series se formuló un filtro en base a DCT/IDCT, donde se filtran las muestras mayores y menores a σ que suele ser ruido. El algoritmo 4.4 muestra el proceso de filtrado, usando una ventana con desplazamientos de una muestra a la derecha, almacenando los valores para cada posición en F que cumplen con las condiciones y calculando el promedio para cada posición de la muestra,. El algoritmo devuelve la serie después de aplicar la transformada inversa.

Algoritmo 4.4 Filtrado de una señal con DCT/IDCT y una ventana de corrimientos

Entrada: serie de tiempo T
Salida: serie de tiempo filtrada

1. Calcular el número de muestras de la serie de tiempo **T** que llamaremos **N**
2. Calcular el tamaño de la ventana, que esta definido por $2^p \approx N/2$ que llamaremos s
3. Calcular la desviación estándar de la serie de tiempo σ
4. Sea **F** el arreglo que almacenará la serie de tiempo filtrada del tamaño de la serie de tiempo
5. Sea **NM** el arreglo que alamacen el numero de muestras alamacenadas para cada posición en **F**, del tamaño de la serie de tiempo
6. Mientras la ventana no llegue al final de la serie
   a) Almacenar el subserie de tamaño **s** en **W** de la serie T
   b) Calcular **DCT** para **W** y almacenar en W_{DCT}
   c) Para cada muestra en W_{DCTi}
      1) Si $W_{DCTi}<\sigma$, almacenar la muestra en $F_i = F_i + W_{DCTi}$y $NM_i = NM_i + 1$
   d) Avanzar una posición de inicio una unidad para **F,T y NM**
7. Para cada posición F_i/NM_i
8. Aplicar **IDCT** a **F** que será la serie filtrada

FIN

4.8. Interpolación lineal

La interpolación lineal es el método más simple en uso hoy en día para la reconstrucción de señales. Es el método usado por los programas de generación de gráficas, donde se interpola con lineas rectas entre una serie de puntos que el usuario quiere graficar.

La idea básica es conectar los 2 puntos dados en x_i , es decir (x_0 ,y_0) y (x_1 , y_1). La función interpola una linea recta entre los dos puntos. Para cualquier punto entre los dos valores de x_0 y x_1 se debe seguir la ecuación de la linea.

$$\frac{y - y_0}{y_1 - y_0} = \frac{x - x_0}{x_1 - x_0}$$

que se puede derivar geométricamente. En lo anterior, el único valor desconocido es y , que representa el valor desconocido para x , despejando queda:

$$y = y_0 + (y_1 - y_0)\frac{x - x_0}{x_1 - x_0} \tag{4.1}$$

donde se asume que $x_0 < x < x_1$, de otra forma esto se conocería como extrapolación.

Si se tienen más de dos puntos para la interpolación, es decir N > 2, con puntos $x_0, x_1, ..., x_n$, se hace la interpolación y generación de la nueva señal con los nuevos puntos calculados.

4.9. Detección de objetos constantes

Para detectar objetos constantes, se usó la desviación estándar que se calculó sobre la serie de tiempo con un umbral de desviación bajo. Este umbral se fijo en 0.05. Todas las serie de tiempo que tengan una desviación estándar típica menor al umbral se etiquetaron como estrellas sin pasar al clasificador.

La siguiente ecuación ilustra el calculó de la desviación estándar típica.

$$\sigma = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2} < 0.05$$

Para toda $x_i \epsilon$ Serie de tiempo a clasificar.

4.10. Clasificación en estrellas, objetos variables y candidatos a supernova

Para realizar el proceso de clasificación ese necesario determinar algunos elementos fundamentales que componen el modelo. Primeramente se debe seleccionar el algoritmo de clasificación, esto dependerá de la naturaleza del problema. En este caso se realiza una clasificación supervisada para poder separar los objetos que son candidatos a supernovas. En las siguientes secciones se define el algoritmo con el que se trabaja y la forma en que se calcula la similitud entre las series de tiempo.

4.10.1. Algoritmo de clasificación: K-NN

Una forma práctica y de fácil aplicación para predecir o clasificar un nuevo dato, basado en observaciones conocidas o pasadas, es la técnica del vecino más cercano. Es una técnica supervisada que se implementa para la clasificación,el algoritmo 4.4 muestra los pasos que sigue este proceso.

Algoritmo 4.5 Algoritmo K-NN

Entrada: Conjunto de instancias,umbral,k vecinos
Salida: clasificación de cada instancia
Entrenamiento:

1. **Almacenar el conjunto de ejemplos** D=$\{(x_1,c_1),...,(x_n,c_n)\}$ para cada clase c

Clasificación:
Sea x_q un nuevo caso a clasificar

1. Para cada clase c

   a) Para todo **objeto** del conjunto D que pertenece a la clase c , calcular la **distancia** $d_i = d(x_i, x_q)$

   b) Ordenar $d_i = (i = 1, ..., n)$ en orden ascendente

   c) Tomar los k casos D_X^K más cercanos a x_q

   d) Calcular el promedio m de distancia de x_q

   e) Si $m < umbral$, asignar clase a x_q

   Fin

4.10.2. Funciones de distancia

Existen pocas formas de medir la similitud entre dos series de tiempo, ya que dependerá en gran medida de la naturaleza del problema que se desee resolver. En nuestro caso se optó por seleccionar de las medidas comunes, el algoritmo DTW y el coeficiente de correlación. DTW es un algoritmo que permite verificar la similitud de dos series punto a punto de manera dinámica a diferencia de la distancia euclidiana. Por otra parte tenemos el coeficiente de correlación, que es famoso por medir el comportamiento de los puntos que integran las series de tiempo, este representa de manera global la similitud de dos series.

4.10.2.1. Dynamic time warping(DTW)

DTW(dynamic time warping) es un algoritmo que mide las similitud entre dos secuencias temporales. A diferencia de la distancia euclidiana, DTW calcula la distancia entre dos series sin importa si estas estan rotadas o desplazadas y determina sus similitud, aun con las variaciones localizadas que aumentan o disminuyen la duración del tramo de análisis. Formalmente podemos definir matemáticamente el algoritmo de la siguiente manera.

Sean las secuencias X=$(x_1,x_2,...,x_N)$, de longitud N, e Y=$(y_1,y_2,...,y_M)$, de longitud M.

Definimos un espacio vectorial F formado por estas muestras, de manera que:

$$x_n, y_m \epsilon F, \forall n \epsilon [1, N], \forall m \epsilon [1, M]$$

Medida de costes o distancias locales:

$$c : F \times F \rightarrow \Re \geq 0$$

Esto produce una matriz de costos

$$C \epsilon \Re^{NXM}$$

Esta matriz nos permite encontrar el camino mínimo.
Un camino de alineación se define como:

$$p = (p_1, p_2, ..., p_L) \text{ con } p_{l=}(n_l, m_l) \epsilon [1, N] x [1, M], \forall l \epsilon [1, L]$$

El camino de alineación se define como una secuencia que debe cumplir 3 condiciones: frontera, monotonía y salto.
El costo de un camino es la suma de todos los costos locales

$$c_p(x, y) = \sum_{l=1}^{L} c(x_{nl}, y_{ml})$$

El costo total mínimo se define como DTW

$$DTW(X, Y) = c_{p*}(X, Y) = min\{c_p(X, Y)|p\}$$

Donde **p** es el camino de alineamiento
Se implementa a través de un algoritmo de programación dinámica.
La problemática asociada hace referencia a la dificultad añadida en el proceso de medida de distancia entre patrones. Normalmente se utiliza la distancia euclidiana como función de distancia, en nuestro caso usaremos DTW porque es una medida más flexible para comparar las series de tiempo, la forma en la que se mide la distancia entre las series de tiempo se ilustra en la figura 4.5.

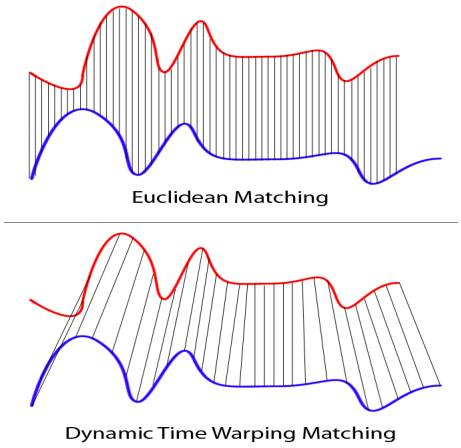


Figura 4.5: DTW vs Distancia euclidiana

Para realizar dicho cálculo de similitud se aplica el siguiente algoritmo de DTW, se divide en dos fases, la primera calcula la matriz de distancia entre todos los puntos. La segunda

traza el "camino" óptimo que es la medida de similitud entre ambas series. El algoritmo 4.5 expone los procesos involucrados en el calculo de DTW.

Algoritmo 4.6 Matriz de costos para dynamic time warping

Entrada:Dadas dos series de tiempo X=($x_1, x_2, .., x_n$), donde nϵN y Y=($y_1, y_2, .., y_n$)
Salida: distancia minima
Construir la matriz de distancias C, a la que llamaremos matriz local de costos.
Sea $c \epsilon C$ el costo en la posición i,j
El algoritmo de calculo de costos con sus respectivos criterios.

1. n$\leftarrow$|X|
2. m$\leftarrow$|Y|
3. dtw$\leftarrow$new[n x m]
4. dtw(0,0)$\leftarrow$0
5. for i=1;i$\leq$n;j++ do
   a) dtw(i,1)$\leftarrow$dtw(i-1,1)+c(i,1)
6. for j=1;j$\leq m; j++$
   a) dtw(1,j)$\leftarrow$dtw(1,j-1)+c(1,j)
7. for i=1;i$\leq n$;j++
   a) for j=1;j$\leq m$;j++
      1) dtw(i,j)$\leftarrow$min{dtw(i-1,j);dtw(i,j-1);dtw(i-1,j-1)}

Para realizar el cálculo de la distancia optima entre las dos series se aplica el algoritmo 4.6, el siguiente algoritmo genera el camino óptimo que representa la distancia entre ambas series.

Algoritmo 4.7 Camino óptimo para DTW

Entrada: Matriz de costos del algoritmo 4.5
Salida: Camino óptimo

1. path←new array
2. i=row(dtw)
3. j=columns(dtw)
4. while(i>1)&(j>1)
   a) if i==1 then
      1) j=j-1
      2) elseif j==1
         i=i-1
      3) else
         if dtw(i-1,j)==min{dtw(i-1);dtw(i,j-1);dtw(i-1,j-1)}
         i=i-1
         else if dtw(i,j-1)==min{dtw(i-1);dtw(i,j-1);dtw(i-1,j-1)}
         j=j-1
         else
         i=i-1;j=j-1
         path.add(i,j)

4.10.2.2. Coeficiente de correlación

El coeficiente de correlación es una medida de asociación entre dos variables y se simboliza con la literal r.

Los valores de la correlación van de + 1 a - 1, pasando por el cero, el cual corresponde a ausencia de correlación. Los primeros dan a entender que existe una correlación directamente proporcional e inversamente proporcional, respectivamente.

El valor del coeficiente correlación varía en el intervalo [-1,1],:

- Si r = 1, existe una correlación positiva perfecta. El coeficiente indica una dependencia total entre las dos variables denominada relación directa: cuando una de ellas aumenta, la otra también lo hace en proporción constante.
- Si $0 < r < 1$, existe una correlación positiva.
- Si r = 0, no existe relación lineal. Pero esto no necesariamente implica que las variables son independientes: pueden existir todavía relaciones no lineales entre las dos variables.
- Si $-1 < r < 0$, existe una correlación negativa.

- Si r = -1, existe una correlación negativa perfecta. El índice indica una dependencia total entre las dos variables llamada relación inversa: cuando una de ellas aumenta, la otra disminuye en proporción constante.

Los coeficientes de correlación significan esa asociación entre los cambios que se observan en la variable dependiente con respecto a la variable independiente.

En la figura 4.6 se muestran los ejemplos típicos de correlación.La gráfica (a) representa una correlación positiva, es decir, conforme los valores de X aumentan, también aumentan los valores de Y. A su vez, la gráfica (b) muestra una correlación negativa, de modo que al incrementarse los valores de la variable independiente, los valores de la dependiente disminuyen. La gráfica (c) no indica correlación.

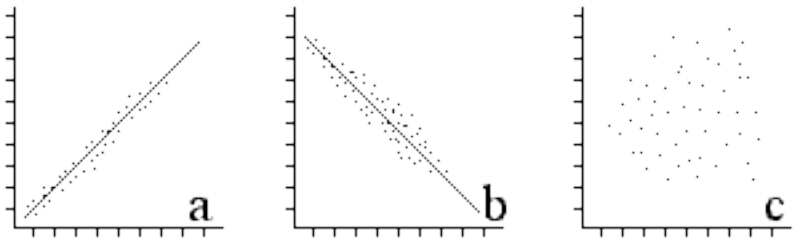


Figura 4.6: Ejemplos típicos de correlación. a) Indica una correlación positiva. b) Indica una correlación negativa. c) No existe una correlación

El coeficiente de correlación lineal de Pearson se define matemáticamente con la ecuación siguiente:

$$r = \frac{N\sum xy - \sum x \sum y}{\sqrt{(N\sum x^2 - (\sum x)^2)(N\sum y^2 - (\sum y)^2)}} \tag{4.2}$$

Donde:
r = coeficiente de correlación de Pearson.
$\sum xy$ = sumatoria de los productos de ambas variables.
$\sum x$ = sumatoria de los valores de la variable independiente.
$\sum y$ = sumatoria de los valores de la variable dependiente.
$\sum x^2$ = sumatoria de los valores al cuadrado de la variable independiente.
$\sum y^2$ = sumatoria de los valores al cuadrado de la variable dependiente.
N = tamaño de la muestra en función de parejas.

4.10.3. Minimización de la distancia

Se minimiza la distancia entre las series de tiempo, esto a través del desplazamiento de series, ya que puede existir un desfase en tiempo. La distancia mínima es la que se almacenará en la matriz de distancias. Un ejemplo de este tipo de problemas se ilustra en la figura 4.7.

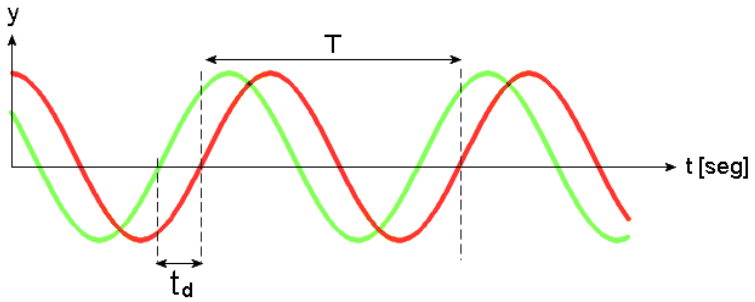


Figura 4.7: Desfase de señales

Para corregir este problema se calcula la distancia entre las series, desplazando la primera y dejando fija la segunda hasta encontrar la distancia mínima entre ellas. Este proceso se ilustra en el algoritmo 4.6.

Algoritmo 4.8 Minimización de la distancia

Entrada: función de distancia f(a,b), Serie 1 y Serie 2
Salida: Distancia mínima

1. Sea min la distancia mínima entre dos series
2. Para la posición n de la serie 1
   a) Calcular f(serie1,serie2)
   b) Si es menor a min almacenar la nueva distacia
   c) desplazar la serie 1
3. Devolver la distancia mínima

4.11. Clasificación en las clases de supernova Ia,Ibc,II y no-supernova

Ya que se han eliminando las estrellas y los objetos variables del conjunto de fuentes es necesario agrupar los candidatos en tres clases de supernova: Ia, Ibc y II. Para ello se cuenta con ejemplos de series de tiempo de 512 supernovas, que se tomaron de la literatura y se usaron como conjunto de entrenamiento.

El problema ya ha sido estudiado con anterioridad,se ha experimentado con una base de datos sintéticos llamada Galaxy Zoo Supernovae, creada por S.Bailey[3], donde se usan tres algoritmos:

- SVM(Support vector machine)
- Random forest
- Boosted tree

Sin embargo en nuestro trabajo implementó el algoritmo de K-NN como primer acercamiento a la solución del problema, ya que en el caso de nuestra base de datos aun se deben revisar las etapas de preprocesamiento.Se probará con las funciones de distancia ya mencionadas, el capítulo 6 expondrá los resultados de la etapa experimental y de la clasificación total de los candidatos.

Para esta clasificación se usaron 512 ejemplos de supernova de la literatura. Estos datos contienen la magnitud aparente de cada estrella que puede convertirse a la cantidad de flujo. Es importante destacar que este conjunto de datos se filtra con los siguientes criterios:

- Cantidad de puntos mayor que 2
- Flujo creciente, que vayan de menor a mayor

Después de filtrar los datos con los criterios anteriores, se asigna la clase a cada ejemplo por un experto en el área.Se seleccionaran los mejores 12 ejemplos de cada clase, para realizar la comparación. Esta selección se hace acorde a los criterios de un experto en astronomía, este conjunto de 36 ejemplos será el conjunto de entrenamiento. El resto es el conjunto de prueba para validar el clasificador.

4.12. Interfaz gráfica del usuario

4.12.1. Interfaz gráfica para PPP

Respecto a la construcción de la interfaz gráfica, se usó la biblioteca Tkinter de Python. Tkinter es una biblioteca gráfica Tcl/Tk. Se eligió este lenguaje de programación por la facilidad de uso en los diferentes sistemas operativos existentes, además que Python es compatible con muchas de las herramientas existentes para la astronomía. Como primera parte se desarrolla la interfaz para PPP.

El diseño de la interfaz está basado en la forma en que se agrupan los comandos en PPP, con un menú parara las siguientes opciones:

- Etiquetado de objectos
- Fotometría
- Edición de scripts de etiquetado y fotometría

Estos módulos trabajan con imágenes de tipo FITS como entrada.

4.12.2. Detector de supernovas

También se agregó un interfaz para los siguientes módulos:

- Construcción de una Bases de datos con los catálogos
- Interfaz procesamiento de series de tiempo(filtrado, normalización, ruido de fondo)
- Interfaz para clasificación de series de tiempo

4.13. Reportes de clasificación

El software genera reportes de clasificación de cada región que se encuentre en la base de datos. La salida será en formato HTML mostrando para cada objeto:

- Coordenadas en AR y DEC
- Clasificación(Estrella, Objeto Variable y Candidato a Supernova)
- En caso de ser supernova su gráfica y su porcentaje de pertenencia.
- También se generará un reporte en formato de texto plano

Para la clasificación de candidatos a supernova, el reporte se hará en formato de texto plano, con las siguientes características:

- Coordenadas en AR y DEC
- Clasificación(Ia, Ibc, II, No-supernova)
- Serie de tiempo

Capítulo 5

Implementación del modelo: desarrollo de software

El software que se desarrolló está basado en el modelo de detección de los proyectos que existen para encontrar objetos variables como supernovas o asteroides. Este software esta compuesto por los siguientes módulos:

- Etiquetado de objetos
- Fotometría
- Integración de base de datos
- Muestreo de estrellas de referencia
- Construcción de series de tiempo
- Limpieza de ruido de fondo con estrellas de referencia
- Tratamiento de series de tiempo(filtrado y normalización)
- Clasificación de series de tiempo

Estos módulos realizan el procesamiento a partir de imágenes hasta la detección de los objetos variables y su respectiva clasificación En las siguientes secciones se comentará el desarrollo en python y las diferentes herramientas utilizadas.

5.1. Etiquetado de objetos

Se desarrolló una interfaz gráfica para realizar el procesamiento de las imágenes FITS, esta herramienta genera un script que es ejecutado por PPP. Como salida se obtiene un archivo que contiene las etiquetas y las coordenadas X,Y del centroide del objeto. A continuación de muestra la interfaz gráfica en la figura 5.1.

Figura 5.1: PPP . Interfaz gráfica de usuario

El proceso que se sigue para llevar a cabo el etiquetado está descrito en el trabajo del Dr. Yee[32][32], creador de PPP. Para dicho proceso se requieren algunos parámetros de entrada que se muestran en la figura 5.2, estos los define el usuario para realizar correctamente el etiquetado de los objetos.

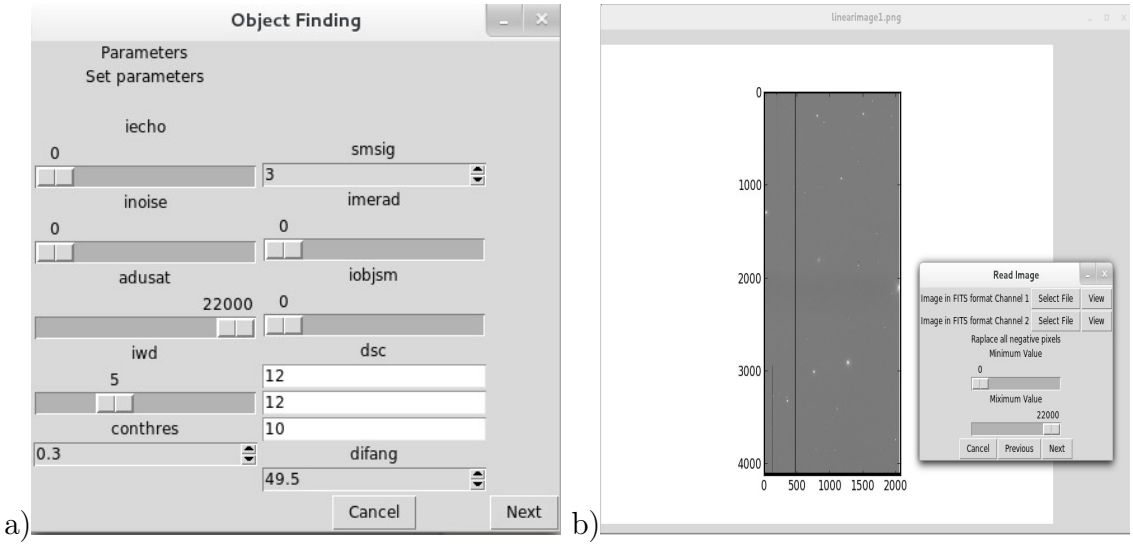


Figura 5.2: a)Parámetros de PPP b)Selección de imagen a etiquetar

Después de realizar los ajustes de parámetros se puede realizar la ejecución del script generado la interfaz. En la figura 5.3, se muestra dicha interfaz.

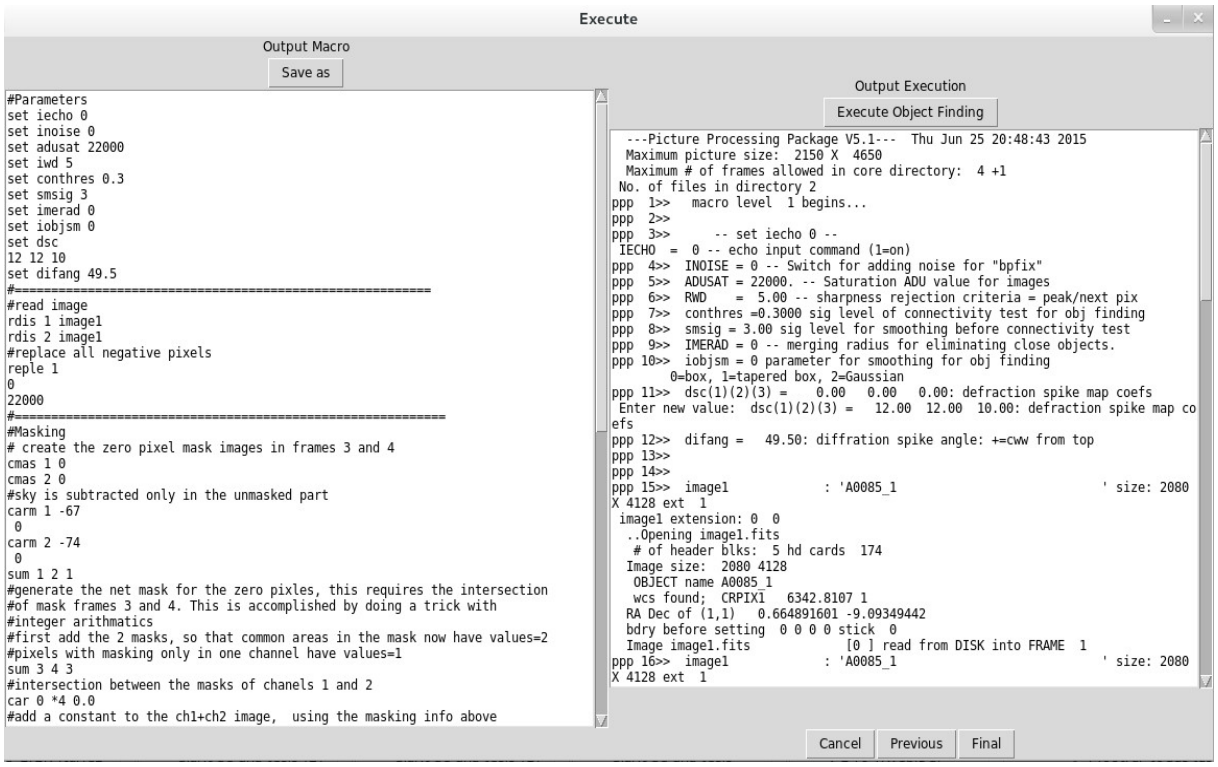


Figura 5.3: Ejecución del etiquetado

La salida de esta ejecución es un catálogo que contiene las coordenadas x,y de cada objeto detectado.

5.2. Fotometría

La fotometría se realiza con base en la salida del etiquetado, tomando como centroide las coordenadas x,y de cada objeto y realizando aperturas circulares, tomando como centro el centroide, intentando hacer el cubrimiento del objeto hasta alcanzar el fondo de la imagen. El número de aperturas se puede ajustar en las opciones de la interfaz. Esto depende de la calidad que se desea la medición del flujo. La interfaz gráfica construida para este propósito se observa en la figura 5.4.

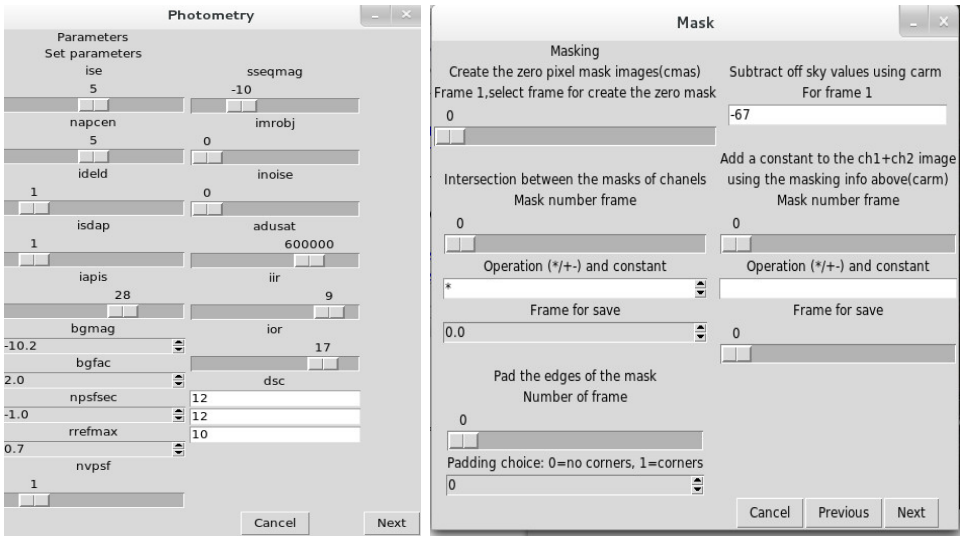


Figura 5.4: Parámetros para fotometría y enmascaramiento de la imagen

Esto parámetros de la figura 5.5 ayudan a acortar la forma en que se realizará la fotometría. Como se muestra en la figura, es necesario seleccionar el archivo de centroides para poder hacer la medición de flujo.

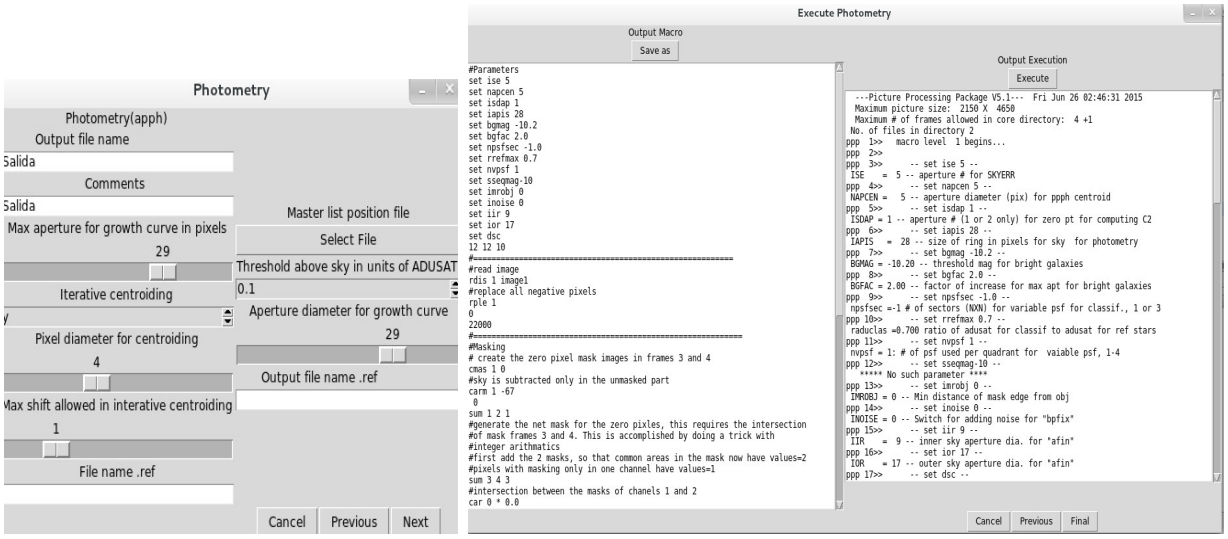


Figura 5.5: Selección de archivo de posiciones y ejecución de fotometría

5.3. Procesamiento de catálogos

Para realizar el procesamiento de los catálogos se implementaron algunas funciones en python, principalmente para procesar archivos. Se usó el manejador de la base de datos MySQL. Cada catálogo fue almacenado acorde a la región a la que pertenece, para cada región se guarda el un catálogo del 2MASS para calcular las estrellas de referencia,así mismo se construyen el conjunto de estrellas de referencia y las series de tiempo. Todos estos procesos

se resumen en una sola interfaz gráfica que se muestra en la figura 5.6. Como se puede observar es necesario indicar la ubicación de los archivos de cada catálogo. También hay que indicar los datos de conexión a la base a la base de datos donde se almacenarán los datos. Es importante indicar el directorio para las series de tiempo, ya que después del proceso de construcción se almacena un archivo final con las series de tiempo correspondiente a cada región.

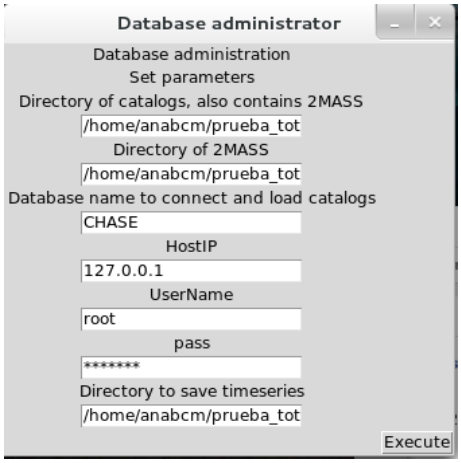


Figura 5.6: Interfaz para el procesamiento de catálogos en la base de datos

5.3.1. Construcción de la base de datos

El programa organiza la base de datos como se modeló en el capítulo 4. El software ejecuta las consultas correspondientes para cada región, la ejecución se hace en orden temporal, ya que de esa manera están organizados los archivos. Para organizar por región, cada catálogo se consultó el encabezado, ya que éste contiene las coordenadas iniciales de ascensión recta y declinación, un ejemplo de la consulta a la base de datos construida con el software se muestra en la figura 5.7.

```
mysql> create database Ejemplo_Tesis;
Query OK, 1 row affected (0.00 sec)

mysql>  SELECT COUNT(*) FROM information_schema.tables WHERE table_schema = 'Ejemplo_Tesis';
+----------+
| COUNT(*) |
+----------+
|      140 |
+----------+
1 row in set (0.01 sec)

mysql>  SELECT COUNT(*) FROM information_schema.tables WHERE table_schema = 'Ejemplo_Tesis' and table_name like "%2MASS%";
+----------+
| COUNT(*) |
+----------+
|        2 |
+----------+
1 row in set (0.00 sec)

mysql>  SELECT COUNT(*) FROM information_schema.tables WHERE table_schema = 'Ejemplo_Tesis' and table_name like "%ar%";
+----------+
| COUNT(*) |
+----------+
|        3 |
+----------+
1 row in set (0.00 sec)
```

Figura 5.7: Salida de consulta de base de datos en MySQL

En la figura 5.7 se muestra el ejemplo de ejecución para tres regiones, el total de tablas contiene las tablas de serie de tiempo y 2MASS. Se crea previamente la base de datos en la que se almacenarán los datos. En el ejemplo se pude ver que la base de datos contiene dos catálogos 2MASS, esto se debe a que un solo catálogo puede abarcar más de una región. La última consulta muestra la cantidad de tablas que se tienen para las estrellas de referencia.

```
mysql>  SELECT table_name  FROM information_schema.tables WHERE table_schema = 'Ejemplo_Tesis';
+---------------------------------+
| table_name                      |
+---------------------------------+
| flux_list_of_tables_1764_284    |
| flux_list_of_tables_954_201     |
| flux_list_of_tables_955_201     |
| id_names_tables                 |
| list_of_tables_1764_284         |
| list_of_tables_1764_284_ref_star|
| list_of_tables_954_201          |
| list_of_tables_954_201_ref_star |
| list_of_tables_955_201          |
| list_of_tables_955_201_ref_star |
| pgc18889201401041csv            |
| pgc18889201401071csv            |
| pgc18889201401101csv            |
| pgc18889201401121csv            |
| pgc18889201401141csv            |
| pgc18889201401181csv            |
| pgc18889201401211csv            |
| pgc18889201401261csv            |
| pgc18889201402091csv            |
| pgc18889201402121csv            |
| pgc18889201402171csv            |
| pgc18889201402211csv            |
| pgc18889201402241csv            |
| pgc18889201404091csv            |
```

Figura 5.8: Listado de tablas en el ejemplo de la base de datos

La figura 5.8 muestra la lista de nombres de la base de datos ejemplo. Se puede observar que existe una tabla que enlista las tablas que pertenecen a cada región, las tablas para las

estrellas de referencia y las épocas de cada región con su fecha respectiva.

5.3.2. Selección de estrellas de referencia

El software realiza automáticamente el cálculo de estrella de referencia, tomando los datos de las tablas de 2MASS y las N épocas, construyendo una tabla temporal con los objetos y calculando las estrellas de referencia con base en la media. Estas estrellas se almacenan en una tabla dentro de la base de datos para posteriormente normalizar las series de tiempo, esto aplicando el algoritmo 4.1. Continuando con el ejemplo en la figura 5.9 se muestra una tabla de estrellas de referencia.

```
mysql> select * from list_of_tables_954_201_ref_star ;
+----------+-----------+-------------+
| ar       | decl      | fluxmean    |
+----------+-----------+-------------+
| 95.34000 | -20.08000 |  3101.82750 |
| 95.34000 | -20.07000 |  1301.61500 |
| 95.35000 | -20.21000 |  7052.92429 |
| 95.35000 | -20.19000 | 23774.52818 |
| 95.35000 | -20.18000 |  7256.82045 |
| 95.35000 | -20.13000 | 56506.65750 |
| 95.36000 | -20.20000 |  6217.75464 |
| 95.36000 | -20.17000 |  3839.71000 |
| 95.37000 | -20.13000 |  2836.61607 |
| 95.37000 | -20.10000 |  4134.62000 |
| 95.38000 | -20.20000 |  1898.65000 |
| 95.38000 | -20.10000 |  2008.26500 |
| 95.41000 | -20.21000 | 22320.22429 |
| 95.41000 | -20.18000 |  2201.44000 |
| 95.41000 | -20.05000 |  3901.72528 |
| 95.43000 | -20.07000 |  1094.85000 |
| 95.44000 | -20.20000 |  5624.85077 |
| 95.44000 | -20.09000 |  3585.18258 |
| 95.44000 | -20.06000 |  3418.44286 |
| 95.45000 | -20.21000 |  6185.82833 |
| 95.45000 | -20.20000 |  4856.36750 |
| 95.45000 | -20.09000 |  7140.89486 |
| 95.45000 | -20.05000 |  3034.35133 |
| 95.46000 | -20.14000 |  1452.92333 |
| 95.46000 | -20.10000 |  2240.61560 |
| 95.47000 | -20.13000 |  2804.93095 |
| 95.49000 | -20.16000 |  2820.72138 |
| 95.49000 | -20.05000 |  1265.68500 |
+----------+-----------+-------------+
28 rows in set (0.00 sec)
```

Figura 5.9: Ejemplo de estrellas de referencia

Un punto importante aquí fue hacer coincidir los sistemas de coordenadas, ya que los catálogos de cada época y los catálogos 2MASS manejaban dos sistemas diferentes.Para solucionar este problema se optó por usar la herramienta **skycoor** del **WTCtools** desarrollado por la Universidad de Harvard, llevando todas las coordenadas al sistemas J2000.

5.3.3. Construcción de las series de tiempo

Al finalizar la selección de las estrellas de referencia se construyó la series de tiempo. Para este proceso se aplicará el algoritmo 4.2. Primeramente generando el conjunto de objetos que aparece en todos los catálogos, para ello se usó las propiedades del **JOIN** para hacer la unión de todos los conjuntos menos la intersección. Se crea una tabla que contiene cada objeto y el flujo en cada época, si no existe se agrega un 0, esto dependerá del número de épocas que existan por cada región.

5.4. Tratamiento de las series de tiempo

Este módulo limpia las series de tiempo aplicando técnicas de análisis de señales. Primero se elimina el ruido de fondo producido por la atmósfera, se interpola con un número de puntos dado y al finalizar se aplica el filtrado con DCT. EN este caso el programa se encarga de realizar todo el procesamiento, el usuario no requiere dar datos de entrada.

La figura 5.10 muestra la interfaz que sirve para realizar el filtrado y clasificación de las series de tiempo. También se pude indicar el directorio del conjunto de datos en caso de tener un archivo con series de tiempo que se deseen clasificar.

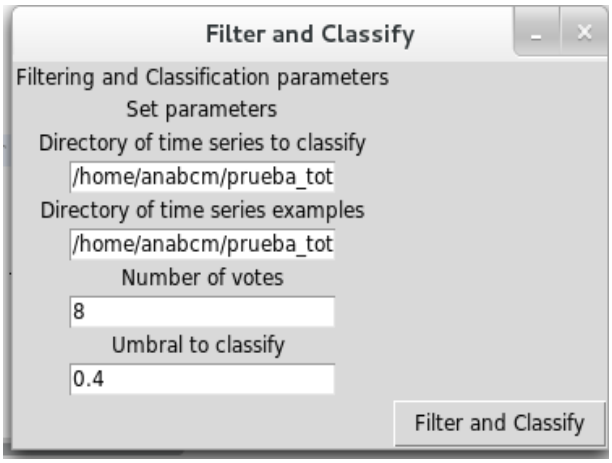


Figura 5.10: Interfaz para el filtrado y clasificación de las regiones existentes en la base de datos.

La figura 5.10 muestra la interfaz para realizar el filtrado y clasificación, los parámetros solicitados son:

- El directorio de las series de tiempo a clasificar
- El directorio donde se encuentran los ejemplos a usar para el conjunto de prueba
- El número de votos a usar para el algoritmos K-NN
- El umbral de clasificación o la distancia mínima de pertenecía a una clase

5.4.1. Ruido de fondo en series de tiempo

El ruido de fondo esta relacionado directamente con las condiciones atmosféricas de observación, que son impredecibles. Para lidiar con este problema se aplica el algoritmo 4.3 al conjunto de series de tiempo que fueron creadas como se comentó en la sección 5.3.3. Para ellos se consultó el conjunto de estrellas de referencia y el conjunto de series de tiempo. Se calculan a continuación los factores de normalización, que dependen de la cantidad de objetos que coincidan con el catálogo 2MASS.

5.4.2. Interpolación

En este caso en particular es necesario determinar el número de puntos que se necesita recuperar de cada serie. Para aplicar el proceso de interpolación se usó que es una extensión de Python, que agrega soporte para trabajar con vectores y matrices, constituyendo una biblioteca de funciones matemáticas de alto nivel para operar con esos vectores o matrices. El ancestro de Numpy, Numeric, fue creado originalmente por Jim Hugunin con algunas contribuciones de otros desarrolladores. En 2005, Travis Oliphant creó NumPy incorporando características de Numarray en NumPy con algunas modificaciones.

De Numpy principalmente se usan las funciones de interpolación lineal y de operación con vectores, como se definió en la sección 4.8.

5.4.3. Filtrado con DCT

En este caso se usó el paquete Scipy que contiene las funciones para calcular la transformada rápida de Fourier, además del calculo de DCT/IDCT.

5.5. Clasificador de objetos(variables, estrellas, candidatos a supernova)

Para realizar la clasificación se usó un criterio de clasificación de la siguiente manera:

1. Se calcula la semejanza a los ejemplos seleccionados
2. Se calcula la desviación estándar, para verificar si se trata de un elemento que tiende a ser constante
3. Si la desviación estándar >0.025 entonces se etiqueta como constante
4. Si no se cumple la condición 3 entonces se calcula la cercanía a las clases, si supera el umbral dado se clasifica como candidato a supernova.
5. En caso de no cumplir con los criterios 3 o 4 se etiqueta como objeto variable.

5.5.1. Conjunto de entrenamiento y prueba

Para realizar el entrenamiento y prueba del clasificador se usaron los ejemplos de supernovas ya existentes. Seleccionando de ellos los 12 ejemplos más significativos por cada clase. En cuanto al conjunto de prueba se seleccionaron 2 regiones del cielo, en donde cada objeto fue clasificado de manera visual para poder aplicar la técnica de ground truth y verificar la clasificación. Las características de los conjuntos acorde al ground truth (datos de campo, datos obtenidos de la realidad) se observan en el cuadro 5.1.

Región	Total de objeto	Candidatos	Constantes	Variables
pgc18889	122	5	40	77
pgc36664	238	21	48	169

Cuadro 5.1: Regiones seleccionadas para el entrenamiento y prueba del clasificador

Cabe resaltar que el número de objetos variables es mayor a los otros conjuntos, ya que muchas de las estrellas son variables y tiene períodos muy erráticos.

5.5.2. Funciones de distancia

Se implemento el coeficiente de correlación de Pearson, aplicando las ecuaciones que se mostraron en la sección 4.10.2. Para hacer el cálculo del DTW se usó la librería Fastdtw elaborada por Stan Salvador y Philip Chan, que es un algoritmo de aproximación al DTW. Se implementan ambas funciones para determinar con cuál se obtienen mejores resultados de clasificación.

La matriz de distancia se calcula minimizando la distancia entre cada serie, es decir se hace un desplazamiento de las series hasta encontrar la distancia mínima entre las series.

5.5.3. Clasificación y K-NN

Se implemento el algoritmo de clasificación K-NN, tomando en cuenta el número de vecinos más cercanos con los que se desea trabajar que se definen en la entrada de la interfaz gráfica, estos serán los vecinos con mayor cercanía a cada objeto de cada clase.

Posteriormente se compara la cercanía que brinda el K-NN con respecto a los criterios expuestos al inicio de esta sección para hacer una clasificación global de cada serie de tiempo. En caso de ser clasificado como supernova pasa a la segunda etapa de clasificación que divide este conjunto en sus subclases.

5.6. Clasificador de supernova(Ia, Ibc, II)

El conjunto que de candidatos a supernova que se obtuvo en el clasificador de la sección 5.5 pasa a una segunda etapa de clasificación en sus correspondientes clases de supernova. Este clasificador es más exigente en cuanto a la semejanza, en caso de no cumplir con ella es descartado como un objeto variable.

5.6.1. Conjunto de entrenamiento y prueba

Para este procedimiento, primero se clasificó toda la base de datos para obtener los candidatos a supernova. Posteriormente, se clasificaron 500 objetos en su correspondiente clase, dejando también en claro que se encontraron objetos que tienden a ser variables. Este conjunto se dividió en dos conjuntos :entrenamiento y prueba. Asignando el 30 % al primero y el resto para las pruebas, con el fin de seleccionar los mejores parámetros de clasificación(Número de vecinos, umbral de clasificación). El conjunto quedo constituido de la siguiente manera(cuadro 5.2).

Conjunto	Ia	Ibc	II	Variable
Entrenamiento	58	23	37	32
Prueba	184	43	86	67

Cuadro 5.2: Distribución de los conjuntos de entrenamiento y prueba para clasificación de candidatos.

En cuanto al clasificador y a las funciones de distancia se usaron las ya implementadas para el clasificador de objetos variables.

5.7. Reportes del sistema

Los reportes de realizan en un formato HTML(figura 5.11), ilustrando la curva de luz de cada objeto. En caso de haber sido clasificada como supernova se agrega la cercanía que tiene respecto a su clase. Esta salida es opcional ya que se requiere tiempo de máquina para la graficación de cada curva.

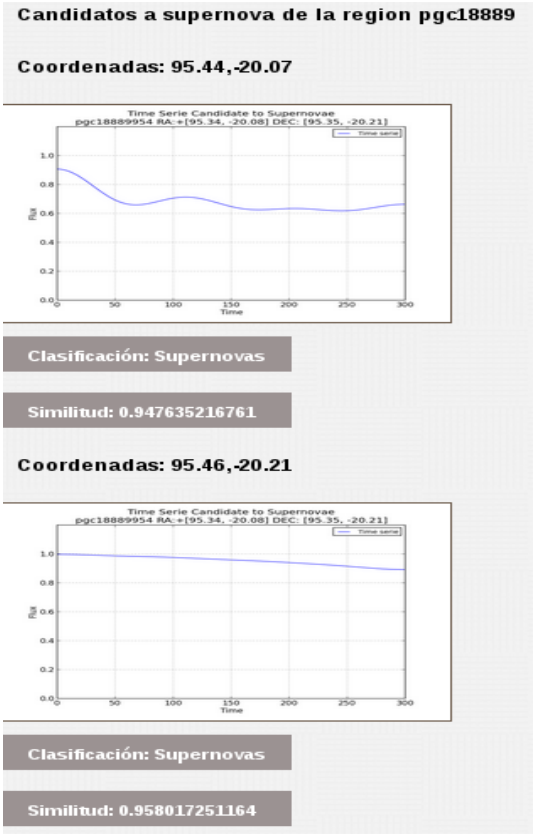


Figura 5.11: Ejemplo HTML del reporte de clasificación

Un reporte adicional es un archivo de texto que contiene las coordenadas del objeto, la clasificación y cercañía a su clase correspondiente. Esta clasificación también se almacena en la base de datos.

Capítulo 6

Resultados experimentales

6.1. Estructura final de la base de datos

Al finalizar el procesamiento se obtuvo una base de datos con 60,782 tablas organizadas en 3,546 regiones celestes, que comprenden más de 2 años de observación.

Éstas ocupan un espacio de 4 GBytes, únicamente los catálogos sin imágenes FITS. Se sabe exactamente cuantas épocas tiene cada región y se puede mapear cada objeto por sus coordenadas en todas las épocas disponibles. La figura 6.1 muestra la estructura general de la base de datos CHASE, creada por el software SDS.

```
mysql>  SELECT COUNT(*) FROM information_schema.tables WHERE table_schema = 'CHASE';
+----------+
| COUNT(*) |
+----------+
|    60782 |
+----------+
1 row in set (0.16 sec)

mysql>  SELECT COUNT(*) FROM information_schema.tables WHERE table_schema = 'CHASE' and table_name like "%ref%";
+----------+
| COUNT(*) |
+----------+
|     3546 |
+----------+
1 row in set (0.12 sec)

mysql>  SELECT COUNT(*) FROM information_schema.tables WHERE table_schema = 'CHASE' and table_name like "%2MASS%"
;
+----------+
| COUNT(*) |
+----------+
|     2583 |
+----------+
1 row in set (0.13 sec)
```

Figura 6.1: Estructura final de la base de datos

El número de tablas de los catálogos 2MASS no coincide con le número de regiones ya que un catálogo 2MASS puede usarse para más de una región.

6.2. Tratamiento de las datos

Al realizar el preprocesamiento de los datos se obtuvieron 272,846 objetos diferentes a los cuales se les construyó su serie de tiempo. Para realizar esta tarea se uso un modelo distribuido de trabajo en cuatro equipos de cómputo en donde se instalo SDS y distribuyo la base de datos CHASE. Cada equipo proceso 887 regiones del cielo. Ésto fue posible ya que al construir la base de datos se almacenan las tablas por regiones que son independientes una de otra, y no requiere hacer algún tipo de reducción. Por lo cual no se requirió una herramienta para cómputo paralelo o distribuido como hadoop.

Se muestra en la figura 6.2, un conjunto de ejemplos de las series de tiempo construidas con los datos de los catálogos. SDS también pude graficar las series de tiempo.

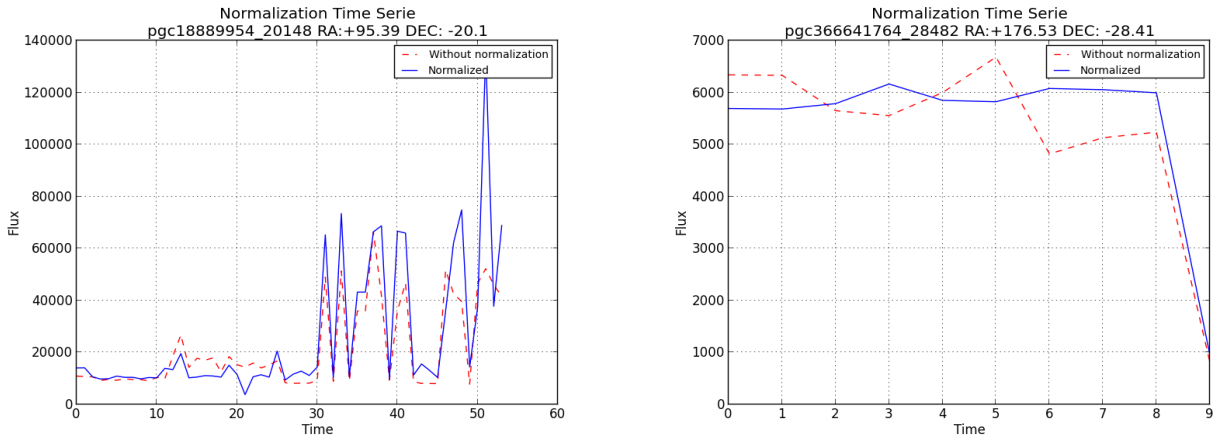


Figura 6.2: Ejemplos de filtro de ruido de fondo sin cambios significativos.

Las gráficas de la figura 6.2, muestran ejemplo de dos regiones, la PGC366641764 y la PGC18889954. La gráfica punteada muestra los datos originales, sin normalizar. La gráfica continua muestra la normalización de la serie de tiempo. En estos ejemplos la normalización no afectó los datos significativamente.

Por otra parte tenemos los ejemplos de las series de tiempo de la figura 6.3, donde se observa que disminuye notoriamente el ruido de fondo.

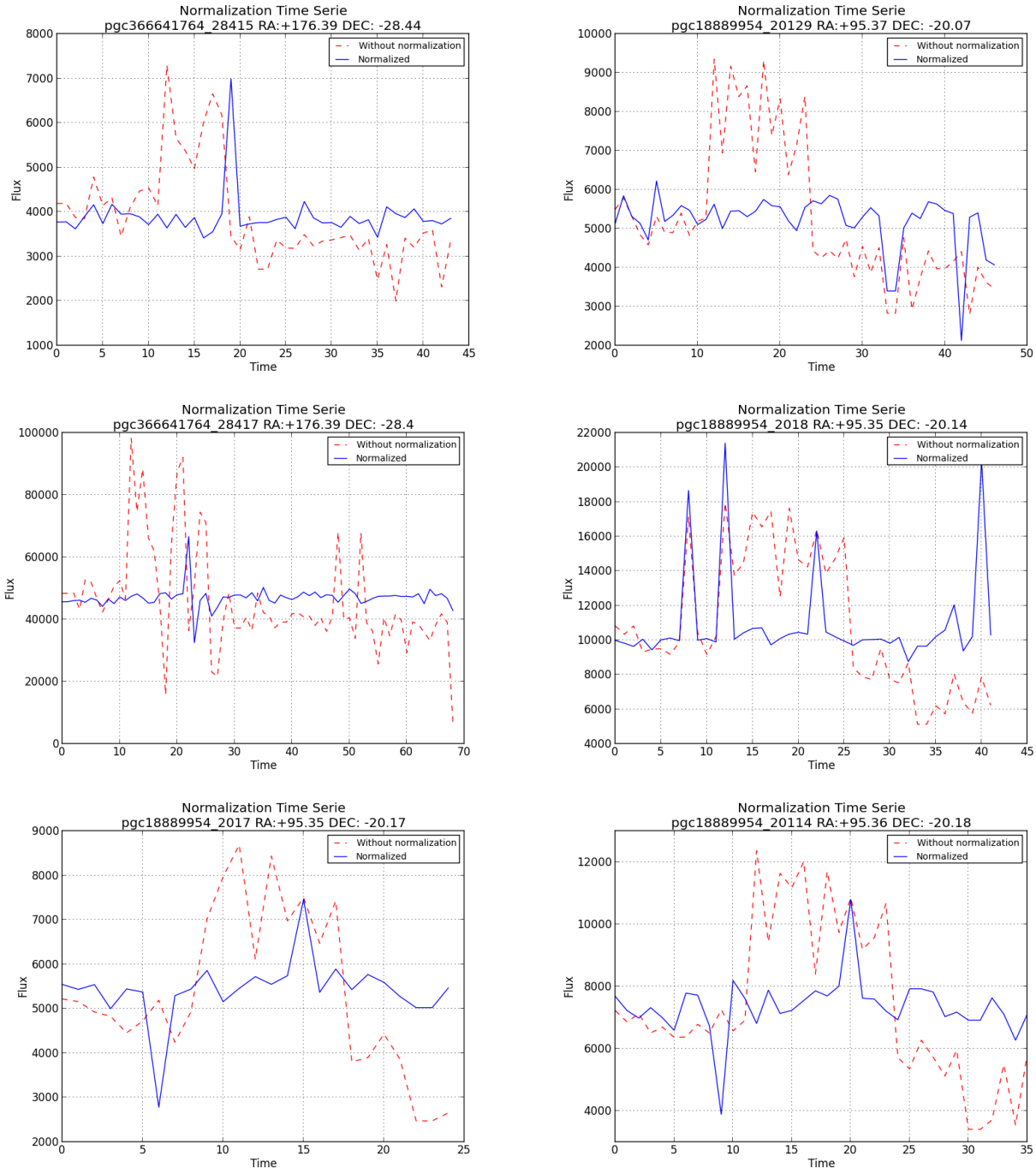


Figura 6.3: Ejemplos de series de tiempo en la base de datos y su corrección con el factor de normalización correspondiente.

Se observa que el número de puntos no es igual en la figura 6.2 y figura 6.3, esto debido a que no en todas las épocas aparece el objeto, por causa del ruido . De ahí la necesidad de normalizar y filtrar el ruido de la señal.

En la figura 6.4 se muestran un par de ejemplo de series de tiempo que al aplicar el filtrado y la interpolación tienden a ser constantes.

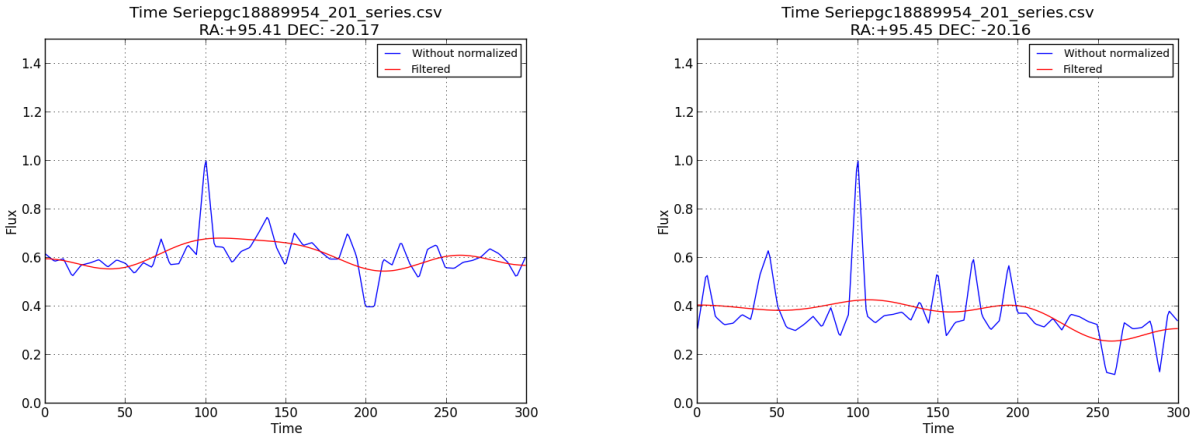


Figura 6.4: Ejemplos de objetos que tienden a ser constantes después de haber sido interpoladas y filtradas con DCT/IDCT

La figura 6.5 muestra dos ejemplos de series de tiempo de posibles candidatos a supernova, después de haber sido filtradas con DCT/IDCT e interpoladas.

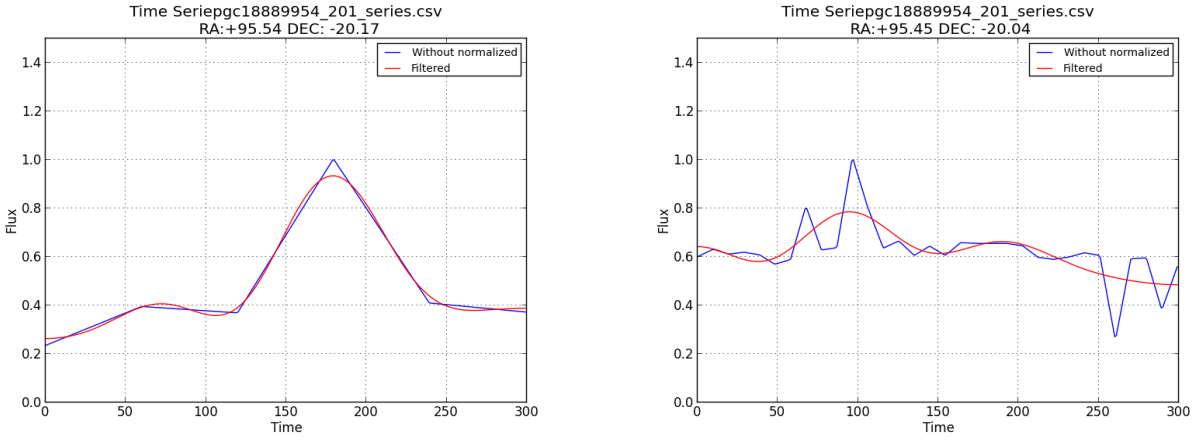


Figura 6.5: Ejemplos de series de tiempo de posibles candidatos a supernova después de ser interpoladas y filtradas con DCT/IDCT

La figura 6.6 muestra series de tiempo de objetos variables después de haber sido filtrados, como ser observa tienen comportamientos muy erráticos, tampoco cuentan con un periodo definido.

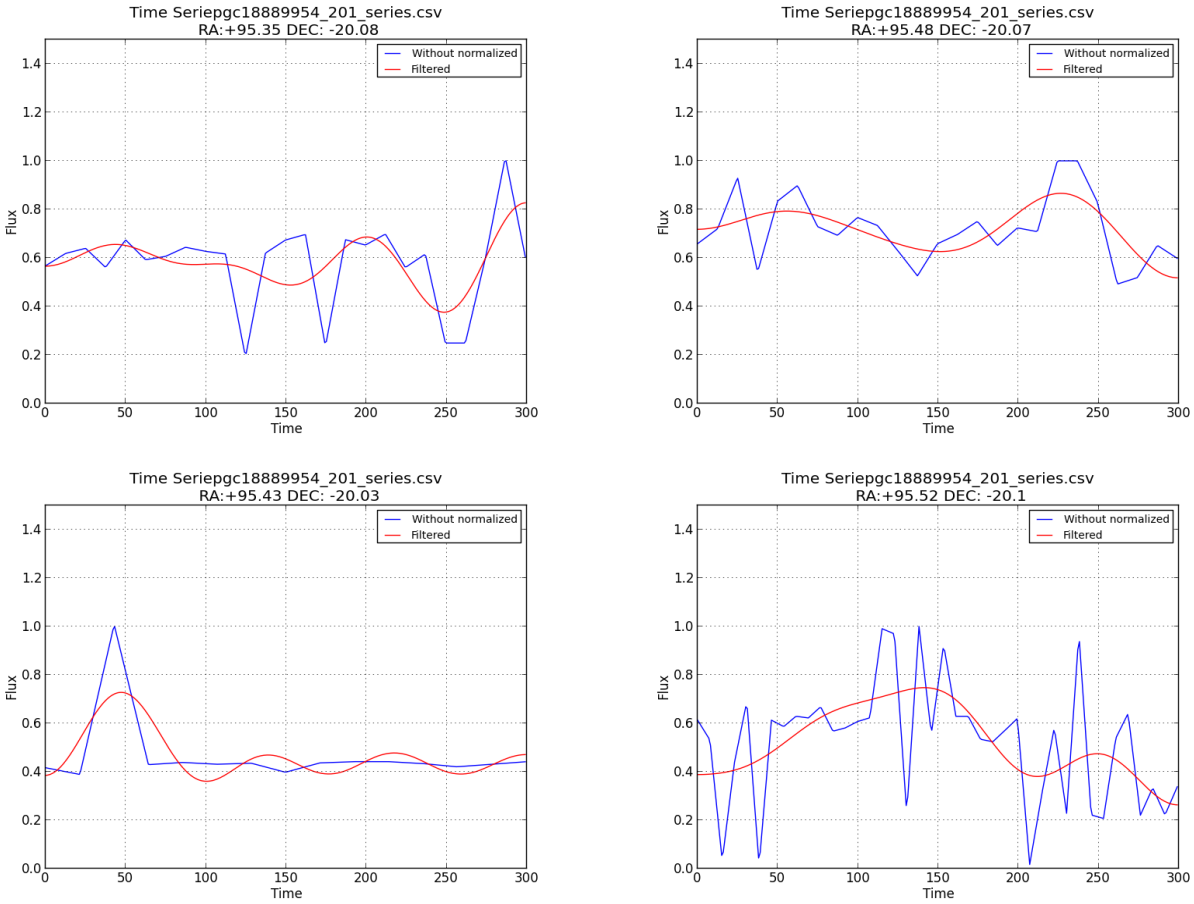


Figura 6.6: Ejemplos de series de tiempo de objetos variables interpoladas y filtradas con DCT/IDCT

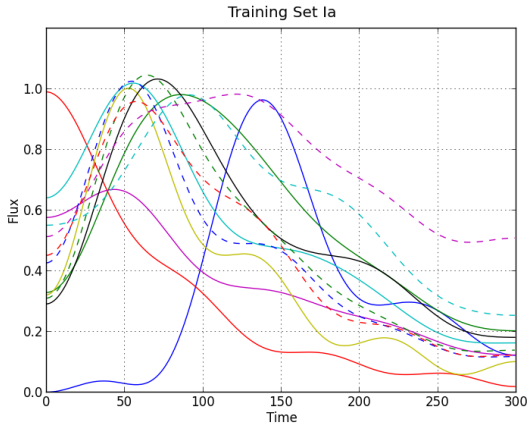
6.3. Selección de función de distancia y parámetros de clasificación

El clasificador K-NN requiere un par de parámetros que indicarán a que clase corresponde cada objeto, estos parámetros son:

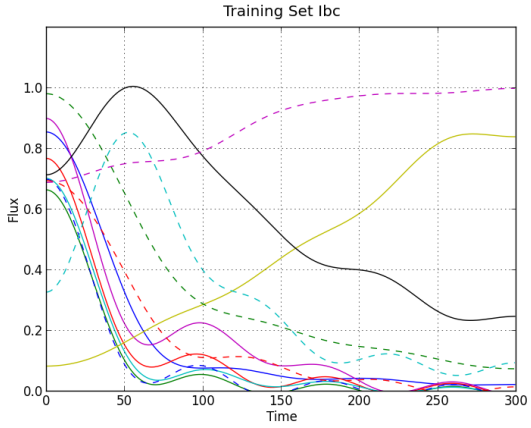
- Umbral de clasificación: Para cada objeto se define a qué distancia promedio debe estar el objeto respecto a la clase para determinar su pertenencia.
- Número de vecinos: debido a la dimensionalidad de los objetos fue necesario determinar con cuántos vecinos es necesario clasificar una serie de tiempo.

Los primeros intentos de clasificación se realizaban con todos los ejemplos, sin embargo el tiempo de cálculo era considerable. Para realizar el calculo se requiere 74s con coeficiente de correlación y 8min28s con DTW para una sola región del cielo con 222 objetos. De ahí que se decidió reducir el conjunto de ejemplos para optimizar la clasificación. La reducción la realizó un experto, quien seleccionó los ejemplos más representativos de cada clase.

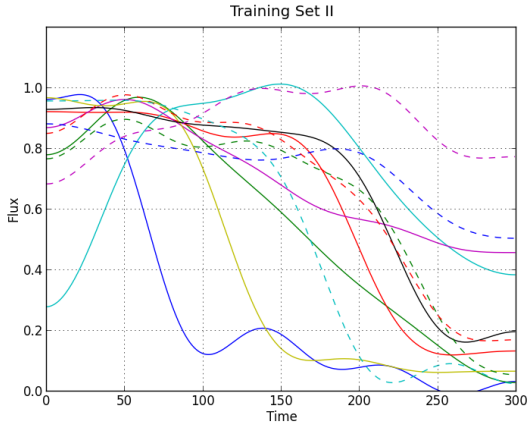
Para seleccionar los parámetros se experimentó con el clasificador, los ejemplos seleccionados y el conjunto de prueba para determinar que parámetros daban un mejor resultado final de clasificación para las tres clases. A continuación se muestran las gráficas de los ejemplos usados en la clasificación, figura 6.7.



Tipos Ia



Tipos Ibc



Tipo II

Figura 6.7: Gráficas del conjunto de entrenamiento de cada tipo de supernova

Como ya se mencionó en el capítulo anterior el conjunto de prueba y validación estuvo constituido de la siguiente manera, tabla 6.1.

Región	**Total de objeto**	**Candidatos**	**Constantes**	**Variables**
pgc18889	122	5	40	77
pgc36664	238	21	48	169

Cuadro 6.1: Regiones seleccionadas para el entrenamiento y prueba del clasificador

Se realizó las pruebas de clasificación con los conjuntos pgc18889 y pgc36664 y ambas funciones de distancia. Los resultados se encuentran en el apéndice A y muestran los resultados de clasificación con las dos funciones de distancia, variando el umbral y el número de vecinos. De los resultados podemos observar que el clasificador recupera muy bien las candidatos a supernova, debido a su flexibilidad al comparar una serie de tiempo con el conjunto de ejemplo. En la siguiente tabla 6.2, se muestra que cuando hay una recuperación de candidatos del 100 %, disminuye la tasa de clasificación, debido a que muchos de los objetos variables son clasificados como supernovas.

Región	**% DTW**	**% Coeficiente**
pgc18889	23.52	44.26
pgc36664	33.6	41.17

Cuadro 6.2: Candidatos a Supernova recuperados con coeficiente de correlación y DTW y el 100 % de candidatos recuperados

Sin embargo, por esta flexibilidad que ofrece el porcentaje de clasificación total de las clases estrella y objeto variable se ven afectadas, en general no se recuperan bien, es por ello que el porcentaje se mantiene constante con DTW. Se observa en la tabla 6.3, que el coeficiente de correlación tiene un comportamiento más diverso, con un umbral alto de clasificación recupera bien los candidatos a supernova.

Región	**% DTW**	**%SN**	**% Coeficiente**	**%SN**
pgc18889	23.52	100 %	83.97	60 %
pgc36664	33.60	100 %	86.88	35 %

Cuadro 6.3: Máximo porcentaje de clasificación correcta de las 3 clases: estrella,variable y candidato a supernova: porcentaje de recuperación de supernovas.

Las medidas de clasificación para el clasificador con coeficiente de correlación para la región pgc18889 se muestran en la siguiente tabla 6.4. Se puede ver que 196 ejemplos de 238 fueron bien clasificados(true positive) y la mayoría son objetos variables.

Clase	TP	FP	FN	TN
Supernova	9	19	11	199
Constante	36	1	12	189
Variable	154	19	16	49

Cuadro 6.4: Medidas para el clasificador de la región pgc18889 para separar las supernovas, objetos constantes y objetos variables.

Para los ejemplos de la región pgc36664, se obtuvieron las medidas del clasificador, con coeficiente de correlación como función de distancia, los resultados se muestran en la tabla 6.5. De 122 ejemplos, se clasifican correctamente 104 en sus respectivas clases, además de detectar correctamente tres candidatos a supernova.

Clase	TP	FP	FN	TN
Supernova	3	6	2	111
Constante	36	7	4	75
Variable	65	5	12	40

Cuadro 6.5: Medidas para el clasificador con coeficiente de correlación como medida de distancia para la región pgc36664

Si se observa en el cuadro 6.3, se verá que no se logró obtener un par de parámetros que en combinación lograran el máximo porcentaje de clasificación y recuperaran todas los candidatos a supernova. DTW por su flexibilidad no ayuda a clasificar correctamente los conjuntos. Por estas razones se decidió elegir un porcentaje menor de clasificación global de las tres clases para poder recuperar correctamente los candidatos a supernova. Si observamos nuevamente las tablas del apéndice A podemos observar que existen parámetros de coeficiente de correlación que dan un porcentaje arriba del 65 % y que recuperan un porcentaje arriba del 60 % de candidatos a supernova. El umbral debe ser mayor al 0.7 de similitud y un número de vecinos mayor a 3.

6.4. Clasificación de la base de datos CHASE en supernovas, objetos variables y objetos constantes

Para terminar de probar el algoritmo y los parámetros de clasificación se eligió como función de distancia el coeficiente de correlación, un umbral del similitud alto, del 0.8 y se usaron 8 ejemplos para determinar la pertenencia a las clases.

Se procesó la base de datos CHASE, clasificando las 3546 regiones de las que está constituida. Los resultados de clasificación se muestran en la siguiente tabla(cuadro 6.6).

Estrellas	Objetos Variables	Candidatos a supernovas
117,784	130,556	24,546

Cuadro 6.6: Clasificación de la base de datos CHASE

La base de dato contiene un poco más de 272,846 objetos, es decir se está clasificando el 12.26 % como candidato a supernova. Este conjunto pasará a la siguiente etapa de clasificación en cuatro clases:Ia,Ibc,II y no-supernova. Este clasificador es más exigente con la similitud respecto a las clases, ya que se usa como filtro de objetos que no son supernovas.

6.5. Clasificación de los candidatos en en las clases I y II

S. Baley y C.Aragon [4], desarrollaron un proyecto de clasificación de supernovas, usando las diferentes bandas e implementaron tres algoritmos de clasificación: máquinas de soporte vectorial, random forest y boosted tree; obteniendo como mejor resultado un 78 % con máquinas de soporte vectorial. Su ventaja fue el uso de más información proporcionada por las tres bandas que en nuestro caso no contamos ya que solo trabajamos con el total de la energía(brillo total emitido).

La base de datos CHASE contiene 272,846 objetos a los que se les construyó una serie de tiempo. Estos datos fueron clasificados como se mostró en la sección 6.4.Para realizar la subclasificación de los candidatos a supernova se desarrolló el clasificador para los diferentes tipos de supernova. El clasificador fue un K-NN con la función de similitud de coeficiente de correlación. Se usaron 500 muestras de los candidatos clasificadas por un experto, éste conjunto se dividió en dos: 40 % para seleccionar los parámetros de clasificación y 60 % para probar el clasificador.

Los resultados para los parámetros con 40 % de la muestras(200) se muestran en la tabla 6.7.

Umbral	**Vecinos**	**%**
0.85	2	72 %

Cuadro 6.7: Parámetros de entrenamiento para la clasificación de los diferentes tipos de supernova

Las medidas para este clasificador para el conjunto de 200 ejemplos se muestra en la tabla 6.8, se observan los verdaderos positivos con un total de 144 aciertos para ambas clases. Otro punto a notar es que no se encontraron objetos variables en el conjunto a pesar del umbral alto, denota entonces que el primer clasificador de objetos variables separa bien los candidatos a supernova.

Clase	**TP**	**FP**	**FN**	**TN**
I	44	13	41	101
II	100	42	8	49
Variables	0	0	6	193

Cuadro 6.8: Medias del clasificador en la etapa de entrenamiento para las diferentes clases de supernovas.

Se observó que se requieren pocos vecinos y un umbral alto de clasificación ya que se cuenta con una amplia diversidad de ejemplos para cada clase. Con los parámetros seleccionados se consiguió una clasificación correcta del 72 %.

Posteriormente se probaron los parámetros de clasificación seleccionados con el 60 % restante de los datos(300), que dieron como resultado un 63 % de clasificación correcta.

Las medidas de clasificación para el conjunto de ejemplos de validación se muestran en la tabla 6.9. De 300 ejemplo, se clasificaron 178 correctamente en sus respectivas clases.

Clase	**TP**	**FP**	**FN**	**TN**
I	65	46	63	125
II	123	64	39	73
Variables	0	1	9	289

Cuadro 6.9: Medidas de clasificación para el clasificador de clases de supernova

Con los parámetros seleccionados y probados se clasificó el conjunto de candidatos (24,000 aprox) en las dos clases generales (I y II), descartando objectos que no alcanzaron el umbral de similitud. Los resultados de esta clasificación se muestran en la tabla 6.10.

Tipo	**I**	**II**	**Variables**
Número de objetos	13,569	10,389	57

Cuadro 6.10: Resultados de clasificación de los candidatos

6.6. Comparación con otros proyectos

Existen precedentes de proyectos que realizaron búsquedas de objetos variables de interés particular, algunos de ellos terminaron su periodo de trabajo, tabla 6.11. Esto trabajo se revisaron en la capitulo de estado del arte. En el caso de nuestro trabajo podemos destacar dos proyectos se con sus propios modelos han trabajado en áreas similares.

Trabajo	**Métodos y técnicas**	**Análisis de cueva de luz**	**Gestor de Base de datos**	**Machine learning**	**Resultados**
Modelo propuesto	**Análisis de señales Quita ruido de fondo con estrellas de referencia. Clasificación supervisada para detección de candidatos.**	**Si**	**Si**	**Si**	**20,000 aprox candidatos que son el 10 % del total de la base de datos.**
CRTS	Implementan una BN (Bayes Network), con varias clases, parámetros y capas.	Si	Si	Si	Clasificación correcta de objetos variables de hasta un 93 %
SuperMacho	Se usó el método llamado análisis de imágenes por diferencia (DIA), de Phillips & Davis	No	Si	No	Detección de objetos variables

Cuadro 6.11: Comparación de resultados del SDS respecto a otros proyectos

Los proyecto que se han desarrollado para la detección de objetos variables y supernovas trabajan directamente sobre las imágenes, realizando registro para hacer coincidir los puntos y realizando restas de imágenes para poder determinar si existe algún cambio de brillo. El proyecto SuperMacho[27] fue pionero en el desarrollo de software complejo para la detección de objetos variables y de los primeros también en usar una base de datos como apoyo para manejar sus datos, aunque no realizan un análisis de la curva de luz, pudieron hacer detecciones en tiempo real. También aplicaron técnicas de minería de datos como clasificación de los diferentes objetos, lo que es adelantado a su época.

Por otra parte tenemos el proyecto más reciente y versátil, el CRTS[7], un proyecto aun vigente del Caltech que esta haciendo exploración en tiempo real, a diferencia del proyecto MACHO, CRTS analiza la curva de luz completa como se puede observar en la página del proyecto se publican alertas de objetos de interés que pueden ser candidatos a supernovas. El grupo de trabajo del proyecto Catalina a implementado diversas técnicas de clasificación[8]. En este caso tienen un porcentaje de clasificación del 93 % contra un 83 % de clasificación de objetos. Sin embargo nuestra herramienta es mucho maś completa en cuanto a las funciones de preprocesamiento de datos.

Capítulo 7

Conclusiones, contribuciones y trabajo a futuro

Este capítulo expone las conclusiones de la tesis para la clasificación de objetos astronómicos, las contribuciones y el trabajo que se propone se puede llegar a realizar en un futuro.

7.1. Conclusiones

La detección de objetos variables, específicamente supernovas ha sido un trabajo que se ha venido desarrollando durante las últimas décadas, ya que la astronomía se interesa por entender nuestro universo. Paralelamente, la cantidad de datos que se obtienen en las observaciones ha crecido exponencialmente, y la detección de objetos de interés se ha vuelto una tarea titánica, que requiere el procesamiento computacional y la aplicación de técnicas de minería de datos.

Este problema se ha atacado en el pasado, trabajando con análisis de imágenes y obteniendo los candidatos al encontrar diferencias significativas, en este caso, pocas veces se analizan los objetos de manera temporal. También en astronomía se trabaja con curvas de luz a través de análisis de Fourier y wavelets. Nuestro trabajo muestra un modelo en el que es posible la aplicación de estas técnicas, ya que ofrecen la ventaja de que dado un conjunto de imágenes de la observación de una región determinada en un periodo de tiempo definido, es posible detectar objetos con una alta similitud a supernovas.

Los resultados obtenidos empleando algoritmos y técnicas de análisis de señales mostraron ser útiles para tratar las series de tiempo y por ende para el trabajo con curvas de luz que son útiles para hacer análisis de objetos en la astronomía.

Los manejadores de bases de datos jugaron un papel fundamental al tratar cientos de miles de registros, gracias a que sus procesos están optimizados para dichas tareas. Esto puede ofrecer algunas ventajas a los astrónomos, desde la consulta de la serie de tiempo de uno o varios objetos, dadas su coordenadas, hasta la administración de la observación de varios años.

Por otra parte el trabajo en astronomía con grandes volúmenes de datos es la nueva forma de hacer ciencia en este campo, grandes proyectos se estan gestando que requieren técnicas computacionales que ayuden a procesar y extraer información útil sobre los datos. Esto se

deberá hacer en tiempo real debido a la importancia y al corto tiempo de duración de algunos de los fenómenos, como las supernovas. Un ejemplo claro de la necesidad del desarrollo de este tipo de herramientas es el telescopio LSST(Large Synoptic Survey Telescope), que estará en marcha en los próximos años, y cubrirá aproximadamente 37 billones de estrellas y galaxias con 10 año de observación, esta observación producirá 15 Terabytes de datos por noche que requerirán ser procesados. De ahí la importancia del desarrollo y aplicación de nuevas técnicas de minería de datos para la astronomía.

7.2. Contribuciones

Este trabajo contribuye de la siguiente manera, con base en los objetivos planteados en el capítulo uno:

- Un modelo de detección de objetos variables, implementado y probado con datos reales de la base de datos CHASE.
- Un software que integra el modelo de extracción, limpieza y procesamiento de los dato. Y permite trabajar con cualquier conjunto de catálogos que contengan el flujo o la magnitud de los objetos observados, así como imágenes en formato FITS.
- Una interfaz gráfica para PPP, programa que cuenta con funciones de etiquetado y fotometría confiables. Esta interfaz es amigable con el usuario y permite procesar un conjunto de imágenes, ya que generá un script con los parámetros seleccionados.
- Una aplicación para la gestión y organización de catálogos en una base de datos, que permite aprovechar las ventajas del manejador de bases de datos.
- Un filtro de ruido de fondo con estrellas de referencia, útil para normalizar catálogos y series de tiempo.
- Curvas de luz de cada objeto, independientemente del proceso de clasificación.Que pueden ser visualizadas para un análisis propio.
- Series de tiempo sin el ruido intrínseco y su reconstrucción por medio de interpolación lineal, esto puede ser útil para el proceso de comparación.
- Una clasificación de los objetos en tres clases que son de interés para el astrónomo, ya que se identifican con un buen porcentaje de precisión, las estrellas, objetos con variabilidad y candidatos a supernova.
- Un clasificador optimizado para la separación en clases de supernova, que también se puede usar externamente.

- Conjunto de reportes de clasificación de la base de datos CHASE.

7.3. Trabajo futuro

Algunas de las lineas de investigación que pueden continuar el presente trabajo son:

- Experimentar con nuevos algoritmos de clasificación de series de tiempo.
- Construir un modelo para cada tipo de supernova, esto podría optimizar el tiempo de clasificación de cada objeto.
- Clasificar objetos variables en sus diferentes tipos, ya que el presente trabajo solo abarca los candidatos a supernovas.
- Experimentar con nuevas bases de datos, ya que este modelo y software lo permiten.
- Utilizar nuevos criterios de validación de la clasificación, que podrían se proporcionados por expertos en astronomía.
- Integrar técnicas para el tratamiento de la señales diferentes a la herramienta, como análisis de Fourier y Wavelets, para el uso de detección de objetos variables y supernovas a través de un análisis espectral.

Bibliografía

[1] N. Ahmed, T. Natarajan, and K. R. Rao. Discrete cosine transform. *IEEE Trans. Computers*, C-23(1):90–93, 1974.

[2] C. Alard. Image subtraction using a space-varying kernel. 144:363–370, jun 2000.

[3] S. Bailey, C. Aragon, R. Romano, R. C. Thomas, B. A. Weaver, and D. Wong. How to Find More Supernovae with Less Work: Object Classification Techniques for Difference Imaging. *apj*, pages 1246–1253, August 2007.

[4] S. Bailey, C. Aragon, R. Romano, R. C. Thomas, B. A. Weaver, and D. Wong. How to Find More Supernovae with Less Work: Object Classification Techniques for Difference Imaging. *apj*, 665:1246–1253, August 2007.

[5] Amanullah Barbary, Aldering. The Hubble Space Telescope Cluster Supernova Survey. II. The Type Ia Supernova Rate in High-redshift Galaxy Clusters. *apj*, 745:32, January 2012.

[6] M. Bolden and P. Kervin. Panoramic-Survey Telescope And Rapid Response System: Leveraging Astronomical Technology for Satellite Situational Awareness. In *38th COSPAR Scientific Assembly*, volume 38 of *COSPAR Meeting*, page 3, 2010.

[7] S. G. Djorgovski, A. J. Drake, A. A. Mahabal, M. J. Graham, C. Donalek, R. Williams, E. C. Beshore, S. M. Larson, J. Prieto, M. Catelan, E. Christensen, and R. H. McNaught. The Catalina Real-Time Transient Survey (CRTS). *ArXiv e-prints*, February 2011.

[8] S. G. Djorgovski, A. A. Mahabal, C. Donalek, M. J. Graham, A. J. Drake, B. Moghaddam, and M. Turmon. Flashes in a Star Stream: Automated Classification of Astronomical Transient Events. *ArXiv e-prints*, September 2012.

[9] C. Donalek, A. Arun Kumar, S. G. Djorgovski, A. A. Mahabal, M. J. Graham, T. J. Fuchs, M. J. Turmon, N. Sajeeth Philip, M. Ting-Chang Yang, and G. Longo. Feature Selection Strategies for Classifying High Dimensional Astronomical Data Sets. *ArXiv e-prints*, October 2013.

[10] Richard O. Duda, Peter E. Hart, and David G. Stork. *Pattern Classification (2Nd Edition)*. Wiley-Interscience, 2000.

[11] A. Gal-Yam, D. Maoz, P. Guhathakurta, and A. Filippenko. Supernovae in Low-Redshift Galaxy Clusters: Observations by the Wise Observatory Optical Transient Search (WOOTS). *apj*, 680:550–567, June 2008.

[12] A. Gal-Yam, D. Maoz, K. Sharon, F. Prada, P. Guhathakurta, et al. Supernovae in galaxy clusters. 2003.

[13] L. M. German. Results of the Mount Stromlo Abell cluster supernova search . *AA*, 415(3):863–878, 2004.

[14] Uriel Giveon. *M.Sc. Thesis. Long-term optical variability properties of the Palomar-Green quasars*. Tel Aviv University, 2000.

[15] E. Hubble and M. L. Humason. The Velocity-Distance Relation among Extra-Galactic Nebulae. 74:43, ju 193.

[16] J. Loveday. The Sloan Digital Sky Survey. *Contemporary Physics*, 43:437–449, June 2002.

[17] LSST Dark Energy Science Collaboration. Large Synoptic Survey Telescope: Dark Energy Science Collaboration. *ArXiv e-prints*, November 2012.

[18] Basa S. Mazure A. *Exploding Superstars Understanding Supernovae and Gamma-Ray Burts*. Springer Berlin Heidelberg New York, 2007.

[19] G. Miknaitis, G. Pignata, A. Rest, W. M. Wood-Vasey, S. Blondin, P. Challis, R. C. Smith, C. W. Stubbs, N. B. Suntzeff, R. J. Foley, T. Matheson, J. L. Tonry, C. Aguilera, J. W. Blackman, A. C. Becker, A. Clocchiatti, R. Covarrubias, T. M. Davis, A. V. Filippenko, A. Garg, P. M. Garnavich, M. Hicken, S. Jha, K. Krisciunas, R. P. Kirshner, B. Leibundgut, W. Li, A. Miceli, G. Narayan, J. L. Prieto, A. G. Riess, M. E. Salvo, B. P. Schmidt, J. Sollerman, J. Spyromilio, and A. Zenteno. The ESSENCE Supernova Survey: Survey Optimization, Observations, and Supernova Photometry. *The Astrophysical Journal*, 666(2):674, 2007.

[20] Smithsonian Astrophysical Observatory NASA. Chandra X-ray Observatory, December 2014.

[21] S. Perlmutter, G. Aldering, G. Goldhaber, R. A. Knop, P. Nugent, P. G. Castro, S. Deustua, S. Fabbro, A. Goobar, D. E. Groom, I. M. Hook, A. G. Kim, M. Y. Kim, J. C. Lee, N. J. Nunes, R. Pain, C. R. Pennypacker, R. Quimby, C. Lidman, R. S. Ellis, M. Irwin, R. G. McMahon, P. Ruiz-Lapuente, N. Walton, B. Schaefer, B. J. Boyle, A. V. Filippenko, T. Matheson, A. S. Fruchter, N. Panagia, H. J. M. Newberg, W. J. Couch, and T. S. C. Project. Measurements of Omega and Lambda from 42 High-Redshift Supernovae. *apj*, 517:565–586, June 1999.

[22] A. C. Phillips and L. E. Davis. Registering, PSF-Matching and Intensity-Matching Images in IRAF. In R. A. Shaw, H. E. Payne, and J. J. E. Hayes, editors, *Astronomical Data Analysis Software and Systems IV*, volume 77 of *Astronomical Society of the Pacific Conference Series*, page 297, 1995.

[23] G. Pignata, J. Maza, R. Antezana, R. Cartier, G. Folatelli, F. Forster, L. Gonzalez, P. Gonzalez, M. Hamuy, D. Iturra, P. Lopez, S. Silva, B. Conuel, A. Crain, D. Foster,

K. Ivarsen, A. Lacluyze, M. Nysewander, and D. Reichart. The CHilean Automatic Supernova sEarch (CHASE). In G. Giobbi, A. Tornambe, G. Raimondo, M. Limongi, L. A. Antonelli, N. Menci, and E. Brocato, editors, *American Institute of Physics Conference Series*, volume 1111 of *American Institute of Physics Conference Series*, pages 551–554, May 2009.

[24] John C. Russ. *The image processing handbook*. CRC Press, 2007.

[25] D.J. Sand, M.L. Graham, C. Bildfell, R.J. Foley, C. Pritchet, et al. Intracluster supernovae in the Multi-epoch Nearby Cluster Survey. *Astrophys.J.*, 729:142, 2011.

[26] M. F. Skrutskie, R. M. Cutri, R. Stiening, M. D. Weinberg, S. Schneider, J. M. Carpenter, C. Beichman, R. Capps, T. Chester, J. Elias, J. Huchra, J. Liebert, C. Lonsdale, D. G. Monet, S. Price, P. Seitzer, T. Jarrett, J. D. Kirkpatrick, J. E. Gizis, E. Howard, T. Evans, J. Fowler, L. Fullmer, R. Hurt, R. Light, E. L. Kopan, K. A. Marsh, H. L. McCallon, R. Tam, S. Van Dyk, and S. Wheelock. The Two Micron All Sky Survey (2MASS). *The Astronomical Journal*, 131(2):1163, 2006.

[27] C Smith R. Real-time time-variability analysis of GB to TB datasets: experience from SuperMacho and Supernova projects at NOAO/CTIO. *Proccedings- Spie The International Society for Optical Engineering*, 4836:395–405, 2003.

[28] Murtag F. Starck J.L. *Astronomical Image and Data Analysis*. Springer Berlin Heidelberg New York, 2002.

[29] Mark Sullivan et al. Rates and properties of type Ia supernovae as a function of mass and star-formation in their host galaxies. *Astrophys.J.*, 648:868–883, 2006.

[30] Austin B. Tomaney and Arlin P.S. Crotts. Expanding the realm of microlensing surveys with difference image photometry. *Astron.J.*, 112:2872–2895, 1996.

[31] D.C. Wells, E.W. Greisen, and R.H. Harten. FITS a Flexible Image Transport System. *aaps*, 44:363, June 1981.

[32] HKC. Yee. A faint-galaxy photometry and image-analysis system. 103:396–411, April 1991.

[33] Fang Yuan and Carl W. Akerlof. Astronomical Image Subtraction by Cross-Convolution. 2008.

Apéndice A

Resultados de clasificación con funciones de distancia

Apéndice A

Resultados del clasificador Región pgc36664176

Clasificación con DTW

Vecinos	Umbral	SN	%
1	0.1	5	0.3360655738
1	0.15	5	0.3360655738
1	0.2	5	0.3360655738
1	0.25	5	0.3360655738
1	0.3	5	0.3360655738
1	0.35	5	0.3360655738
1	0.4	5	0.3360655738
1	0.45	5	0.3360655738
1	0.5	5	0.3360655738
1	0.55	5	0.3360655738
1	0.6	5	0.3360655738
1	0.65	5	0.3360655738
1	0.7	5	0.3360655738
1	0.75	5	0.3360655738
1	0.8	5	0.3360655738
1	0.85	5	0.3360655738
1	0.9	5	0.3360655738
1	0.95	5	0.3360655738
2	0.1	5	0.3360655738
2	0.15	5	0.3360655738
2	0.2	5	0.3360655738
2	0.25	5	0.3360655738
2	0.3	5	0.3360655738
2	0.35	5	0.3360655738
2	0.4	5	0.3360655738
2	0.45	5	0.3360655738
2	0.5	5	0.3360655738
2	0.55	5	0.3360655738
2	0.6	5	0.3360655738
2	0.65	5	0.3360655738
2	0.7	5	0.3360655738
2	0.75	5	0.3360655738
2	0.8	5	0.3360655738
2	0.85	5	0.3360655738
2	0.9	5	0.3360655738
2	0.95	5	0.3360655738
3	0.1	5	0.3360655738
3	0.15	5	0.3360655738
3	0.2	5	0.3360655738
3	0.25	5	0.3360655738
3	0.3	5	0.3360655738
3	0.35	5	0.3360655738
3	0.4	5	0.3360655738

Vecinos	Umbral	SN	%
5	0.1	5	0.336065574
5	0.15	5	0.336065574
5	0.2	5	0.336065574
5	0.25	5	0.336065574
5	0.3	5	0.336065574
5	0.35	5	0.336065574
5	0.4	5	0.336065574
5	0.45	5	0.336065574
5	0.5	5	0.336065574
5	0.55	5	0.336065574
5	0.6	5	0.336065574
5	0.65	5	0.336065574
5	0.7	5	0.336065574
5	0.75	5	0.336065574
5	0.8	5	0.336065574
5	0.85	5	0.336065574
5	0.9	5	0.336065574
5	0.95	5	0.336065574
6	0.1	5	0.336065574
6	0.15	5	0.336065574
6	0.2	5	0.336065574
6	0.25	5	0.336065574
6	0.3	5	0.336065574
6	0.35	5	0.336065574
6	0.4	5	0.336065574
6	0.45	5	0.336065574
6	0.5	5	0.336065574
6	0.55	5	0.336065574
6	0.6	5	0.336065574
6	0.65	5	0.336065574
6	0.7	5	0.336065574
6	0.75	5	0.336065574
6	0.8	5	0.336065574
6	0.85	5	0.336065574
6	0.9	5	0.336065574
6	0.95	5	0.336065574
7	0.1	5	0.336065574
7	0.15	5	0.336065574
7	0.2	5	0.336065574
7	0.25	5	0.336065574
7	0.3	5	0.336065574
7	0.35	5	0.336065574
7	0.4	5	0.336065574

3	0.45	5	0.3360655738
3	0.5	5	0.3360655738
3	0.55	5	0.3360655738
3	0.6	5	0.3360655738
3	0.65	5	0.3360655738
3	0.7	5	0.3360655738
3	0.75	5	0.3360655738
3	0.8	5	0.3360655738
3	0.85	5	0.3360655738
3	0.9	5	0.3360655738
3	0.95	5	0.3360655738
4	0.1	5	0.3360655738
4	0.15	5	0.3360655738
4	0.2	5	0.3360655738
4	0.25	5	0.3360655738
4	0.3	5	0.3360655738
4	0.35	5	0.3360655738
4	0.4	5	0.3360655738
4	0.45	5	0.3360655738
4	0.5	5	0.3360655738
4	0.55	5	0.3360655738
4	0.6	5	0.3360655738
4	0.65	5	0.3360655738
4	0.7	5	0.3360655738
4	0.75	5	0.3360655738
4	0.8	5	0.3360655738
4	0.85	5	0.3360655738
4	0.9	5	0.3360655738
4	0.95	5	0.3360655738

7	0.45	5	0.336065574
7	0.5	5	0.336065574
7	0.55	5	0.336065574
7	0.6	5	0.336065574
7	0.65	5	0.336065574
7	0.7	5	0.336065574
7	0.75	5	0.336065574
7	0.8	5	0.336065574
7	0.85	5	0.336065574
7	0.9	5	0.336065574
7	0.95	5	0.336065574
8	0.1	5	0.336065574
8	0.15	5	0.336065574
8	0.2	5	0.336065574
8	0.25	5	0.336065574
8	0.3	5	0.336065574
8	0.35	5	0.336065574
8	0.4	5	0.336065574
8	0.45	5	0.336065574
8	0.5	5	0.336065574
8	0.55	5	0.336065574
8	0.6	5	0.336065574
8	0.65	5	0.336065574
8	0.7	5	0.336065574
8	0.75	5	0.336065574
8	0.8	5	0.336065574
8	0.85	5	0.336065574
8	0.9	5	0.336065574
8	0.95	5	0.336065574

**********MEJOR RESULTADO**********
Porcentaje: 0.33606557377
Vecinos: 8
Umbral: 0.95
Numero de supernovas: 5

	SN
SN	5
C	4
V	70

**********MEJOR RESULTADO CLASIFICANDO SUPERNOVAS**********
Nsupernovas: 5
Vecinos: 8
Umbral: 0.95
Porcentaje de aciertos: 0.33606557377

Resultados de clasificador Región pgc18889954

Clasificación con DTW

Vecinos	Umbral	SN	%
1	0.1	20	0.2352941176
1	0.15	20	0.2352941176
1	0.2	20	0.2352941176
1	0.25	20	0.2352941176
1	0.3	20	0.2352941176
1	0.35	20	0.2352941176
1	0.4	20	0.2352941176
1	0.45	20	0.2352941176
1	0.5	20	0.2352941176
1	0.55	20	0.2352941176
1	0.6	20	0.2352941176
1	0.65	20	0.2352941176
1	0.7	20	0.2352941176
1	0.75	20	0.2352941176
1	0.8	20	0.2352941176
1	0.85	20	0.2352941176
1	0.9	20	0.2352941176
1	0.95	20	0.2352941176
2	0.1	20	0.2352941176
2	0.15	20	0.2352941176
2	0.2	20	0.2352941176
2	0.25	20	0.2352941176
2	0.3	20	0.2352941176
2	0.35	20	0.2352941176
2	0.4	20	0.2352941176
2	0.45	20	0.2352941176
2	0.5	20	0.2352941176
2	0.55	20	0.2352941176
2	0.6	20	0.2352941176
2	0.65	20	0.2352941176
2	0.7	20	0.2352941176
2	0.75	20	0.2352941176
2	0.8	20	0.2352941176
2	0.85	20	0.2352941176
2	0.9	20	0.2352941176
2	0.95	20	0.2352941176
3	0.1	20	0.2352941176
3	0.15	20	0.2352941176
3	0.2	20	0.2352941176
3	0.25	20	0.2352941176
3	0.3	20	0.2352941176
3	0.35	20	0.2352941176
3	0.4	20	0.2352941176
3	0.45	20	0.2352941176
3	0.5	20	0.2352941176
3	0.55	20	0.2352941176

Vecinos	Umbral	SN	%
5	0.1	20	0.235294118
5	0.15	20	0.235294118
5	0.2	20	0.235294118
5	0.25	20	0.235294118
5	0.3	20	0.235294118
5	0.35	20	0.235294118
5	0.4	20	0.235294118
5	0.45	20	0.235294118
5	0.5	20	0.235294118
5	0.55	20	0.235294118
5	0.6	20	0.235294118
5	0.65	20	0.235294118
5	0.7	20	0.235294118
5	0.75	20	0.235294118
5	0.8	20	0.235294118
5	0.85	20	0.235294118
5	0.9	20	0.235294118
5	0.95	20	0.235294118
6	0.1	20	0.235294118
6	0.15	20	0.235294118
6	0.2	20	0.235294118
6	0.25	20	0.235294118
6	0.3	20	0.235294118
6	0.35	20	0.235294118
6	0.4	20	0.235294118
6	0.45	20	0.235294118
6	0.5	20	0.235294118
6	0.55	20	0.235294118
6	0.6	20	0.235294118
6	0.65	20	0.235294118
6	0.7	20	0.235294118
6	0.75	20	0.235294118
6	0.8	20	0.235294118
6	0.85	20	0.235294118
6	0.9	20	0.235294118
6	0.95	20	0.235294118
7	0.1	20	0.235294118
7	0.15	20	0.235294118
7	0.2	20	0.235294118
7	0.25	20	0.235294118
7	0.3	20	0.235294118
7	0.35	20	0.235294118
7	0.4	20	0.235294118
7	0.45	20	0.235294118
7	0.5	20	0.235294118
7	0.55	20	0.235294118

3	0.6	20	0.2352941176
3	0.65	20	0.2352941176
3	0.7	20	0.2352941176
3	0.75	20	0.2352941176
3	0.8	20	0.2352941176
3	0.85	20	0.2352941176
3	0.9	20	0.2352941176
3	0.95	20	0.2352941176
4	0.1	20	0.2352941176
4	0.15	20	0.2352941176
4	0.2	20	0.2352941176
4	0.25	20	0.2352941176
4	0.3	20	0.2352941176
4	0.35	20	0.2352941176
4	0.4	20	0.2352941176
4	0.45	20	0.2352941176
4	0.5	20	0.2352941176
4	0.55	20	0.2352941176
4	0.6	20	0.2352941176
4	0.65	20	0.2352941176
4	0.7	20	0.2352941176
4	0.75	20	0.2352941176
4	0.8	20	0.2352941176
4	0.85	20	0.2352941176
4	0.9	20	0.2352941176
4	0.95	20	0.2352941176

7	0.6	20	0.235294118
7	0.65	20	0.235294118
7	0.7	20	0.235294118
7	0.75	20	0.235294118
7	0.8	20	0.235294118
7	0.85	20	0.235294118
7	0.9	20	0.235294118
7	0.95	20	0.235294118
8	0.1	20	0.235294118
8	0.15	20	0.235294118
8	0.2	20	0.235294118
8	0.25	20	0.235294118
8	0.3	20	0.235294118
8	0.35	20	0.235294118
8	0.4	20	0.235294118
8	0.45	20	0.235294118
8	0.5	20	0.235294118
8	0.55	20	0.235294118
8	0.6	20	0.235294118
8	0.65	20	0.235294118
8	0.7	20	0.235294118
8	0.75	20	0.235294118
8	0.8	20	0.235294118
8	0.85	20	0.235294118
8	0.9	20	0.235294118
8	0.95	20	0.235294118

**********MEJOR RESULTADO**********
Porcentaje: 0.235294117647
Vecinos: 8
Umbral: 0.95
Numero de supernovas: 20

	SN
SN	20
C	12
V	169

**********MEJOR RESULTADO CLASIFICANDO SUPERNOVAS**********
Nsupernovas: 20
Vecinos: 8
Umbral: 0.95
Porcentaje de aciertos: 0.235294117647

Resultados del clasificador con coeficiente de correlación
Clasificación de Región pgc36664176

Vecinos	Umbral	SN	%
1	0.1	5	0.3360655738
1	0.15	5	0.3360655738
1	0.2	5	0.3360655738
1	0.25	5	0.3360655738
1	0.3	5	0.3360655738
1	0.35	5	0.3606557377
1	0.4	5	0.3852459016
1	0.45	5	0.4098360656
1	0.5	5	0.4180327869
1	0.55	5	0.4344262295
1	0.6	4	0.4508196721
1	0.65	4	0.5
1	0.7	4	0.5491803279
1	0.75	4	0.6229508197
1	0.8	3	0.6557377049
1	0.85	3	0.737704918
1	0.9	3	0.7704918033
1	0.95	3	0.8196721311
2	0.1	5	0.3360655738
2	0.15	5	0.3360655738
2	0.2	5	0.3360655738
2	0.25	5	0.3360655738
2	0.3	5	0.3442622951
2	0.35	5	0.3770491803
2	0.4	5	0.3852459016
2	0.45	5	0.4180327869
2	0.5	5	0.4262295082
2	0.55	5	0.4426229508
2	0.6	4	0.4672131148
2	0.65	4	0.5163934426
2	0.7	4	0.5655737705
2	0.75	4	0.6475409836
2	0.8	3	0.6885245902
2	0.85	3	0.7540983607
2	0.9	3	0.7868852459
2	0.95	3	0.8278688525
3	0.1	5	0.3360655738
3	0.15	5	0.3360655738
3	0.2	5	0.3360655738
3	0.25	5	0.3442622951
3	0.3	5	0.3442622951
3	0.35	5	0.3770491803
3	0.4	5	0.393442623
3	0.45	5	0.4180327869
3	0.5	5	0.4426229508
3	0.55	4	0.4672131148

Vecinos	Umbral	SN	%
5	0.1	5	0.336065574
5	0.15	5	0.336065574
5	0.2	5	0.336065574
5	0.25	5	0.344262295
5	0.3	5	0.368852459
5	0.35	5	0.385245902
5	0.4	5	0.409836066
5	0.45	5	0.43442623
5	0.5	5	0.450819672
5	0.55	4	0.491803279
5	0.6	4	0.508196721
5	0.65	4	0.549180328
5	0.7	4	0.639344262
5	0.75	4	0.696721312
5	0.8	3	0.754098361
5	0.85	3	0.778688525
5	0.9	3	0.803278689
5	0.95	3	0.852459016
6	0.1	5	0.336065574
6	0.15	5	0.336065574
6	0.2	5	0.336065574
6	0.25	5	0.352459016
6	0.3	5	0.368852459
6	0.35	5	0.393442623
6	0.4	5	0.409836066
6	0.45	5	0.43442623
6	0.5	4	0.459016393
6	0.55	4	0.491803279
6	0.6	4	0.532786885
6	0.65	4	0.573770492
6	0.7	4	0.647540984
6	0.75	4	0.696721312
6	0.8	3	0.754098361
6	0.85	3	0.778688525
6	0.9	3	0.81147541
6	0.95	3	0.860655738
7	0.1	5	0.336065574
7	0.15	5	0.336065574
7	0.2	5	0.336065574
7	0.25	5	0.352459016
7	0.3	5	0.368852459
7	0.35	5	0.393442623
7	0.4	5	0.418032787
7	0.45	5	0.43442623
7	0.5	4	0.475409836
7	0.55	4	0.491803279

3	0.6	4	0.4918032787
3	0.65	4	0.5327868852
3	0.7	4	0.5901639344
3	0.75	4	0.6557377049
3	0.8	3	0.7131147541
3	0.85	3	0.762295082
3	0.9	3	0.7950819672
3	0.95	3	0.8442622951
4	0.1	5	0.3360655738
4	0.15	5	0.3360655738
4	0.2	5	0.3360655738
4	0.25	5	0.3442622951
4	0.3	5	0.3524590164
4	0.35	5	0.3770491803
4	0.4	5	0.393442623
4	0.45	5	0.4262295082
4	0.5	5	0.4426229508
4	0.55	4	0.4836065574
4	0.6	4	0.5081967213
4	0.65	4	0.5327868852
4	0.7	4	0.6229508197
4	0.75	4	0.6885245902
4	0.8	3	0.7213114754
4	0.85	3	0.7704918033
4	0.9	3	0.8032786885
4	0.95	3	0.8524590164

7	0.6	4	0.540983607
7	0.65	4	0.581967213
7	0.7	4	0.655737705
7	0.75	3	0.713114754
7	0.8	3	0.754098361
7	0.85	3	0.786885246
7	0.9	3	0.81147541
7	0.95	3	0.868852459
8	0.1	5	0.336065574
8	0.15	5	0.336065574
8	0.2	5	0.336065574
8	0.25	5	0.352459016
8	0.3	5	0.368852459
8	0.35	5	0.393442623
8	0.4	5	0.426229508
8	0.45	5	0.442622951
8	0.5	4	0.475409836
8	0.55	4	0.5
8	0.6	4	0.540983607
8	0.65	4	0.614754098
8	0.7	4	0.663934426
8	0.75	3	0.737704918
8	0.8	3	0.770491803
8	0.85	3	0.803278689
8	0.9	3	0.81147541
8	0.95	2	0.860655738

**********MEJOR RESULTADO**********
Porcentaje: 0.868852459016
Vecinos: 7
Umbral: 0.95
Numero de supernovas: 3

	SN
SN	3
C	0
V	3

**********MEJOR RESULTADO CLASIFICANDO SUPERNOVAS**********
Nsupernovas: 5
Vecinos: 8
Umbral: 0.45
Porcentaje de aciertos: 0.44262295082

Resultados de clasificación con coeficiente de correlación
Clasificación de la región pgc18889954

Vecinos	Umbral	SN	%
1	0.1	20	0.2352941176
1	0.15	20	0.2352941176
1	0.2	20	0.2352941176
1	0.25	20	0.2394957983
1	0.3	20	0.2521008403
1	0.35	20	0.2773109244
1	0.4	20	0.2983193277
1	0.45	20	0.3109243697
1	0.5	20	0.3319327731
1	0.55	20	0.3655462185
1	0.6	20	0.3907563025
1	0.65	20	0.4453781513
1	0.7	19	0.487394958
1	0.75	19	0.5294117647
1	0.8	18	0.5798319328
1	0.85	16	0.6302521008
1	0.9	13	0.7268907563
1	0.95	9	0.8067226891
2	0.1	20	0.2352941176
2	0.15	20	0.2352941176
2	0.2	20	0.2352941176
2	0.25	20	0.243697479
2	0.3	20	0.2647058824
2	0.35	20	0.2857142857
2	0.4	20	0.3025210084
2	0.45	20	0.3109243697
2	0.5	20	0.3529411765
2	0.55	20	0.3739495798
2	0.6	20	0.4201680672
2	0.65	20	0.474789916
2	0.7	18	0.512605042
2	0.75	18	0.5462184874
2	0.8	17	0.6008403361
2	0.85	14	0.6512605042
2	0.9	11	0.7478991597
2	0.95	9	0.8361344538
3	0.1	20	0.2352941176
3	0.15	20	0.2352941176
3	0.2	20	0.2352941176
3	0.25	20	0.2521008403
3	0.3	20	0.2773109244
3	0.35	20	0.2941176471
3	0.4	20	0.3067226891
3	0.45	20	0.3193277311
3	0.5	20	0.3613445378
3	0.55	20	0.3907563025

Vecinos	Umbral	SN	%
5	0.1	20	0.235294118
5	0.15	20	0.235294118
5	0.2	20	0.235294118
5	0.25	20	0.264705882
5	0.3	20	0.285714286
5	0.35	20	0.306722689
5	0.4	20	0.319327731
5	0.45	20	0.340336135
5	0.5	20	0.386554622
5	0.55	20	0.420168067
5	0.6	18	0.466386555
5	0.65	18	0.516806723
5	0.7	16	0.546218487
5	0.75	13	0.571428571
5	0.8	12	0.613445378
5	0.85	11	0.722689076
5	0.9	11	0.781512605
5	0.95	7	0.865546219
6	0.1	20	0.235294118
6	0.15	20	0.235294118
6	0.2	20	0.239495798
6	0.25	20	0.277310924
6	0.3	20	0.289915966
6	0.35	20	0.31092437
6	0.4	20	0.323529412
6	0.45	20	0.357142857
6	0.5	20	0.399159664
6	0.55	20	0.420168067
6	0.6	18	0.478991597
6	0.65	16	0.521008403
6	0.7	15	0.56302521
6	0.75	13	0.588235294
6	0.8	11	0.634453782
6	0.85	11	0.739495798
6	0.9	10	0.794117647
6	0.95	7	0.869747899
7	0.1	20	0.235294118
7	0.15	20	0.235294118
7	0.2	20	0.24789916
7	0.25	20	0.281512605
7	0.3	20	0.289915966
7	0.35	20	0.31512605
7	0.4	20	0.327731092
7	0.45	20	0.37394958
7	0.5	20	0.411764706
7	0.55	20	0.424369748

3	0.6	20	0.4369747899
3	0.65	19	0.4915966387
3	0.7	18	0.5294117647
3	0.75	17	0.5588235294
3	0.8	14	0.5924369748
3	0.85	11	0.6680672269
3	0.9	11	0.756302521
3	0.95	9	0.8445378151
4	0.1	20	0.2352941176
4	0.15	20	0.2352941176
4	0.2	20	0.2352941176
4	0.25	20	0.2605042017
4	0.3	20	0.281512605
4	0.35	20	0.2983193277
4	0.4	20	0.3151260504
4	0.45	20	0.3361344538
4	0.5	20	0.3697478992
4	0.55	20	0.4159663866
4	0.6	20	0.4537815126
4	0.65	18	0.5042016807
4	0.7	17	0.5420168067
4	0.75	16	0.5798319328
4	0.8	13	0.6050420168
4	0.85	11	0.7100840336
4	0.9	11	0.768907563
4	0.95	8	0.8571428571

7	0.6	16	0.483193277
7	0.65	16	0.529411765
7	0.7	13	0.558823529
7	0.75	12	0.592436975
7	0.8	11	0.655462185
7	0.85	11	0.74789916
7	0.9	9	0.81512605
7	0.95	6	0.869747899
8	0.1	20	0.235294118
8	0.15	20	0.235294118
8	0.2	20	0.256302521
8	0.25	20	0.281512605
8	0.3	20	0.289915966
8	0.35	20	0.31512605
8	0.4	20	0.336134454
8	0.45	20	0.37394958
8	0.5	20	0.411764706
8	0.55	19	0.441176471
8	0.6	16	0.491596639
8	0.65	15	0.537815126
8	0.7	13	0.56302521
8	0.75	12	0.617647059
8	0.8	11	0.659663866
8	0.85	11	0.75210084
8	0.9	9	0.827731092
8	0.95	5	0.869747899

**********MEJOR RESULTADO**********
Porcentaje: 0.86974789916
Vecinos: 8
Umbral: 0.95
Numero de supernovas: 5

	SN
SN	5
C	2
V	3

**********MEJOR RESULTADO CLASIFICANDO SUPERNOVAS**********
Nsupernovas: 20
Vecinos: 8
Umbral: 0.5
Porcentaje de aciertos: 0.411764705882

Apéndice B

Glosario de términos y acrónimos

- **2MASS:** Two Micron All Sky Survey. Es un proyecto desarrollado por University of Massachusetts, Infrared Processing and Analysis Center y el California Institute of Technology.
- **Abell cluster:** Cúmulos de galaxias con corrimiento al rojo nominal z $\leq$ 0,2. Cada cúmulo debe tener una población mínima de 50 miembros.
- **AGN:** Active Galactic Nucleus. Una galaxia se dice activa cuando una fracción significativa de la radiación electromagnética que emite no es debida a los componentes "normales" de una galaxia.
- **ANN:** Artificial Neural Network. Se trata de un sistema de interconexión de neuronas que colaboran entre sí para producir un estímulo de salida.
- **AR:** Ascensión recta. En astronomía, la ascensión recta es una de las coordenadas astronómicas que se utilizan para localizar los astros sobre la esfera celeste.
- **CCD:** Charge-Coupled Device. Es el sensor con diminutas células fotoeléctricas que registran la imagen.
- **CHASE:** Chilean Automatic Supernovae SEarch. Es un proyecto de astronomía de la Universidad de Chile.
- **CRTS:** Catalina Real-time Transient Survey. Es un proyecto que busca objetos variables como supernovas, galaxias de núcleo activo, blazares entre otros.
- **DCT:** Discrete Cosine Transform. Es una transformada basada en la Transformada de Fourier discreta, usando solo la parte real.
- **DEC:** Declinación. En astronomía, la declinación es el ángulo que forma un astro con el ecuador celeste.
- **DTW:** Dynamic Time Warping. Es un algoritmo para medir la similitud entre dos series de tiempo.

- **FITS:** Flexible Image Transport System. Es el formato de archivo más utilizado comúnmente en el mundo de la astronomía.
- **FP:** False positive. Es una medida del modelo donde se predice la pertenencia a una clase a la cual no pertenece.
- **FN:** False negative. En clasificación es el número de instancias pertenecientes a otras clases que se clasificaron erróneamente como una clase en particular.
- **GALFIT:** es una herramienta de software para extraer información de galaxias, estrellas, cúmulos globulares, discos estelares de imágenes.
- **Ground truth:** datos reales, tomados por observación de la verdadera naturaleza del fenómeno. Constituyen los datos de prueba, necesarios para determinar el porcentaje de aciertos de un clasificador supervisad
- **GTM:** Gran Telescopio Milimétrico.
- **IDCT:** Inverse Discrete Cosine Transform.
- **IRAF:** Image Reduction and Analysis Facility. Consiste en una gran colección de software escrito por astrónomos y programadores mantenida por el Observatorio Nacional de Astronomía Óptica (NOAO).
- **KNN:** K-Nearest Neighbors. Es un método de clasificación supervisada.
- **PAN-STARSS:** Panoramic Survey Telescope & Rapid Response System.
- **PPP:** Picture Processing Program: Es un software para procesamiento de imágenes de astronomía, principalmente fotometría.
- **PROMPT:** Panchromatic Robotic Optical Monitoring and Polarimetry Telescopes.
- **PSF:** Point Spread Function. Describe la respuesta de un sistema de imagen a una fuente de punto o punto objeto.
- **RGB:** red, green, blue. Es la composición del color en términos de la intensidad de los colores primarios de la luz.
- **SDSS:** Sloan Digital Sky Survey.
- **SDS:** Supernovae Detection Software. Es el nombre del software desarrollado en este trabajo de tesis.
- **SN:** En astronomía es la abreviación de supernova.
- **SQL:** Structured Query Language, es un lenguaje declarativo de acceso a bases de datos relacionales.
- **SVM:** Support Vector Machines.

- **TP:** True positive. Número de estancias que un modelo predijo correctamente.
- **TN:** True negative. Es el número de instancias que no se clasificaron como una clase en particular.
- **WOOTS:** Wise Observatory Optical Transient Search.