



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE INGENIERÍA

**Reducción de ruido en imágenes
para una cámara-espectrógrafo en
infrarrojo medio**

TESIS

Que para obtener el título de
Ingeniero Eléctrico Electrónico

P R E S E N T A

Enrique Escobedo Cousin

DIRECTOR DE TESIS

Elfego Ruiz Schneider



Ciudad Universitaria, Cd. Mx., 2003

Índice

I. Instrumentación astronómica 2

1.1 Telescopios 2

1.2 Detectores 5

1.2.1 Dispositivo de carga acoplada (CCD) 5

1.3 Otros instrumentos 8

1.3.1 Fotómetros 8

1.3.2 Interferómetros 8

1.3.3 Espectrómetros 9

II. Astronomía infrarroja 10

2.1 La región infrarroja del espectro electromagnético 10

2.1.1 Ventanas atmosféricas y emisión térmica 12

2.2 Telescopios infrarrojos 13

2.3 Detectores de infrarrojo 13

2.3.1 Historia 14

2.3.2 Clasificación 15

2.3.3 Detectores de conducción en la banda de impurezas (IBC) 17

2.3.4 Arreglos infrarrojos 19

2.5 La cámara-espectrógrafo de infrarrojo medio 21

2.5.1 Fuentes de ruido 21

III. Marco teórico 22

3.1 Generalidades sobre imágenes digitales 22

3.1.1 Restauración de imágenes 22

3.2 Transformación *wavelet* 24

3.2.1 Análisis de Fourier 24

3.2.2 Transformación *wavelet* continua 27

3.2.3 Transformación *wavelet* discreta 28

3.3 Las *wavelets* en imágenes 29

3.3.1 Reducción de ruido con *wavelets* 31

IV. Algoritmo de reducción de ruido	32
4.1 De la compresión a la reducción de ruido	32
4.2 Árboles de ceros	32
4.3 Modelado estadístico de los coeficientes <i>wavelet</i>	34
4.3.1 Estimador del mínimo error cuadrático medio (MMSE)	34
4.3.2 Estimador de máxima similitud (ML)	35
4.4 Integración del algoritmo	36
V. Resultados y conclusiones	38
5.1 Discusión de los resultados	38
5.2 Conclusiones	41
5.3 Trabajo a futuro	42
Apéndice A	44
A.1 Resultados en diversas imágenes	44
A.2 Pruebas con otros tipos de ruido	49
Referencias bibliográficas	51

Presentación

La Astronomía es la más antigua de las ciencias y, sin embargo, no ha dejado de despertar la curiosidad de los seres humanos; es una precursora de la tecnología y aún hoy está a la vanguardia; es una piedra rodante del conocimiento a la que no se le ha pegado el polvo. Asociados a la investigación y a los descubrimientos astronómicos siempre pueden encontrarse proyectos de desarrollo tecnológico que involucran a distintas áreas del conocimiento, cuyo impacto, usualmente, se ve reflejado en una gran cantidad de aplicaciones en diversos ámbitos.

El presente trabajo es parte de un proyecto mayor que se lleva a cabo en el Instituto de Astronomía de la Universidad Nacional Autónoma de México. Se trata de la construcción de una cámara-espectrógrafo de infrarrojo medio para el observatorio de San Pedro Mártir, con lo que se busca proveer a la comunidad astronómica mexicana de una herramienta que ofrezca la posibilidad de explorar el universo de una manera distinta, con la motivación de observar lo que es invisible en el intervalo óptico del espectro. Será posible estudiar regiones de formación de estrellas, núcleos de galaxias, atmósferas estelares y nubes planetarias, entre otras áreas de interés. Cabe destacar que las particularidades de la astronomía infrarroja exigen métodos y técnicas especiales que la convierten en una disciplina disímil de la actividad astronómica óptica, con sutilezas y dificultades singulares que suponen retos interesantes para el desarrollo tecnológico como herramienta de la investigación. El objetivo de esta tesis es el diseño y la implementación de un algoritmo de reducción de ruido, enfocado a las cualidades específicas que habrán de presentarse en las imágenes obtenidas por la cámara-espectrógrafo de IR medio. Dado que este instrumento continúa en las primeras etapas del diseño, únicamente será posible contar con imágenes semejantes a las que se espera adquirir, de tal suerte que será necesario hacer consideraciones y suposiciones cuya validez estará en función del bagaje científico acumulado en cuestiones de la observación astronómica, del funcionamiento de los equipos y de los dispositivos de los que ésta se vale, de los fenómenos físicos relacionados con la radiación electromagnética, de la recolección y almacenamiento de los datos, etc.

El primer capítulo de este documento presenta las generalidades de la observación astronómica a través de la descripción de las metodologías y de los equipos de uso más frecuente, así como de la definición de los conceptos básicos, con lo que se pretende dar el primer paso en la ensambladura del marco de referencia de este trabajo. El segundo paso consiste en mostrar los detalles de la astronomía infrarroja, para lo cual se parte con una reseña de dicha región del espectro, para concluir con el registro de los distintos tipos de detectores infrarrojos y la utilización de uno de ellos en la cámara-espectrógrafo que se proyecta. La tercera parte de la tesis se examina la teoría de análisis de señales con *wavelets* como una herramienta viable en la solución del problema. En el capítulo cuarto es presentada la integración del algoritmo de reducción de ruido a partir de técnicas novedosas para la compresión de imágenes recientemente introducidas. Finalmente se evalúan los resultados obtenidos al cotejarlos con los métodos incluidos en un *software* comercial.

Capítulo I. Instrumentación astronómica

1.1 Telescopios

El funcionamiento básico de un telescopio se ilustra en la figura 1.1. La imagen de interés u objeto, situada a una distancia finita a partir del lente objetivo es refractada formando una imagen intermedia que será el objeto del siguiente sistema de lentes, el ocular. Esta imagen debe situarse entre el segundo sistema y su foco f_e . La imagen final se forma después del ocular.

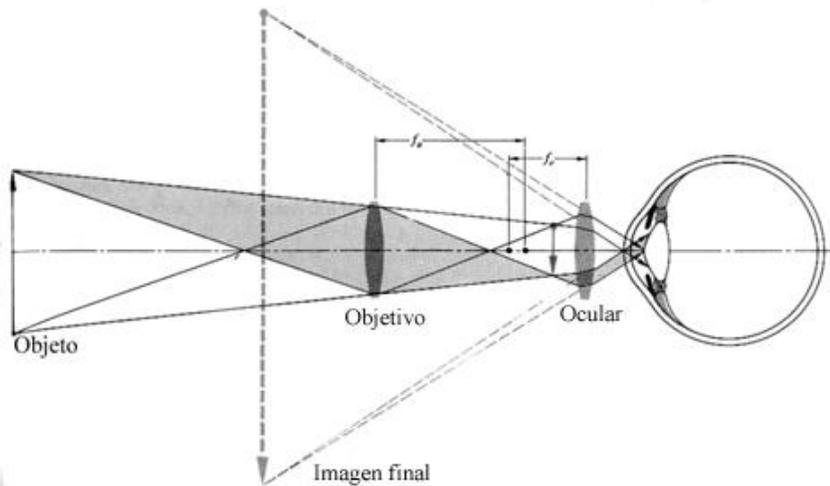


Figura 1.1 Diagrama de un telescopio [7]

Es importante considerar que para objetos muy distantes los rayos incidentes son paralelos, de modo que la imagen intermedia coincide con el segundo foco del objetivo. También es común que el primer foco del ocular coincida en el mismo punto. Si esto sucede, la distancia d entre ambos lentes es igual a la suma de sus distancias focales f_e y f_o dando como resultado un sistema afocal, es decir, tanto los rayos incidentes como los que están a la salida son paralelos entre sí, como puede verse en la figura 1.2.

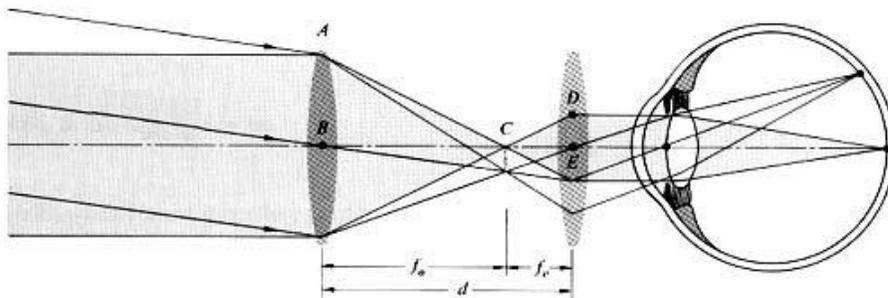


Figura 1.2 Diagrama de un telescopio astronómico [7]

Los telescopios se dividen en refractores y reflectores. Los primeros utilizan lentes mientras los segundos utilizan espejos y son éstos últimos los de mayor uso en la Astronomía moderna. Existen cuatro versiones principales de telescopios reflectores y en todos los casos el principio de funcionamiento es el mismo, las funciones de los espejos primario y secundario son análogas, respectivamente, a las de los lentes objetivo y ocular en los telescopios refractores. Hay que mencionar que a pesar de que el espejo secundario bloquea cierta zona del campo visual del primario, la imagen no presenta “agujeros” porque los rayos de luz que inciden libremente son suficientes para formar una imagen completa al ser enfocados en conjunto.

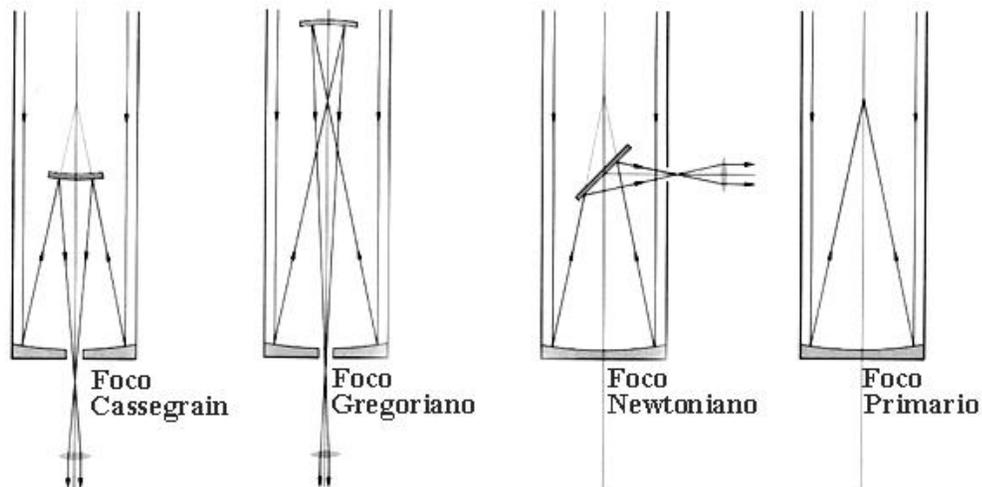


Figura 1.3 Versiones de telescopios [7]

La *resolución* de un telescopio es su capacidad de distinguir detalles finos en las imágenes, por ejemplo, el hecho de que dos estrellas pequeñas separadas por una distancia corta se vean como un sistema binario y no como una estrella grande, implica que un telescopio tiene suficiente resolución para detectarlas. Esta medida depende de la *apertura* del telescopio, que es el diámetro del haz de luz que incide y normalmente coincide con el diámetro del espejo secundario. Una apertura grande no sólo mejora la resolución, sino también la *capacidad de recolección de luz* del telescopio, que es una de las características más importantes. Con una buena capacidad de recolección de luz es posible que un objeto opaco se vea más brillante. La relación entre el diámetro del telescopio D y el foco del primario f se conoce como *apertura relativa*, y su inverso es la *razón focal*. Por ejemplo, un telescopio de diámetro de 1m y distancia focal de 10m tiene una razón focal de 10, que usualmente se denota $f/10$. Este parámetro se usa para especificar a los espejos primarios.

La acción de los espejos (o de los lentes en el caso de los telescopios refractores) puede resumirse en cambiar la forma del frente de onda incidente de manera que pueda concentrarse tanta energía como sea posible en un área pequeña, sin embargo, aún es necesario hacer modificaciones a porciones pequeñas del frente de onda debido a las imperfecciones que pueden presentarse como consecuencia de causas diversas. En primer lugar las irregularidades en la superficie de los espejos y errores en la alineación del sistema deterioran la calidad de la imagen (aberraciones). Sus efectos no se conocen con

anticipación, de modo que es necesario un sistema capaz de compensarlos. Hay que mencionar que es una tendencia en el desarrollo de los telescopios modernos el sustituir los grandes espejos primarios por arreglos de espejos pequeños (espejos primarios segmentados) como el del telescopio Keck, que se muestra en la figura 1.4. Esto simplifica la manufactura y disminuye los costos de fabricación, sin embargo, el pulido y la alineación de las partes resultan más complicados. Los sistemas de corrección de estos errores se conocen como *Óptica Activa*.

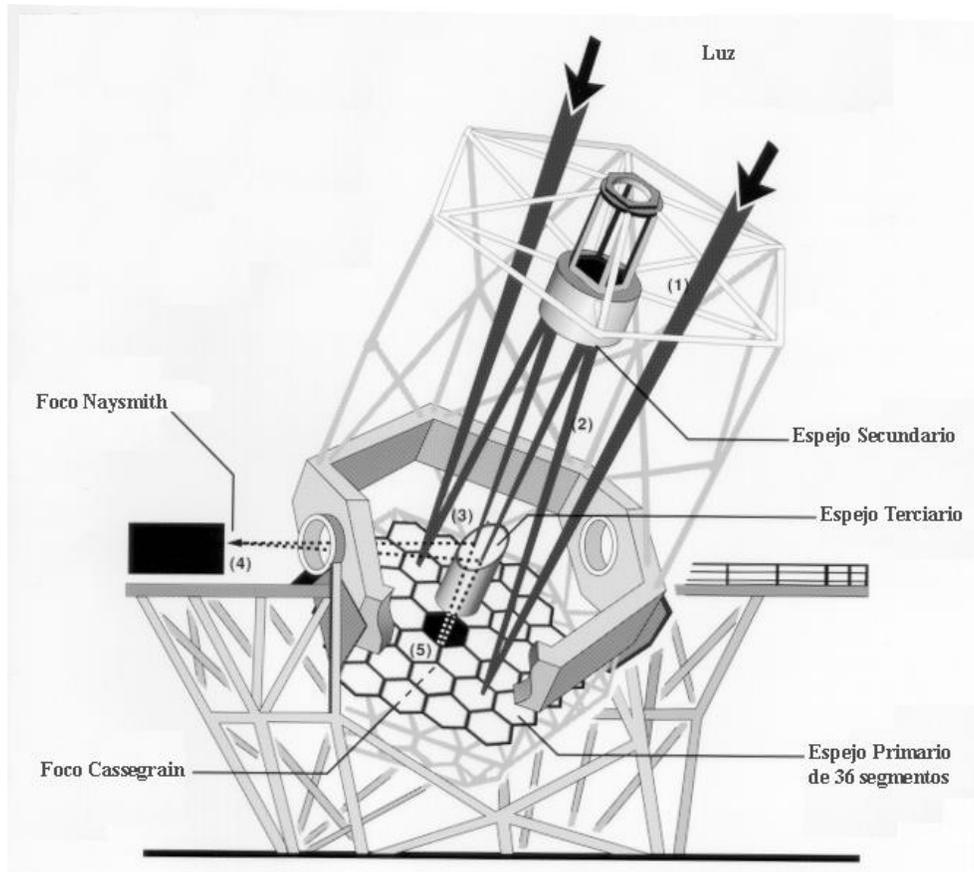


Figura 1.4 El telescopio Keck [25]

Otro factor que influye de manera importante la calidad de la imagen es la acción de la atmósfera, que además de deformar el frente de onda, absorbe algunas componentes del espectro electromagnético en virtud de las características físico-químicas de los gases que las constituyen, particularmente regiones del infrarrojo y del ultravioleta. Con el fin de mejorar el funcionamiento de los telescopios, se emplean sistemas activos para medir las aberraciones introducidas por la turbulencia atmosférica y compensarlas a través de actuadores que ajustan la superficie de los espejos que componen al primario. Estos sistemas se conocen como *Óptica Adaptativa*.

En los telescopios modernos la imagen reflejada por el secundario es capturada por sistemas complejos de lentes y dispositivos electrónicos que la adecuan a las necesidades del observador. Estos sistemas son intercambiables, de manera que el usuario puede elegir

cuáles utilizar según los requerimientos de sus observaciones, por ejemplo para hacer fotometría, interferometría o espectroscopía.

1.2 Detectores

Los estudios de Einstein acerca del *efecto fotoeléctrico* constituyen el principio básico de los detectores que algunas décadas más tarde comenzaron a desarrollarse con fines de observación astronómica. Según tal efecto, hay materiales que emiten electrones al estar expuestos a radiación electromagnética. El *tubo fotomultiplicador*, desarrollado en la década de los 40, fue el primer dispositivo fotoeléctrico. Un fotón incidente en la superficie de cierto material emite un electrón, éste es acelerado por un campo eléctrico hasta incidir en otra superficie fotoemisora logrando emitir dos o tres electrones que igualmente son acelerados hacia otra superficie. El proceso se repite varias veces hasta que la superficie final colecta todos los electrones. El tubo fotomultiplicador llegó a tener popularidad para hacer mediciones precisas de luminosidad.

En la década de los 60, la introducción de otros dispositivos fotoeléctricos abrió nuevas puertas a la observación astronómica. El impacto de los semiconductores dio lugar a una capacidad de recolección de datos que ha crecido enormemente en los últimos años, logrando que las nuevas generaciones de detectores tengan límites de sensibilidad cercanos a los máximos teóricos para algunas longitudes de onda. Casi la totalidad de la investigación astronómica es debida a equipos fotoelectrónicos que convierten las señales electromagnéticas de objetos distantes en datos digitales para ser tratados por una computadora.

1.2.1 Dispositivo de carga acoplada (CCD)

La estructura de los átomos y la manera en que éstos están dispuestos determinan algunas propiedades importantes de los sólidos. Cuando varios átomos conforman una estructura cristalina, los electrones en las órbitas más energéticas o *electrones de valencia* interactúan beneficiando la unión de los átomos y alterando las órbitas exteriores. El resultado de la interacción es que los núcleos atómicos comparten a los electrones exteriores en la llamada *banda de valencia*, dejando vacíos los niveles de más energía o *banda de conducción*. Para que un electrón pueda abandonar la banda de valencia para alcanzar la banda de conducción necesita una cantidad determinada de energía, en otras palabras, puede decirse que entre las dos bandas mencionadas anteriormente existe un *intervalo prohibido de energía* E_g , cuya extensión puede variar en función de la temperatura y del tipo de material.

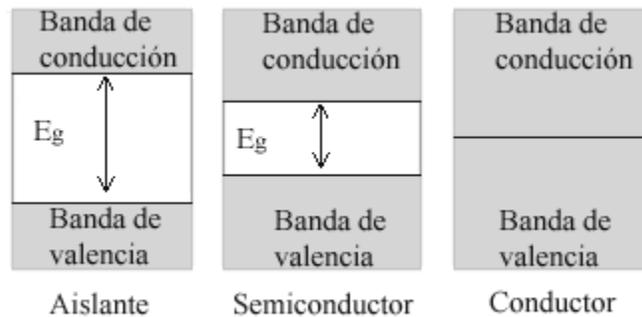


Figura 1.5 Diagrama de bandas [13]

Se sabe que los metales son buenos conductores de electricidad, esto es porque las bandas de valencia y de conducción se traslapan, por lo que los electrones pueden encontrarse libres en la banda de conducción. Por otro lado, en los malos conductores o aislantes como la madera, el intervalo prohibido es muy amplio y la banda de conducción permanece vacía siempre. Cuando el intervalo prohibido es estrecho puede ocurrir que un electrón pueda elevarse a la banda de conducción. Los elementos del grupo IV de la tabla periódica presentan esta característica y se conocen como *semiconductores*. También puede obtenerse semiconductores por medio de aleaciones entre elementos del grupo III con elementos del grupo V. Por convención se usa la notación *semiconductor : impureza*, por ejemplo, Si:As se refiere al silicio con el arsénico como impureza.

Cuando un fotón es absorbido por un semiconductor, su energía se transfiere a uno de los electrones de la banda de valencia, que puede alcanzar la banda de conducción. En el lugar que ocupaba en la banda de valencia queda un *hueco* cargado positivamente. Para cada semiconductor existe una longitud de onda de corte, en función de su intervalo prohibido, a partir de la cual el material deja de ser sensible a la radiación.

Pueden reemplazarse átomos de un semiconductor con átomos de un tipo distinto, quedando *dopado*. Si las impurezas tienen más electrones de valencia que los átomos del semiconductor original se dice que se trata de un material *tipo n*, en el que los portadores minoritarios de carga son los huecos; en el caso contrario, si tiene menos electrones de valencia, es un material *tipo p*, donde los portadores minoritarios son los electrones. En el primer caso el material *tipo n* dona los electrones excedentes a la banda de conducción.

En una unión *p-n* los electrones del material *tipo n* tienden a llenar los huecos del material *tipo p* y similarmente los huecos del segundo tienden a ir hacia el material *tipo n*. Esto produce que en una región alrededor de la unión se genere una barrera de potencial eléctrico que termina restringiendo la migración de los portadores de carga. Cuando un fotón de una frecuencia adecuada incide cerca de la unión, se crea un par electrón-hueco. El electrón es atraído a la región de mayor potencial inmersa en el material *tipo n*. La barrera de potencial o *región libre de carga* se reduce conforme aumente el número de electrones almacenados.

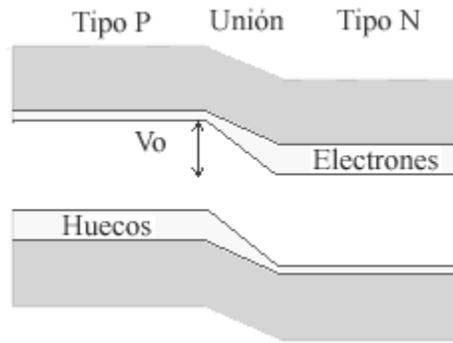


Figura 1.6 Diagrama de la unión P-N [13]

El CCD se conforma de un arreglo de renglones y columnas de elementos de imagen o *pixeles*, que pueden absorber radiación que libera a un electrón del sustrato semiconductor. Con el fin de que cada electrón permanezca en el área correspondiente al pixel del que fue liberado, se utilizan campos electrostáticos que evitan que emigre. Los campos son creados con electrodos que se encuentran separados del semiconductor por una capa de material aislante como el óxido de silicio que logra formar un efecto capacitivo. Debido al campo eléctrico generado, se evita que los pares electrón-hueco vuelvan a unirse porque los portadores de carga positiva se repelen en esta región, mientras los portadores de carga negativa son atraídos y almacenados. La capacidad de almacenamiento en esta región, conocida como *pozo de potencial*, puede controlarse regulando la tensión eléctrica que se aplica en el electrodo. La cantidad de electrones almacenados en cada pixel es proporcional a la intensidad luminosa de la imagen en ese punto, de manera que para *ver* la imagen, solo es necesario averiguar cuántos electrones están guardados en cada elemento. En la práctica, los electrodos que forman el campo eléctrico en cada pixel están divididos en grupos de tres y cada uno puede tener un voltaje independiente. Esta estructura de tres fases se usa para transferir información entre pixeles contiguos con el fin de hacer la lectura de todos los elementos del arreglo.

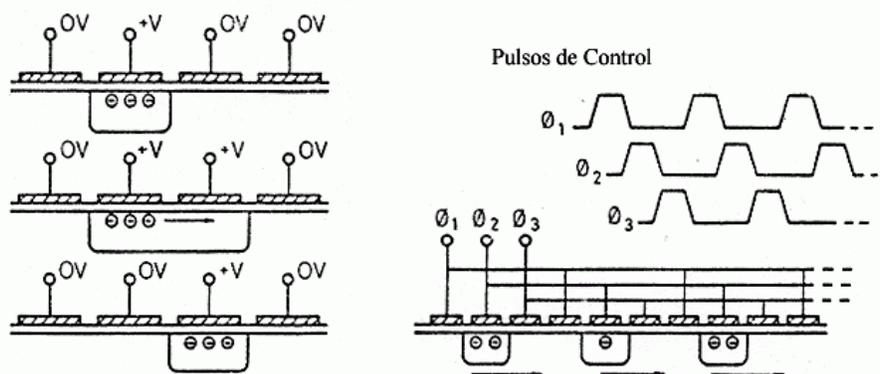


Figura 1.7 Estructura de tres fases del CCD [13]

La lectura de la imagen se hace a través de una salida que amplifica y transmite la información de cada pixel uno a uno, haciendo un barrido por renglones y por columnas. Primero se leen los elementos del primer renglón. Cuando el último ya ha sido leído, todos

los renglones del arreglo se recorren una posición hacia el nivel en donde está el circuito de salida. El proceso se repite hasta que ha sido leído el último elemento del último renglón.

La longitud de onda de los fotones incidentes determina la facilidad de absorción, conforme más cercanos sean los fotones al rojo, mayor será la profundidad que alcancen en el semiconductor antes de ser absorbidos, por esta razón la resolución de los CCD se ve reducida en el rojo. Además, es común hacer que estos dispositivos funcionen en ambientes fríos para disminuir la corriente oscura, pero por otro lado, se produce una extensión del intervalo de energía prohibida, haciendo aún más difícil la absorción en bajas frecuencias.

1.3 Otros instrumentos

1.3.1 Fotómetros

La fotometría es la exposición directa de un instrumento de detección a la energía colectada por el telescopio, con lo que puede medirse la intensidad de la radiación electromagnética. Las cámaras de exposición directa o fotómetros tienen una respuesta espectral limitada y deben calibrarse precisamente. Se usan filtros convencionales con respuestas espectrales estándar. En general, la ventaja de hacer fotometría radica en la observación de fenómenos que requieren tiempos de exposición cortos para poder ser detectados ya que presentarían una imagen saturada en otros detectores.

1.3.2 Interferómetros

Según el principio de superposición, la intensidad de campo eléctrico E resultante en un punto en el espacio en el que dos o más ondas luminosas se traslapan, es igual a la suma de las intensidades individuales. La interferencia óptica resulta de la interacción de más de una ondas de luz. Los interferómetros son dispositivos que producen interferencia entre varias ondas al sobreponerlas para posteriormente proyectarlas sobre una pantalla o un detector. Si la fase relativa de las ondas de luz es diferente para diferentes puntos en la pantalla (o en el detector) se forman franjas de interferencia.

Existen muchos tipos de interferómetros, pero todos basan su funcionamiento en el mismo principio. En la práctica se recurre a la división de un mismo frente de onda para después hacer la superposición de las partes resultantes. En la figura 1.8 se muestra el interferómetro de Michelson en el que un rayo de luz proveniente de una fuente emisora que incide en el divisor O , produce un rayo que se reflejará en el espejo $M1$ y otro en el $M2$. La onda reflejada proveniente de $M2$ atraviesa el separador mientras que la que viene del otro espejo es reflejada y desviada en dirección del detector de manera que ambas ondas son reunidas produciendo interferencia observable en el detector.

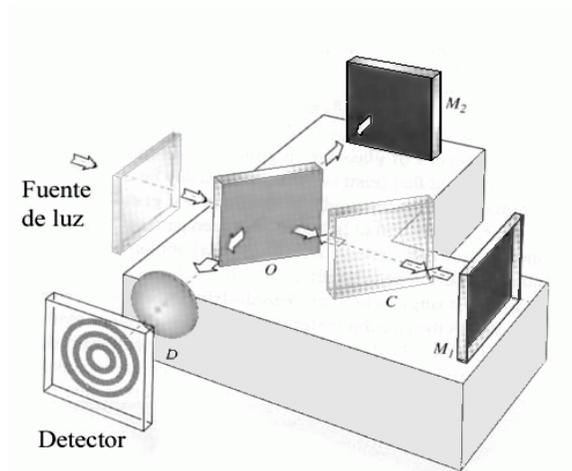


Figura 1.8 Interferómetro de Michelson [7]

Es posible hacer mediciones muy precisas de distancias y ángulos gracias a la interferometría y esto, combinado con las capacidades de las computadoras modernas y el uso de los detectores adecuados da lugar a la obtención de imágenes de alta resolución. Actualmente, el potencial científico más grande es el estudio de estrellas locales y sistemas de estrellas, pero el desarrollo de arreglos con mayor sensibilidad puede abrir nuevas puertas a la investigación, como el estudio de las zonas que rodean a los hoyos negros en los centros de las galaxias.

1.3.3 Espectrómetros

El análisis del espectro de una señal proporciona información importante que no está disponible a "simple vista". La distribución de las componentes en frecuencia de las señales electromagnéticas es el campo de la espectroscopía. Al analizar las componentes de interés, los astrónomos pueden conocer las propiedades físicas de los objetos de estudio, tales como temperatura, luminosidad, composición, masa, etc. En general, el funcionamiento de un espectrómetro puede describirse con base en la figura 1.9. La luz pasa por una ranura de entrada para ser refractada (o reflejada) por un colimador, cuya función es hacer paralelos los rayos de luz antes de que incidan en la rejilla de difracción. Este elemento dispersa los rayos paralelos en sus componentes frecuenciales y envía cada una de estas componentes con un ángulo ligeramente distinto a la salida, de modo que al ser enfocadas por el lente de la cámara, sólo una pueda pasar a través de la segunda ranura. Se hace rotar a la rejilla para que cada frecuencia pase la ranura y su radiación sea registrada.

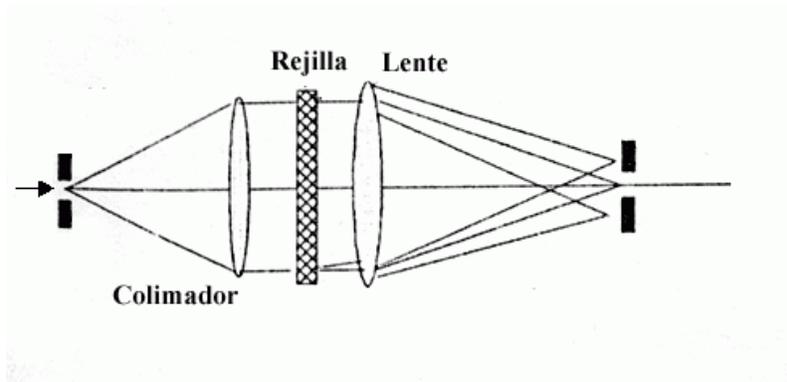


Figura 1.9 Diagrama del espectrómetro [11]

En la espectroscopía no se pretende ver toda la luz de un objeto, sino sólo una región pequeña de su espectro, de hecho, la proyección de esa región es distribuida a lo largo del detector. Esto significa que de la luminosidad efectiva en el detector es mucho menor que la que habría si simplemente se tomaran imágenes del objeto, por lo que se requiere un mayor tiempo de integración para obtener un buen espectro. Así mismo, mientras mayor sea la dispersión de la luz en la rejilla la calidad del espectro será mejor.

Capítulo II. Astronomía infrarroja

2.1 La región infrarroja del espectro electromagnético

Cuando James C. Maxwell propuso la teoría electromagnética, la banda de longitudes de onda entonces conocida del espectro electromagnético se extendía solamente desde el infrarrojo hasta el ultravioleta, lo cual corresponde únicamente a la región visible del espectro, una porción muy pequeña de la que se conoce actualmente, la cual va desde las ondas de radio, en las frecuencias más bajas, hasta los rayos gamma, en frecuencias que rebasan los 10^{22} Hz.

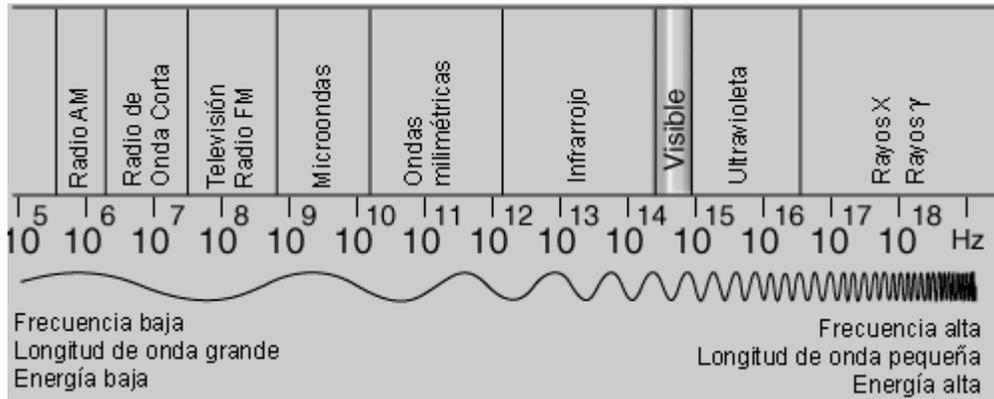


Figura 2.1 Espectro electromagnético

La región infrarroja se extiende aproximadamente entre las longitudes de onda de $1\ \mu\text{m}$ a $300\ \mu\text{m}$. Las moléculas de cualquier objeto a temperaturas mayores al cero absoluto emiten radiación infrarroja, y a medida que la temperatura aumenta la producción de radiación también es mayor. Las moléculas presentan varios tipos de movimientos, algunos de ellos son modos vibracionales con distintos niveles de energía los cuales tienen espectros de emisión y de absorción asociados que, como regla, están incluidos en el intervalo de $1\ \mu\text{m}$ a $300\ \mu\text{m}$. Esto fue caracterizado por Kirchhoff como coeficiente de emisión ϵ_λ y coeficiente de absorción α_λ de los materiales. El primero es la energía emitida por unidad de área por unidad de tiempo en un intervalo pequeño de longitud de onda λ ; el segundo es la fracción de energía absorbida en relación con la incidente en λ . La ley de la radiación de Kirchhoff establece que la energía absorbida por un cuerpo es igual que emite. La relación entre ϵ_λ y α_λ se conoce como *función de distribución de la radiación* o *irradiancia espectral* I_λ , cuya fórmula fue propuesta por Planck:

$$I_\lambda = \frac{2\pi hc^2}{\lambda^5} \left[\frac{1}{e^{\frac{hc}{\lambda T}} - 1} \right] \quad (2.1)$$

donde h es la constante de Planck, c es la velocidad de la luz y k es la constante de Boltzmann, de manera que la irradiancia espectral es una función de la temperatura y de la

longitud de onda de la radiación y no depende del tipo de material, en otras palabras, la radiación en una longitud de onda λ de cualquier material a temperatura T es la misma. Se conoce como *radiación de cuerpo negro* a la radiación de un material que ha absorbido toda la energía incidente. Con la fórmula de Planck puede generarse una familia de curvas que describen la radiación de cuerpo negro en función de la longitud de onda, para distintas temperaturas.

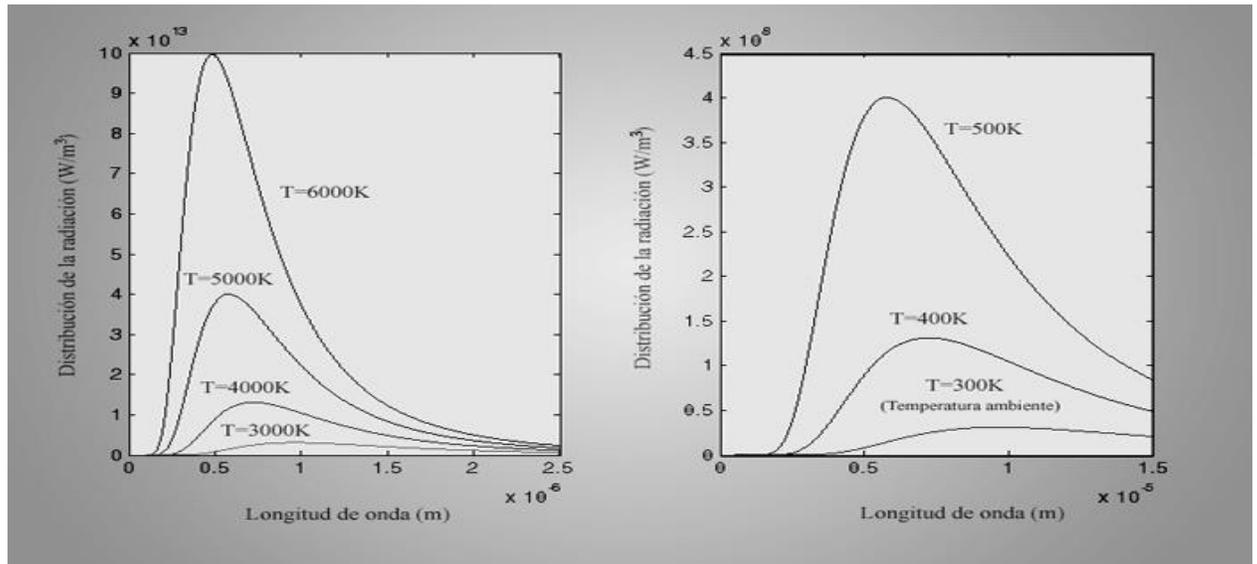


Figura 2.2 Curvas de radiación de cuerpo negro

En la segunda gráfica se muestra la curva correspondiente a la temperatura ambiente, que tiene un máximo alrededor de $\lambda = 10 \mu\text{m}$, lo cual significa que es en esa longitud de onda a la que un cuerpo negro irradia la mayor parte de su energía. Esto es de particular interés para la astronomía infrarroja ya que todas las observaciones se verán contaminadas por la radiación de todos los objetos que rodeen al detector y al telescopio. Las imágenes obtenidas en longitudes de onda cercanas a los $10 \mu\text{m}$ presentan una señal de interés muy tenue en comparación con la cantidad de ruido en las que están sumergidas, resultando en una calidad de imagen muy pobre, por lo que tanto los instrumentos de detección utilizados para astronomía infrarroja así como el tratamiento de las observaciones difieren notablemente de los empleados en el visible.

Frecuentemente se establecen subdivisiones de la región infrarroja en tres intervalos: infrarrojo cercano ($1 \mu\text{m}$ a $5 \mu\text{m}$); infrarrojo medio ($5 \mu\text{m}$ a $30 \mu\text{m}$); e infrarrojo lejano ($30 \mu\text{m}$ a $300 \mu\text{m}$). Esta división es arbitraria y carece de nomenclatura universal, sin embargo es útil, por ejemplo, en la Astronomía se distinguen objetos y fenómenos observables en cada intervalo como se muestra en la siguiente tabla.

Región espectral	Longitud de onda (μm)	Temperatura (K)	Ejemplos observables
Infrarrojo cercano	1 – 5	4000 – 750	Gigantes rojas, Estrellas rojas de menor temperatura. El polvo es transparente.
Infrarrojo medio	5 – 30	750 – 100	Planetas, cometas, asteroides,

			discos protoplanetarios, polvo calentado por luz estelar.
Infrarrojo lejano	30 - 300	100 - 15	Emisiones de polvo frío, regiones centrales de galaxias, nubes moleculares de baja temperatura.

2.1.1 Ventanas atmosféricas y emisión térmica

Algunas moléculas componentes de la atmósfera como el vapor de agua y el CO₂ absorben la radiación infrarroja en ciertos intervalos de longitud de onda. El resultado de este fenómeno se muestra en la siguiente figura, en la que se observan zonas del espectro en las que la transmisión de energía es nula, en contraste con las zonas de transparencia o *ventanas*, que tienen buena transmisión. La presencia de vapor de agua en la atmósfera disminuye conforme la altitud aumenta, de manera que resulta conveniente que los telescopios infrarrojos se localicen en sitios altos y especialmente secos.

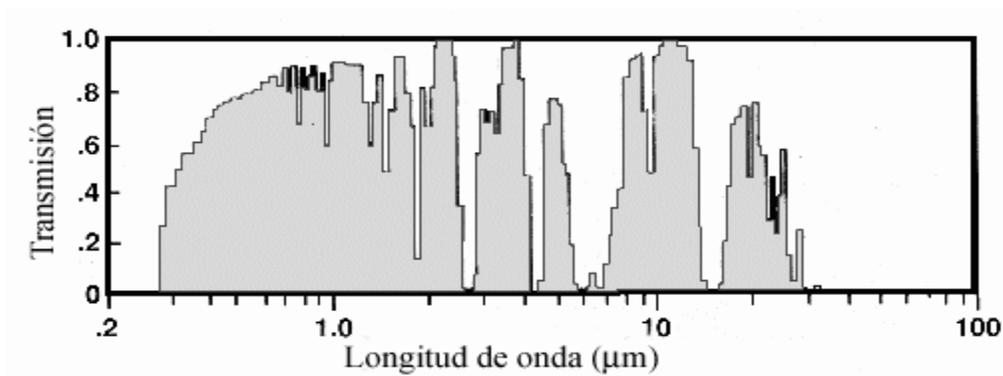


Figura 2.3 Transmisión atmosférica

Otro fenómeno que afecta a la Astronomía infrarroja es la emisión de radiación de cuerpo negro en los intervalos de opacidad de la atmósfera. En consecuencia, no sólo resulta imposible realizar observaciones en los intervalos opacos, sino que contribuyen con una cantidad importante de ruido en el proceso de adquisición de imágenes. En la gráfica siguiente puede observarse un incremento exponencial de la intensidad de la radiación emitida en un intervalo del espectro en el que inicia una zona opaca.

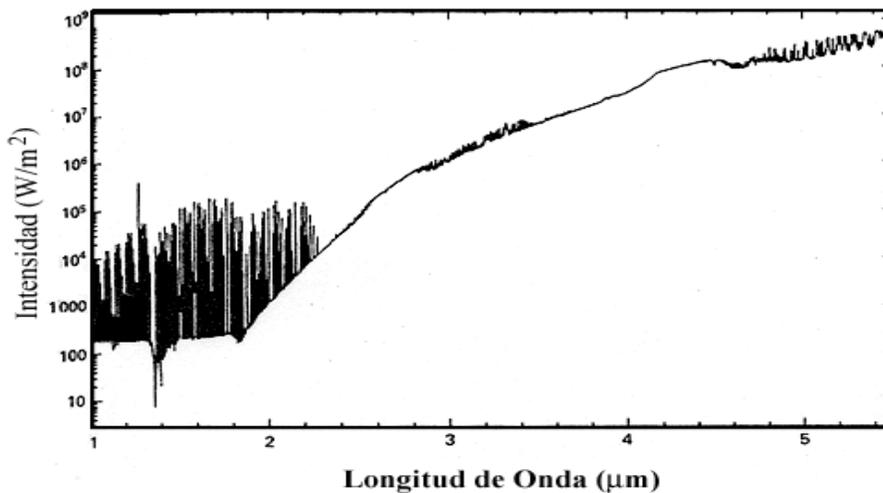


Figura 2.4 Emisión atmosférica [14]

2.2 Telescopios infrarrojos

En los telescopios infrarrojos, tanto los métodos de detección como la función del espejo primario como superficie recolectora de luz, difieren de los usados en los telescopios ópticos. En primer lugar se buscan diseños ópticos totalmente reflectores para eliminar algunos efectos no deseados, como el incremento en la absorción que presentan algunos materiales en este intervalo de longitudes de onda, y un índice de refracción grande en la unión aire-cristal. Adicionalmente, la capacidad de reflexión de algunos metales es mayor en la región espectral del infrarrojo, de modo que resulta conveniente incorporarlos en estos telescopios. Sin embargo, es el control de temperatura de cada parte del telescopio infrarrojo un aspecto crítico en su diseño con el fin de minimizar el ruido térmico generado en su interior. Mientras en el caso del visible las fuentes de ruido son objetos localizados en el campo de visión, en el infrarrojo cualquier objeto fuera de campo puede contribuir con ruido en el proceso de detección, por eso se colocan baffles que bloquean la radiación proveniente de fuentes de ruido.

Una de las soluciones comunes para la reducción de ruido en los telescopios infrarrojos es el movimiento del secundario, un proceso que se conoce como *chopping*. Oscila rápidamente de una posición a otra, en sincronía con las exposiciones que toma el detector para intercalar imágenes del objeto de interés con imágenes de cielo y hacer una comparación entre ellas para obtener una primera disminución del ruido. El secundario es de menor tamaño que en los telescopios para longitudes de onda del visible, ya que debe tener una frecuencia de oscilación de entre 10 y 20 Hz. Es conveniente que las muestras de cielo sin objetos a la vista sean en varios puntos alrededor del objeto de interés para minimizar cualquier sesgo o gradiente en la intensidad del ruido. A la imagen resultante de la diferencia entre la que incluye al objeto de interés y la que no lo incluye se le llama *imagen cosumada*.

2.3 Detectores de infrarrojo

Después del éxito de los CCD, el desarrollo de un arreglo infrarrojo representaba una nueva meta de la instrumentación astronómica motivada por la importancia de la información que puede obtenerse a partir de observaciones en las bandas infrarrojas del espectro. La radiación infrarroja es más penetrante que la visible y la ultravioleta, por lo que puede revelar procesos difícilmente observables en ellas, como regiones de formación de estrellas inmersas en nubes de polvo y otros ejemplos como los mencionados anteriormente. Además de que, debido a la expansión del universo demostrada por Hubble, la radiación proveniente de objetos lejanos sufre un corrimiento hacia frecuencias menores, y en el caso de los objetos más lejanos, el corrimiento alcanza al infrarrojo.

2.3.1 Historia

Los primeros detectores infrarrojos modernos fueron desarrollados durante la Segunda Guerra Mundial, cuando se descubrió que el PbS es un material fotoconductor con una respuesta alrededor de los 3 μm . La producción de detectores de PbS y de Ti_2S tuvo resultados exitosos, motivando que el desarrollo de esta tecnología continuara. En la década de los cincuenta se construyeron detectores policristalinos Pb:Se y Pb:Te para buscadores antimisiles, pero sus métodos de fabricación nunca fueron totalmente dominados. El descubrimiento del transistor estimuló el progreso de las técnicas de purificación de materiales y dio lugar a los primeros detectores extrínsecos basados en germanio. La respuesta fotoconectiva de impurezas de cobre, oro y zinc en germanio hizo posible detectar radiación de hasta 30 μm . Estos deben operarse a temperaturas bajas para que alcancen los niveles de rendimiento de los detectores intrínsecos, además de que es necesario sacrificar eficiencia cuántica con el fin de evitar que los detectores sean demasiado gruesos. Paralelo a los detectores de germanio fueron desarrollados materiales como el In:Sb, cuyas características principales son la facilidad de fabricación en un cristal único y la reducción de la banda prohibida, con lo que se mejora notablemente la sensibilidad. El In:Sb es el primer material de la familia de semiconductores III-V. Siguió aleaciones de materiales III-V, IV-VI y II-VI para aplicaciones específicas con base en la manipulación de la extensión de la banda prohibida. En 1959 Lawson desarrolló aleaciones de Hg:Cd:Te que ofrecían ventajas sin precedentes en virtud de un intervalo prohibido variable, sin embargo su producción en masa presentaba problemas tecnológicos importantes. Una década después los detectores de Pb:Sn:Te representaban la mayor competencia para los de Hg:Cd:Te, ya que su preparación era más sencilla y su funcionamiento parecía más estable, pero con la inconveniencia de tener coeficientes de expansión térmica muy altos que hacen imposible su aplicación en sistemas híbridos. A finales de los sesenta Soref publicó un artículo de detectores extrínsecos de silicio, semejantes a los de germanio pero con las ventajas de tener mayor solubilidad de

impurezas, menor índice de refracción, una sección de fotoionización más amplia para mejor eficiencia cuántica y menor constante dieléctrica para reducir la capacitancia interna. Shepherd y Yang propusieron en 1973 el detector de barrera Shottky, de metal y silicio, cuyos esquemas de lectura podían ser implementados junto a los de detección en un chip común de silicio en configuraciones de dos dimensiones.

El diseño de circuitos de lectura basados en transistores de efecto campo (FET) impulsó la popularidad de los detectores cuyas altas impedancias internas resultaban adecuadas en interfaces con tales circuitos. Este grupo de detectores incluye a los de barrera Shottky de Pt:Si, los fotodiodos de In:Sb y Hg:Cd:Te, los fotoconductores de Pb:Se y PbS y los extrínsecos de silicio. Durante la década de los ochenta el desarrollo de la tecnología de detectores fotovoltaicos de Hg:Cd:Te alcanzó su punto más alto con arreglos de dos dimensiones con millares de elementos que, a pesar de su baja impedancia, gracias a la incorporación de circuitos de integración y de retraso de tiempo presentaron niveles altos de rendimiento. En los últimos años ha sido posible detectar radiación de frecuencias muy bajas con arreglos de pozos cuánticos múltiples de Ga:As y Ga:Al:As que han aprovechado una configuración muy novedosa conocida como arreglo de plano focal, con la que se espera obtener ventajas en la uniformidad y en el control, así como un incremento importante en el número de elementos en los detectores.

2.3.2 Clasificación

En función de los procesos de interacción de la radiación con el material semiconductor, pueden distinguirse cuatro tipos principales de detectores de fotones:

- ⇒ intrínsecos,
- ⇒ extrínsecos,
- ⇒ fotoemisores (barrera Shottky) y
- ⇒ de pozo cuántico.

Los detectores intrínsecos fotoconductores se componen de un material resistivo sensible a la radiación. Un fotón cuya energía $h\nu$ es mayor que el intervalo prohibido de energía E_g del material es absorbido para producir un par electrón-hueco, provocando un cambio en la conductividad. Un par de electrodos conectados a una fuente de energía se colocan a ambos lados del semiconductor para medir la variación de la corriente eléctrica con ayuda de una resistencia sensor.

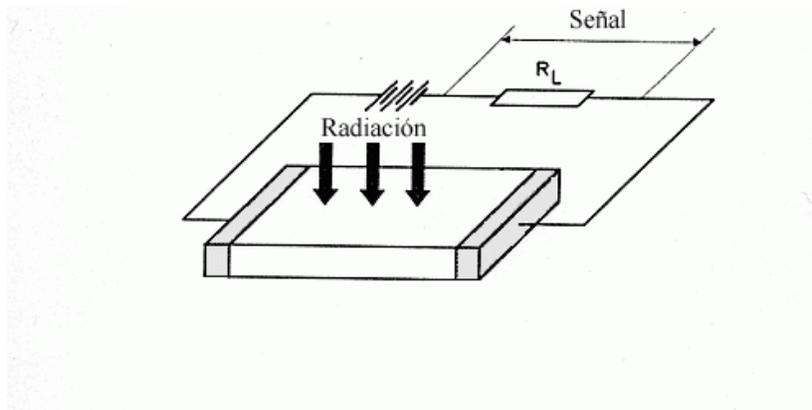


Figura 2.5 Geometría de un detector intrínseco [18]

En los detectores extrínsecos, a diferencia de los anteriores, cuando un fotón de energía $h\nu$ incide, se crea un portador de carga cuya carga complementaria reside en un átomo ionizado que permanece como parte de la estructura del cristal, por lo que permanece inmóvil y no puede contribuir con corriente eléctrica, de manera similar a lo que ocurre en un CCD.

La respuesta de los detectores presenta una frecuencia de corte que, como se ha mencionado, está en función de la amplitud de la banda de energía prohibida y ésta, a su vez, depende de la cantidad de impurezas en el semiconductor. Puede pensarse en incrementar indefinidamente el número de impurezas con el fin de mejorar la eficiencia cuántica, sin embargo, esto puede afectar las características eléctricas del material que dan lugar a modos de operación que no son deseables. Un ejemplo son saltos de electrones o *hopping*. Cuando dos impurezas tipo n están lo suficientemente cercanas, las funciones de onda de sus electrones libres se traslapan, lo cual permite la conducción sin necesidad de que los electrones alcancen la banda de valencia. Según lo anterior se establece un límite máximo de impurezas de alrededor de 10^{15} cm^{-3} . Además, el resultado de utilizar un gran número de impurezas es un detector voluminoso.

Los detectores de barrera Shottky están constituidos de una unión semiconductor-metal. Si el semiconductor es tipo n los electrones emigran al metal dejando una región libre de carga en el semiconductor, con lo que los niveles de Fermi del metal y del semiconductor se igualan y se establece una barrera de potencial ϕ que puede ser usada para la detección de fotones cuya energía es menor a la de la banda de energía prohibida del semiconductor. Esto se conoce como *fotoemisión interna*. Cuando la energía de un fotón es absorbida por un electrón, éste vaga aleatoriamente sobre la superficie del metal hasta que encuentra la unión con el semiconductor para ser emitido sobre la banda de valencia.

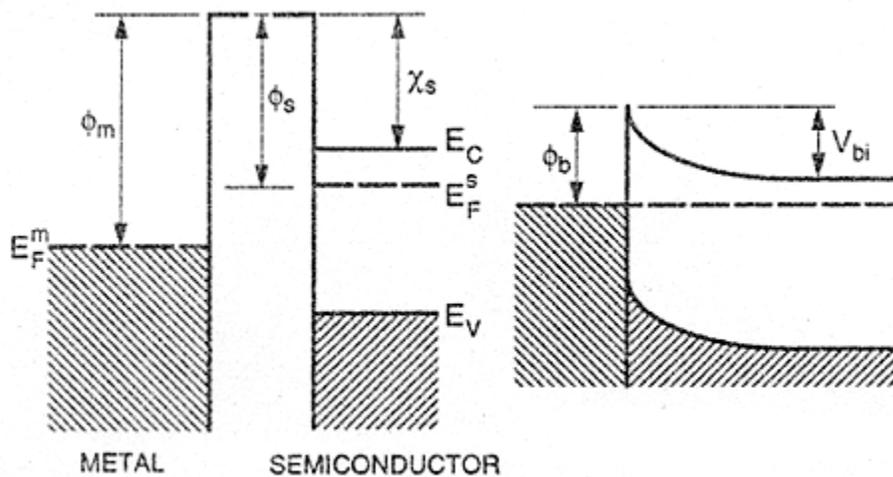


Figura 2.6 Detector de barrera Shottky a) sin contacto; b) con contacto [18]

El principio de funcionamiento de los detectores de pozo cuántico no difiere mucho de los de barrera Shottky. En este caso, la unión de dos semiconductores distintos provoca que las bandas de energía prohibida, de valencia y de conducción no sean homogéneas. Si se acomodan varias capas de dos semiconductores distintos como Ga:As y Al:Ga:As el diagrama de bandas queda de la siguiente manera.

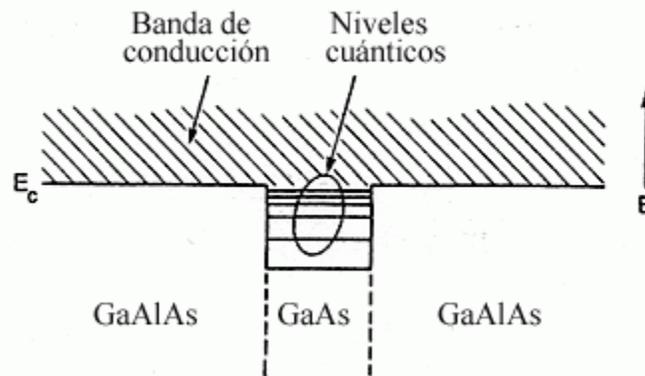


Figura 2.7 Formación de pozos cuánticos [17]

En las regiones correspondientes al Ga:As se forman pozos cuánticos en los que los niveles de energía están bien definidos y pueden ser ocupados. Los electrones que alcancen la banda de conducción pueden encontrar un pozo y quedar atrapados en él. Con el fin de producir una corriente eléctrica aprovechando estos electrones, se procura que cada nivel de energía de un pozo corresponda con el nivel de energía superior del pozo contiguo, de manera de que por medio de efecto túnel un electrón puede pasar de un pozo a otro sucesivamente. Además, con el fin de disponer de una gran concentración de electrones de

conducción que puedan tener interacción con los fotones, el Ga:As es dopado hasta que el nivel de Fermi casi coincide con el fondo de los pozos.

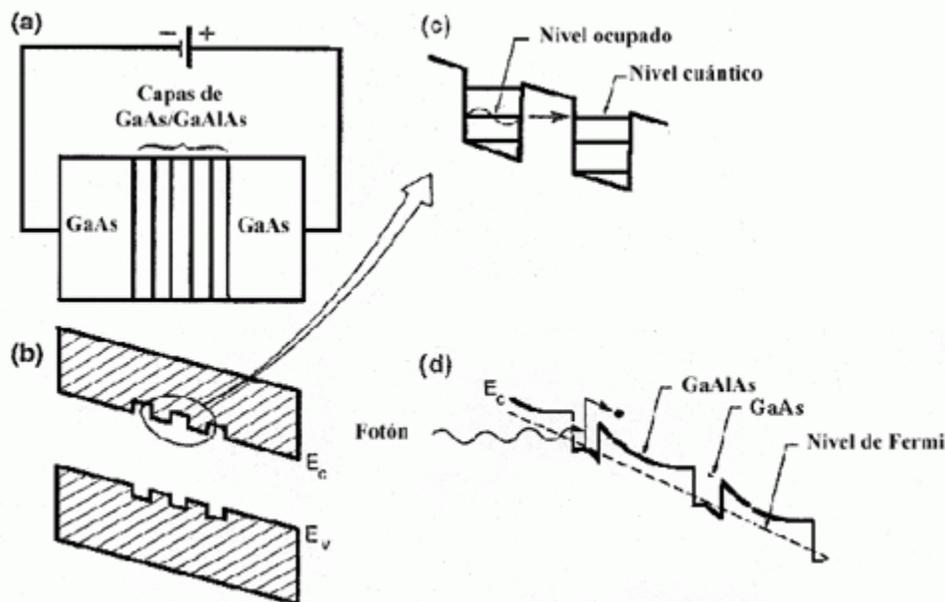


Figura 2.8 Operación de un detector de pozos cuánticos [17]

2.3.3 Detectores de conducción en la banda de impurezas (IBC)

También conocidos como detectores de Banda de Impurezas Bloqueada (BIB), los detectores IBC son un caso particular de los detectores extrínsecos. Como se mencionó en la sección 2.3.2, tanto su eficiencia cuántica como su frecuencia de corte están limitadas por la cantidad de impurezas con que puede doparse al semiconductor, ya que por un lado puede generarse *hopping* si las impurezas están muy cercanas entre sí, y por otro lado el grosor del detector puede alcanzar centenas de micras, lo cual acarrea problemas operacionales. Los detectores IBC superan ambas desventajas.

El semiconductor que conformará la región reactiva al infrarrojo es dopado hasta con 5×10^{17} átomos de impureza por cada centímetro cúbico, una concentración que rebasa ampliamente el límite máximo conveniente, sin embargo, la corriente oscura debida al *hopping* se previene con una capa de bloqueo constituida de un material semiconductor puro con impedancia muy alta. Con la capa de bloqueo no sólo se incrementará la eficiencia en virtud de la gran concentración de impurezas, sino que será posible la detección de radiación de menor energía, además el grosor del dispositivo se ve reducido en más de un orden de magnitud. Usualmente se construyen detectores que son iluminados por detrás, a través de un sustrato transparente separado de la región activa por un contacto metálico. Un segundo contacto metálico se coloca, aislado por un óxido, sobre la capa de bloqueo. La tensión V_{bias} aplicada en estos contactos provoca una migración de cargas positivas o huecos hacia el contacto negativo, creando una región libre de carga en la cual se generarán los pares electrón-hueco a partir de los fotones incidentes. Esta es otra ventaja de los

detectores IBC sobre los detectores extrínsecos convencionales, ya que el electrón liberado, al estar en una zona sin cargas positivas cercanas, no corre peligro de recombinación, lo cual es un fenómeno común en otros dispositivos que además constituye una fuente de ruido. Según lo anterior, puede concluirse que la eficiencia está en función de la extensión de la zona libre de cargas y, por lo tanto, en función de la tensión V_{bias} aplicada, sin embargo, si ésta es muy grande se registrará una ganancia $G > 1$ en el proceso de detección, es decir, un fotón absorbido producirá más de un electrón libre [17], lo cual es poco deseable si se toma en cuenta que la relación señal a ruido S/N también será amplificada. Es necesario aclarar que las imágenes en el infrarrojo medio suelen tener una $S/N < 1$.

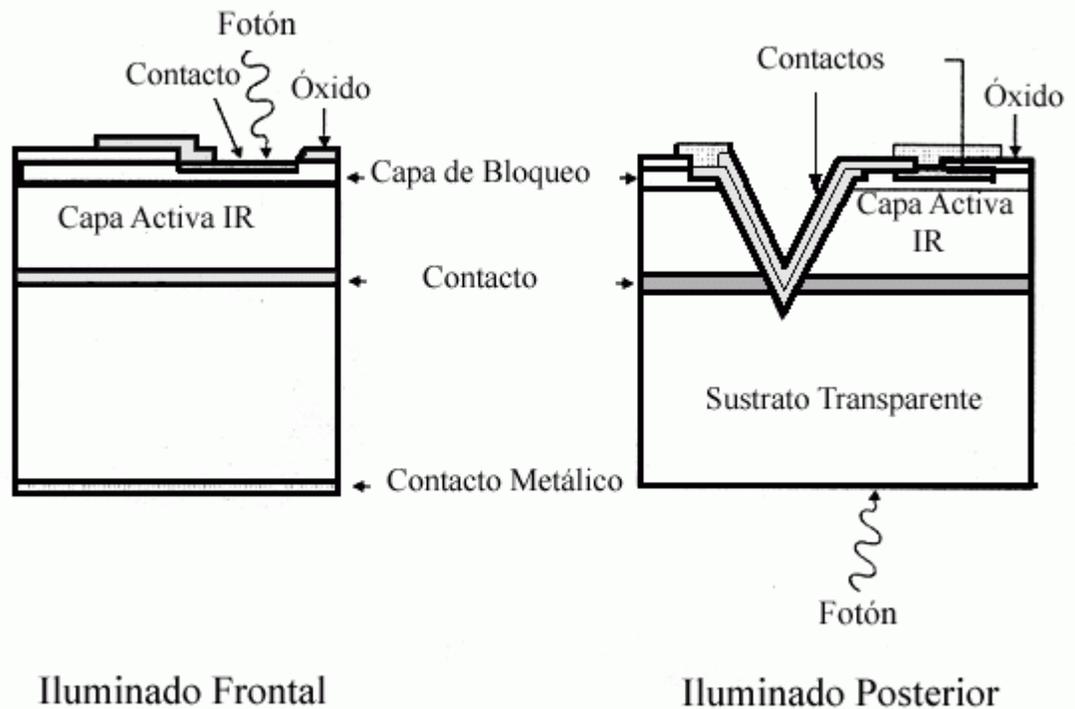


Figura 2.9 Estructura de un detector IBC [17]

El funcionamiento del detector se describe brevemente a continuación. Por excitación térmica, los electrones de la banda de impurezas tienden a saltar hacia la capa de bloqueo pero no la alcanzan si no tienen la energía suficiente. Solamente pueden adquirirla cuando un fotón es absorbido en la región activa. Una vez en la banda de conducción, el electrón puede pasar libremente hacia la banda de bloqueo porque las bandas de conducción de las regiones activa y de bloqueo son continuas.

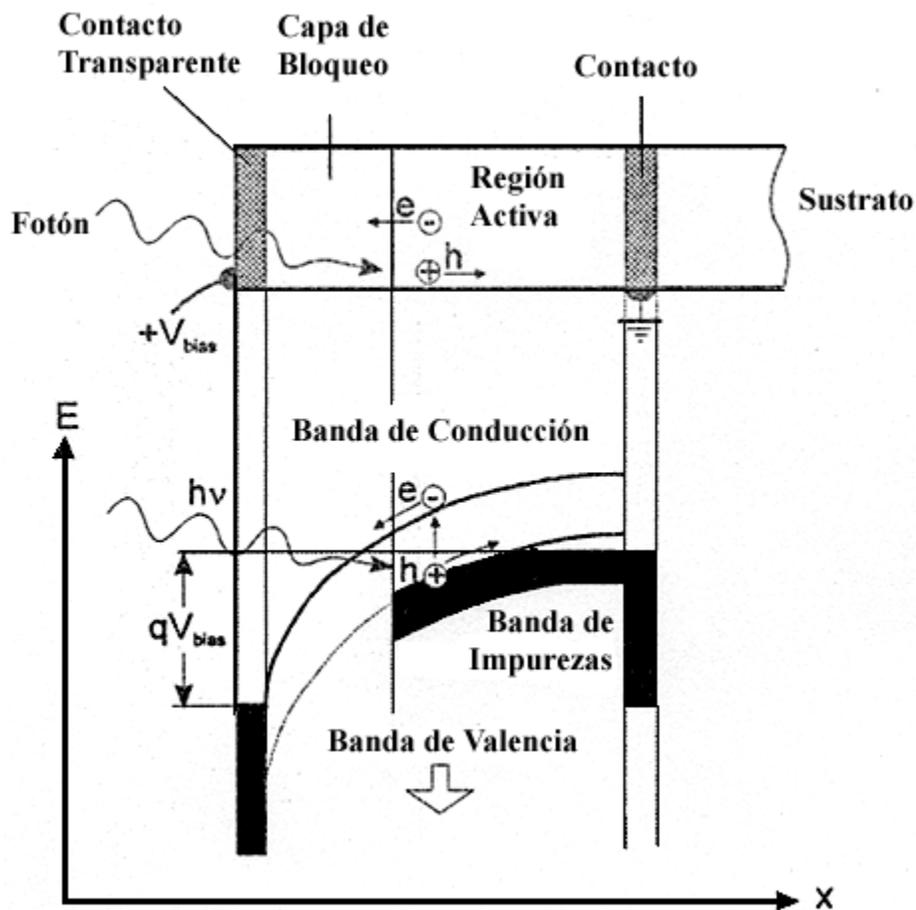


Figura 2.10 Esquema de funcionamiento de un detector IBC [18]

2.3.4 Arreglos infrarrojos

Aún cuando la historia de los detectores infrarrojos modernos se remonta a los años de la posguerra, fue hasta la década de los setenta cuando los primeros arreglos infrarrojos fueron desarrollados. Los primeros formatos eran muy pequeños, la mayoría lineales de 1 x 8 elementos, pero no dejaban de ser novedosos, lo cual impulsó su desarrollo hasta que en la década siguiente se lograron arreglos más robustos de 32 x 32 y 32 x 64 elementos, entre otras configuraciones. En la actualidad, en lo que se refiere a los detectores de banda de impurezas, los fabricantes Rockwell (BIB) y Hughes (IBC) han integrado arreglos de 256 x 256 y 240 x 320 pixeles. Es gracias a los últimos detectores desarrollados que la espectroscopía infrarroja se hizo posible.

Con la integración de los sensores infrarrojos y la electrónica de control y de lectura en un mismo circuito se pretende que el arreglo realice las mismas funciones básicas de un CCD: almacenamiento de las cargas, transferencia de la carga de cada pixel a un circuito de salida y su extracción en forma secuencial. El circuito completo es una estructura híbrida compuesta por dos placas principales. La primera es en donde se lleva a cabo la detección.

Es una malla de pixeles independientes con una pequeña separación entre sí en la que, en el caso de los detectores IBC, tiene la estructura de la figura 2.11. La segunda placa es un arreglo de silicio que realiza la función de lectura. La conexión entre ambas placas es por medio de una rejilla de elementos independientes de indio contenidos en una sustancia que conserva la consistencia de la estructura en los cambios de temperatura. El almacenamiento de carga puede hacerse en la capacitancia de la unión del sensor infrarrojo o en un capacitor alojado en el circuito de silicio. A la lectura registrada en cada pixel de la placa de silicio puede accederse independientemente o por grupos, en cuyo caso es necesario un circuito multiplexor. La estructura ensamblada se muestra en la siguiente figura.

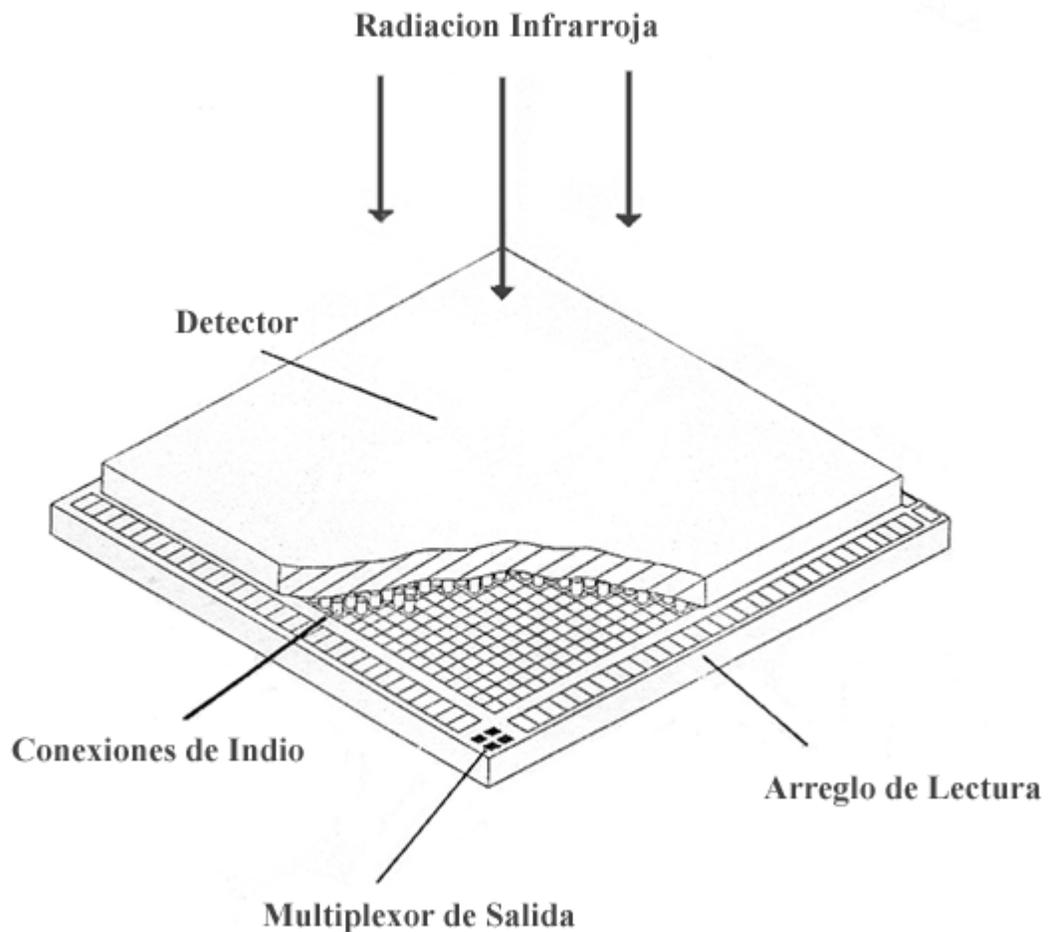


Figura 2.11 Estructura de un arreglo infrarrojo [14]

2.5 La cámara-espectrógrafo de infrarrojo medio

El Instituto de Astronomía ha adquirido un detector IBC para construir una cámara-espectrógrafo que se acoplará al telescopio en el observatorio de San Pedro Mártir. Las características más importantes del detector se presentan en la siguiente tabla.

Detector IBC	
Respuesta espectral	2 - 28 μm
Formato	320 x 240 pixeles
Tiro del pixel	50 μm
Eficiencia cuántica	40 - 50 %
Material	Si:As
Temperatura de operación	4 - 10 K
Uniformidad de respuesta	< 10 %
Corriente oscura	100 e^-/s
Capacidad de pozo	$1.4 \times 10^7 \text{e}^-$ ó $4.5 \times 10^7 \text{e}^-$
Ruido de lectura	130 e^- ó 340 e^-

El detector será colocado en el interior de un criostato enfriado con helio líquido y será la parte final de un sistema óptico conformado por un carrusel con filtros, un colimador, una rejilla y un espejo parabólico. La rejilla y el espejo compartirán un sistema de rotación, con el que será posible elegir una u otro, para hacer espectroscopía o fotometría. Un conjunto de motores capaces de operar a ultrabajas temperaturas proveerá de movimiento a todos los componentes. El funcionamiento adecuado del detector depende de una gran cantidad de señales de control generadas con un DSP, que incluyen el control del proceso de lectura. El detector está dividido en 32 regiones, cada una con una salida de lectura independiente que, después de ser muestreada y amplificada, entra en un multiplexor de 8 canales que cuentan con convertidores A/D de 16 bits. Finalmente, un DSP, cuya función estará enfocada al procesamiento de la imagen, la transmitirá a una estación de trabajo en la que será almacenada y tratada.

2.5.1 Fuentes de ruido

En la imagen final, las fuentes de ruido pueden clasificarse en dos categorías principales. En primer lugar está el ruido fotoelectrónico debido a la naturaleza de la radiación incidente, expuesta con anterioridad en este capítulo. La segunda fuente es el ruido térmico introducido por los dispositivos electrónicos en la lectura de la señal y usualmente es modelado con una distribución de Gauss. El ruido fotoelectrónico también se modela como una variable aleatoria, pero en este caso se utiliza la función de densidad de Poisson [1], en la que la desviación estándar de la señal es igual a la raíz cuadrada de la media. Esto implica que la amplitud del ruido es dependiente de la señal resultante del proceso de detección. En el caso de los detectores IBC, como se mencionó en la sección 2.3.3, la ganancia que presentan está en función de la tensión V_{bias} que se les aplica, y una tensión elevada da lugar a una S/N elevada, de tal suerte que es necesario alimentar al dispositivo con una tensión adecuada para que el ruido efectivamente pueda ser modelado con base en

una distribución de Poisson. En el caso de un detector BIB, con $V_{bias} = 0.5 V$, se obtuvo que el ruido es proporcional a la raíz cuadrada de la señal [24].

Capítulo III. Marco teórico

3.1 Generalidades sobre imágenes digitales

Una *imagen* es una función bidimensional de la intensidad de la luz $f(x_1, x_2)$, donde el valor de f es el valor de la intensidad o brillo en el punto de coordenadas espaciales (x_1, x_2) . Dado que la luz es una forma de energía, f debe ser una cantidad positiva y finita. Nótese que la definición anterior es para una imagen monocromática. Al intervalo $[L_{\min}, L_{\max}]$, en el que L_{\min} es el valor mínimo de $f(x_1, x_2)$ y L_{\max} es el valor máximo, se le conoce como *escala de grises*, en el cual, usualmente L_{\min} es cero y representa al negro, mientras que L_{\max} es el blanco y los valores intermedios son tonalidades de grises que varían desde el negro hasta el blanco. En las imágenes digitales estos valores son discretos, es decir, existe un número finito de valores que $f(x_1, x_2)$ puede adquirir. La digitalización de una imagen toma en cuenta tanto las coordenadas espaciales (*muestreo de la imagen*) como la amplitud de la función f (*cuantificación de los niveles de grises*), con lo que se forma un arreglo de números reales de dos dimensiones con $N \times M$ elementos, o píxeles:

$$f(x_1, x_2) = \begin{bmatrix} f(0,0) & f(0,1) & \dots & f(0, M-1) \\ f(1,0) & f(1,1) & & f(1, M-1) \\ \vdots & \vdots & & \vdots \\ f(N-1,0) & f(N-1,1) & \dots & f(N-1, M-1) \end{bmatrix} \quad (3.1)$$

Es común que tanto N y M como la cantidad de niveles de grises G sean potencias de 2. La resolución de la imagen depende de estos parámetros, ya que conforme N , M y G sean más grandes, la imagen digital se aproximará cada vez más a la imagen real, sin embargo, el número de bits b necesarios para almacenar estos datos también crece:

$$b = M \times N \times \log_2 G \quad (3.2)$$

En el caso de la cámara-espectrógrafo de infrarrojo medio, la imagen obtenida por el detector IBC es de 320x240 píxeles, y el convertidor analógico digital es de 16 bits para 65,536 niveles de gris, con lo que el ruido de cuantificación resulta despreciable.

3.1.1 Restauración de imágenes

Los problemas básicos en el procesamiento de imágenes digitales son los siguientes:

- ⇒ representación y modelado (caracterización de la información de cada pixel y su modelado matemático, generalmente por medio de series ortogonales);

- ⇒ mejoramiento (resaltar determinadas características de la imagen utilizando algoritmos iterativos);
- ⇒ restauración (minimizar degradaciones conocidas en imágenes difusas, con ruido y con errores geométricos);
- ⇒ análisis (mediciones cuantitativas de la imagen para obtener información);
- ⇒ reconstrucción (rescate de objetos de dos dimensiones a partir de sus proyecciones) y
- ⇒ compresión de datos (para facilitar la transacción de datos procurando que las pérdidas de información no sean apreciables).

El objetivo de este trabajo se centra en uno de los puntos relacionados con la restauración de imágenes: la reducción de ruido. Las técnicas de restauración están orientadas al modelado de la degradación para, de esta manera, diseñar el proceso inverso que ayude a recuperar la imagen más cercana posible a la original. El siguiente es un modelo que representa el proceso de degradación.

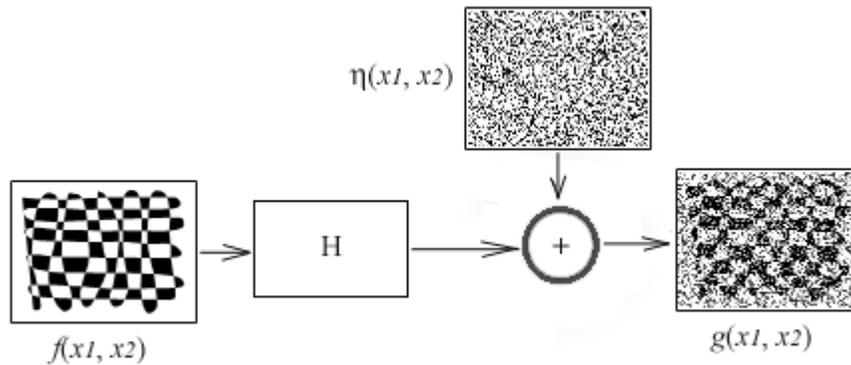


Figura 3.1 Degradación de una imagen.

$$g(x_1, x_2) = H[f(x_1, x_2)] + \eta(x_1, x_2) \quad (3.3)$$

La ecuación anterior expresa la relación entrada-salida del sistema. Si se considera que H es un operador lineal invariante en la posición y con la definición de la función impulso, de (3.3) puede derivarse la *ecuación general de restauración* que establece que si se conoce la respuesta de H a un impulso unitario en las coordenadas (α, β) , la respuesta a cualquier entrada $f(\alpha, \beta)$ puede calcularse [6].

$$g(x_1, x_2) = \int \int_{-\infty-\infty}^{\infty \infty} f(\alpha, \beta) h(x_1 - \alpha, x_2 - \beta) d\alpha d\beta \quad (3.4)$$

La versión discreta se construye tomando en cuenta que las imágenes $f(x_1, x_2)$ y $h(x_1, x_2)$ se completan con ceros para formar matrices de $M \times N$ elementos.

$$g_e(x_1, x_2) = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} f_e(m, n) h_e(x_1 - m, x_2 - n) \quad (3.5)$$

Dentro de los métodos tradicionales existen tanto los que utilizan conceptos en el dominio de la frecuencia como los que utilizan conceptos en el dominio del tiempo. En general, los primeros ofrecen ventajas en cuanto a una baja complejidad de procesamiento, sin embargo, es el método de Wiener, de dominio temporal, el que da el mejor estimado del objeto de observación. La siguiente es la ecuación para implementar el filtro de mínimos cuadrados de Wiener, con la que puede obtenerse un estimado de la imagen.

$$\hat{F}(u, v) \approx \left[\frac{1}{H(u, v)} \frac{|H(u, v)|^2}{|H(u, v)|^2 + K} \right] G(u, v) \quad (3.6)$$

3.2 Transformación *wavelet*

Las transformaciones matemáticas se aplican a las señales para obtener de ellas más información de la que puede disponerse cuando solamente se tiene la señal original. En la práctica, es común encontrar señales en el dominio del tiempo. En el caso de las imágenes, el dominio es el espacio y las señales tienen dos dimensiones, pero con el fin de presentar una explicación más sencilla y más clara, en el desarrollo de este tema se hará uso de señales temporales unidimensionales y posteriormente los resultados serán extrapolados a los arreglos bidimensionales. Para facilitar la explicación del uso de las funciones *wavelets*, se empezará con una revisión del análisis de Fourier, que es más conocido.

3.2.1 Análisis de Fourier

La transformación de Fourier (FT) es la más popular en el análisis de señales. Se trata de una transformación reversible que traslada una función definida en el dominio del tiempo, $f(t)$, a una función en el dominio de la frecuencia $F(\omega)$ cuya base se conforma de exponenciales complejas (senos y cosenos), en otras palabras, es posible representar a la señal original a través de combinaciones lineales de los elementos de esta base. De esta manera, la función puede analizarse espectralmente porque los coeficientes de la función transformada representan la contribución de los senos y los cosenos en cada frecuencia.

$$F(\omega) = \int_{-\infty}^{\infty} f(t) e^{-j\omega t} dt \quad (3.7)$$

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\omega) e^{j\omega t} d\omega \quad (3.8)$$

La información que proporciona la FT es independiente del tiempo, es decir, no dice en qué momento contribuye cada frecuencia, de ahí que el campo de acción de la FT se vea limitado solamente a señales estacionarias. Para resolver este problema, puede asumirse que pequeños intervalos de una señal no estacionaria son estacionarios, lo que hace posible el análisis de Fourier de la señal dividida en segmentos. Para esto es necesaria una *función ventana* cuya extensión debe ser igual o menor al tamaño de los intervalos que se

consideran estacionarios. La ventana se localiza inicialmente en $t = 0$, se multiplica por la función original y se calcula la FT del segmento resultante. El proceso se repite a lo largo del tiempo de análisis de toda la señal. El resultado es un *espectrograma*, una representación de las componentes en frecuencia a través del tiempo. Este procedimiento lleva a la definición de la transformación de Fourier de tiempo limitado (STFT), cuya expresión matemática es la siguiente:

$$STFT(u, \omega) = \int_{-\infty}^{\infty} f(t) g(t-u) e^{-j\omega t} dt \quad (3.9)$$

La desventaja principal de la STFT proviene del *principio de incertidumbre de Heisenberg*. Este principio, aplicado originalmente a la determinación del momento y de la localización espacial de partículas elementales en movimiento, puede aplicarse también a la información en tiempo y frecuencia de una señal. El principio establece que no puede saberse con exactitud el instante en el que existe cada componente espectral. En cambio, pueden determinarse intervalos de tiempo en los que se presentan ciertas bandas de frecuencia, lo cual es, en el fondo, un problema de cantidad de resolución, que está relacionado con la extensión de la ventana temporal utilizada en la ecuación anterior, técnicamente conocida como *soporte*. Una ventana angosta, o *soporte compacto*, permite que la resolución sea buena en el tiempo, pero mala en la frecuencia, y viceversa. La figura 3.2 presenta tres espectrogramas para una misma función (3.10), donde cada uno ha sido calculado con una ventana de extensión distinta, según la ecuación (3.11).

$$f(t) = \begin{cases} \text{sen } 2\pi 50t; & 0 \leq t < 300 \text{ ms} \\ \text{sen } 2\pi 150t; & 300 \leq t < 600 \text{ ms} \\ \text{sen } 2\pi 250t; & 600 \leq t < 800 \text{ ms} \\ \text{sen } 2\pi 300t; & 800 \leq t < 1000 \text{ ms} \end{cases} \quad (3.10)$$

$$g(t) = e^{-\frac{1}{2}at^2} \quad (3.11)$$

El primer espectrograma, calculado con la ventana más angosta, presenta buena resolución en tiempo pero en frecuencia es burda. En cambio, en el tercer espectrograma, la resolución en frecuencias es excelente, mientras la del tiempo es muy mala. El principio de incertidumbre puede escribirse según la siguiente ecuación, y se ilustra en la figura 3.3. El rectángulo de la figura se conoce como *caja de Heisenberg*, y su área mínima está definida por la ecuación (3.12):

$$\sigma_t \sigma_f \geq \frac{1}{4\pi}$$

o bien

$$\sigma_t \sigma_\omega \geq \frac{1}{2} \quad (3.12)$$

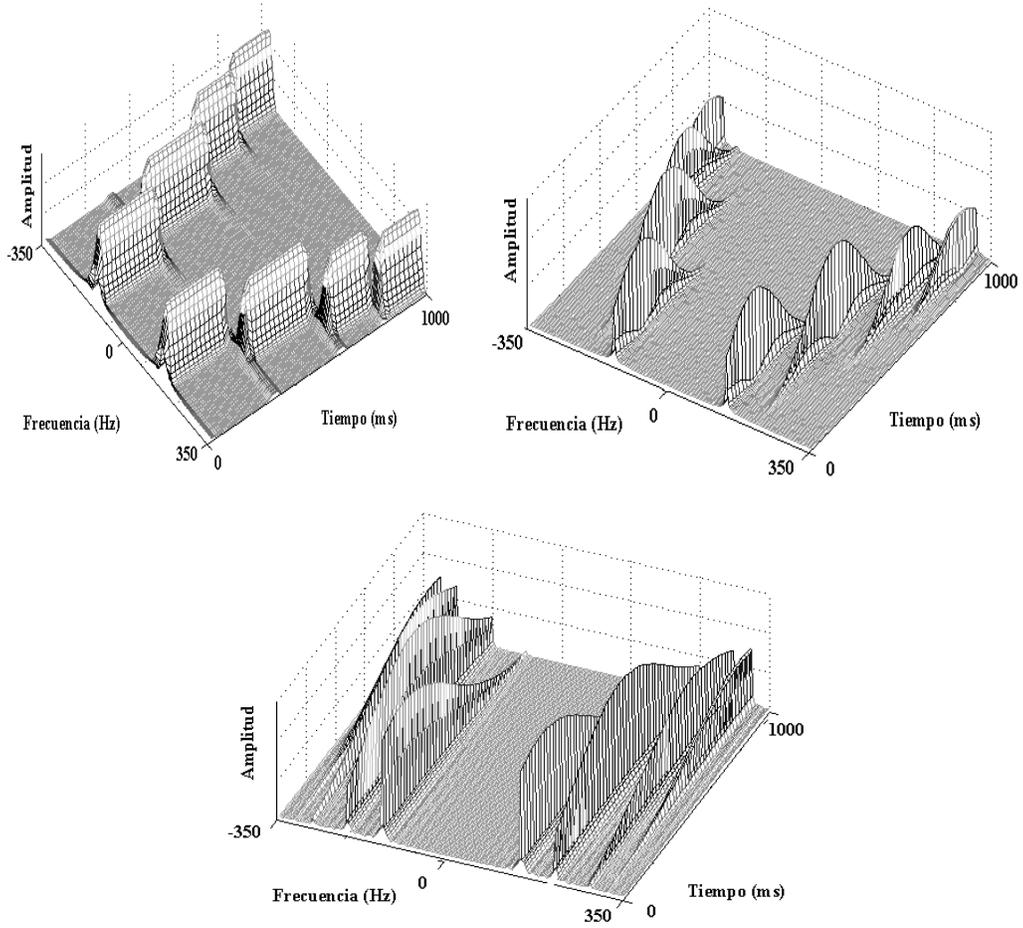


Figura 3.2 Espectrogramas para a) $a = 0.001$, b) $a = 0.0001$ y c) $a = 0.00001$.

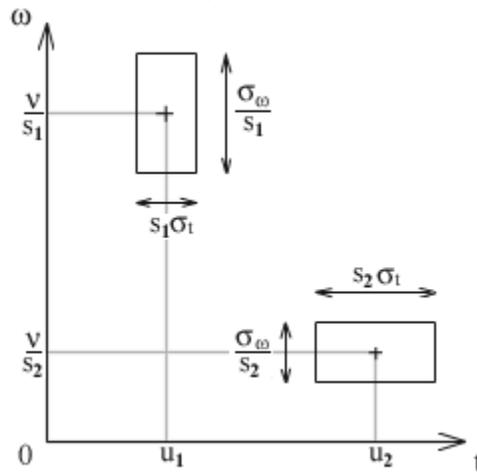


Figura 3.3 Caja de Heisenberg.

3.2.2 Transformación *wavelet* continua

La transformación *wavelet* continua CWT es una alternativa a la STFT que resuelve de mejor manera el problema de la resolución. El análisis es similar en el sentido de que la señal es multiplicada por una función y la transformada se calcula de manera individual para varios segmentos de la función original. La diferencia principal es que la extensión de la ventana es variable, de manera que la señal $f(t)$ se descompone en una suma ponderada de *wavelets* dilatadas y trasladadas. Una *wavelet* $\varphi(t)$ es una función de energía finita centrada en u y con un factor de escala s :

$$\varphi_{u,s}(t) = \frac{1}{\sqrt{s}} \varphi\left(\frac{t-u}{s}\right) \quad (3.13)$$

De forma semejante a la ecuación (3.9), la transformada *wavelet* de $f(t)$ en el tiempo u a una escala s queda expresada de la siguiente manera:

$$CWT(u, s) = \langle f, \varphi_{u,s} \rangle = \int_{-\infty}^{\infty} f(t) \frac{1}{\sqrt{s}} \varphi^*\left(\frac{t-u}{s}\right) dt \quad (3.14)$$

De acuerdo a la integral, puede observarse que esta ecuación también puede ser escrita como la convolución de la señal $f(t)$ con el conjugado de la *wavelet* $\varphi_{u,s}(t)$. La escala es inversamente proporcional a la frecuencia. Una escala grande corresponde a una vista global no detallada de la señal, mientras una escala pequeña proporciona una visión con detalles. Las *wavelets* de escalas altas tienen una duración larga, y viceversa, lo cual resulta conveniente en el análisis de señales que presentan componentes de alta frecuencia de poca duración, así como componentes prolongadas de frecuencia baja. En general, el uso de

escalas de la forma $s = 2^j$ y de desplazamientos $u = n2^j$ es lo más frecuente porque facilita el cálculo numérico. Estas formas son llamadas diádicas.

Es necesario aclarar que una transformada *wavelet* calculada a una escala a^j no es una representación completa, ya que hace falta la parte correspondiente a las escalas mayores que a^j , es decir, las de baja frecuencia. Por esta razón, se introduce una función complementaria paso-bajas, conocida como *función escala* ϕ , que permite que el análisis contemple también las bajas frecuencias. Su definición es similar a la ecuación (3.13).

$$\phi_{u,s}(t) = \frac{1}{\sqrt{s}} \phi\left(\frac{t-u}{s}\right) \quad (3.15)$$

3.2.3 Transformación *wavelet* discreta

En 1976, Croiser, Esteban y Galand desarrollaron una técnica de descomposición de señales discretas. En el mismo año, Crochere, Flanagan y Weber publicaron trabajos similares en el campo de la codificación de voz, y llamaron *codificación sub-banda* a la nueva técnica, la cual sería perfeccionada por Vetterli y Le Gall más de una década después. Unos años después, Burt desarrolló una técnica similar a la que llamó *análisis de resolución múltiple* (MRA).

La transformación *wavelet* discreta (DWT) sigue el mismo principio que la CWT, entonces puede escribirse la forma discreta de la ecuación (3.14), dada por:

$$DWT[n, a^j] = \sum_{m=0}^{N-1} f[m] \psi_j^*[m-n] = f * \psi^*[n] \quad (3.16)$$

Según lo anterior, el filtrado es el primer paso en el procesamiento de señales digitales con *wavelets*. Se usan filtros que dividen la señal $f[n]$ en componentes de alta y de baja frecuencia. Posteriormente, de acuerdo con la regla de Nyquist, la mitad de las muestras pueden ser removidas en un submuestreo. Cada proceso de filtrado y submuestreo constituye un *nivel de descomposición*. El resultado de cada nivel de descomposición de $f[n]$ consta de una aproximación burda $a_j[p]$ y una componente de detalles $d_j[p]$, que se obtienen de los filtros paso-bajas $h[n]$ y paso-altas $g[n]$, respectivamente. Este procedimiento se llama *codificación sub-banda*. La expresión matemática de la descomposición y un esquema asociado se muestran en seguida.

$$\begin{aligned}
 a_j[p] &= \sum_n f[n] h[n-2p] = f[n] * h^*[2p] \\
 d_j[p] &= \sum_n f[n] g[n-2p] = f[n] * g^*[2p]
 \end{aligned}
 \tag{3.17}$$

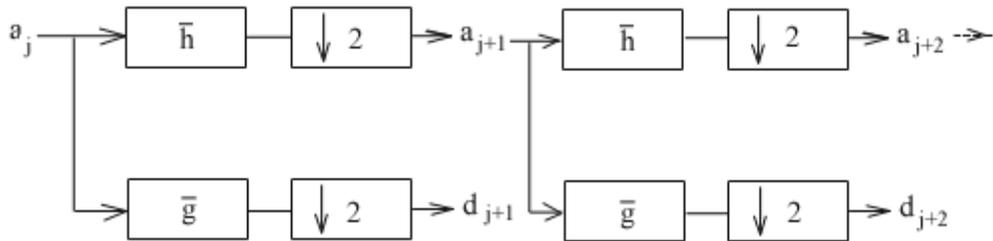


Figura 3.4 Bancos de filtros de dos bandas en la DWT.

Los valores de $a_j[p]$ y de $d_j[p]$ se conocen como *coeficientes wavelet*. Las frecuencias más significativas en la señal original aparecerán como los coeficientes de mayor amplitud en la región correspondiente de la señal transformada. Para el cálculo del siguiente nivel de descomposición basta con repetir el procedimiento para la última $a_j[p]$ obtenida, colocada en el lugar de $f[n]$. La diferencia entre la DWT y la FT es que la localización temporal de esas frecuencias queda registrada. Las bandas de frecuencia que no son muy prominentes serán representadas con amplitudes pequeñas en la DWT, y pueden ser ignoradas con el fin de reducir la cantidad de datos sin que signifique una pérdida importante de información.

Al igual que la de Fourier, la transformación *wavelet* es una operación reversible. En este caso, la señal original puede obtenerse a partir de los coeficientes wavelet de cada nivel de descomposición. La reconstrucción es la operación inversa a la descomposición y su expresión matemática es la siguiente.

$$f[n] = \sum_n h[p-2n] a_j[n] + \sum_n g[p-2n] d_j[n]
 \tag{3.18}$$

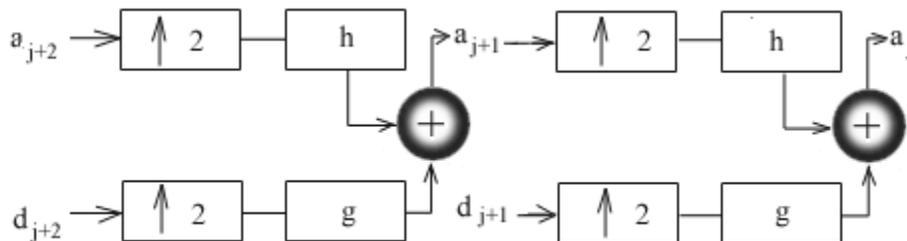


Figura 3.5 Bancos de filtros de dos bandas en la DWT inversa.

3.3 Las *wavelets* en imágenes

Para extrapolar el análisis con *wavelets* a las señales de dos dimensiones, se considera al conjunto de funciones $f(x_1, x_2)$ como una expansión lineal de las funciones separables $a(x_1)$ y $b(x_2)$.

$$f(x, y) = \sum_{m=-\infty}^{\infty} c[m] a_m[x_1] b_m[x_2] \quad (3.19)$$

Con base en lo anterior, y si se considera a ϕ como la función escala y φ es la *wavelet* correspondiente, para generar una base de *wavelets* en el espacio de funciones finitas de dos dimensiones se definen tres *wavelets* básicas:

$$\varphi^1(x) = \phi(x_1)\phi(x_2), \quad \varphi^2 = \phi(x_1)\phi(x_2), \quad \varphi^3 = \phi(x_1)\phi(x_2), \quad (3.20)$$

y con ellas, para $k = 1, 2, 3$:

$$\varphi_{j,n}^k = \frac{1}{2} \varphi^k \left(\frac{x_1 - 2^j n_1}{2^j}, \frac{x_2 - 2^j n_2}{2^j} \right) \quad (3.21)$$

Estas tres *wavelets* proporcionan detalles de la imagen a diferentes escalas y orientaciones: φ^1 para frecuencias horizontales ω_1 bajas y frecuencias verticales ω_2 altas; φ^2 para ω_1 altas y ω_2 bajas y, finalmente, φ^3 para ω_1 y ω_2 altas. Solamente hace falta la aproximación para bajas frecuencias, para lo cual se usa la función escala, definida de la siguiente forma:

$$\phi_{j,n}^2 = \phi_{j,n_1}(x_1)\phi_{j,n_2}(x_2) = \frac{1}{2^j} \phi \left(\frac{x_1 - 2^j n_1}{2^j} \right) \phi \left(\frac{x_2 - 2^j n_2}{2^j} \right) \quad (3.22)$$

La siguiente figura ilustra las regiones del plano (ω_1, ω_2) en las que se concentra la energía de cada φ^k y de ϕ^2 .

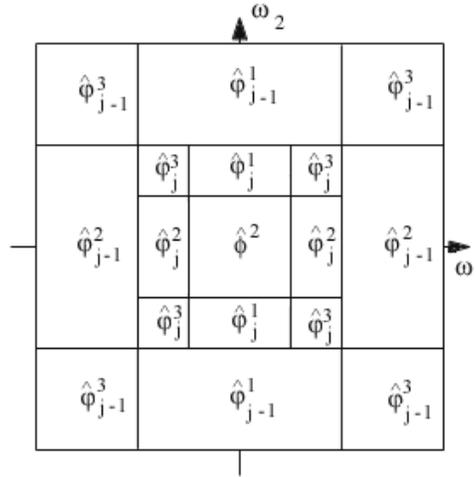


Figura 3.6 Rectángulos de concentración de energía en el plano (ω_1 , ω_2).

Tomando lo anterior como base, el algoritmo para la DWT, presentado en la sección 3.2.3, puede extenderse para dos dimensiones con el siguiente grupo de ecuaciones:

$$\begin{aligned}
 a_j[n_1, n_2] &= f[n_1, n_2] * h[-2n_1]h[-2n_2], \\
 d_j^1[n_1, n_2] &= f[n_1, n_2] * h[-2n_1]g[-2n_2], \\
 d_j^2[n_1, n_2] &= f[n_1, n_2] * g[-2n_1]h[-2n_2], \\
 d_j^3[n_1, n_2] &= f[n_1, n_2] * g[-2n_1]g[-2n_2].
 \end{aligned}
 \tag{3.23}$$

Como ejemplo, a continuación se presenta la descomposición de una imagen cuya dimensión es de 256x256 píxeles. En el área correspondiente a ω_1 alta y ω_2 baja puede observarse que los detalles horizontales de la imagen se han perdido, mientras los verticales están bien definidos y viceversa. Por otro lado, los detalles diagonales sobresalen en la región de frecuencias altas para ambas variables. Finalmente, en la zona correspondiente a los coeficientes de aproximación se observa la imagen original con baja resolución. La dimensión de cada una de las regiones es de 128x128 píxeles. En el caso de un segundo nivel de descomposición, se tienen tres zonas de 128x128 y cuatro de 64x64, estas últimas correspondientes a la parte que fue descompuesta una vez más.

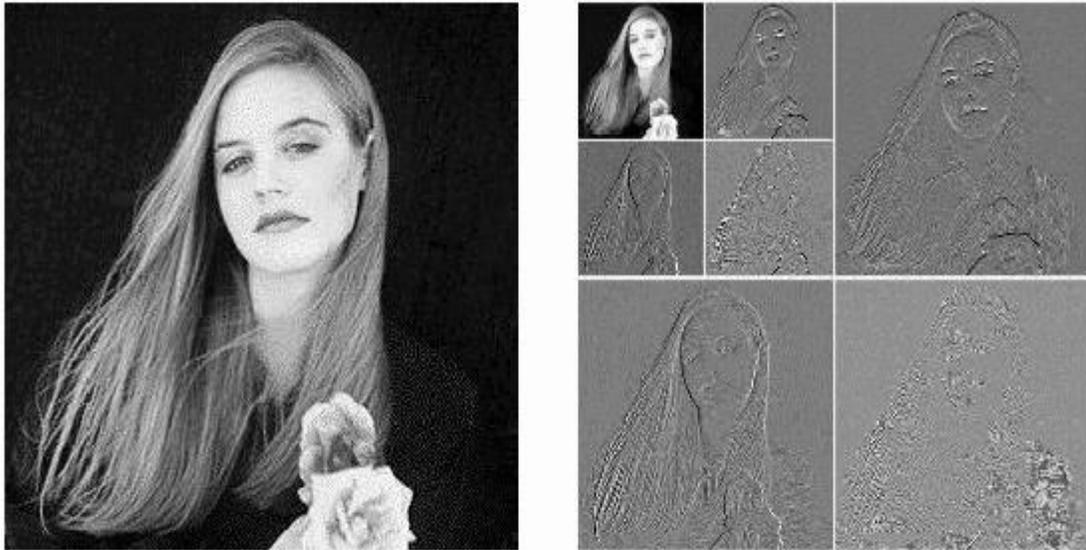


Figura 3.7 Descomposición wavelet de una imagen de 256x256 pixeles en dos niveles

3.3.1 Reducción de ruido con wavelets

Tradicionalmente, las técnicas de reducción de ruido basadas en el análisis con *wavelets* funcionan de la siguiente manera. Cuando se lleva a cabo la descomposición de una imagen, se obtienen coeficientes de aproximación y coeficientes de detalles. Si los coeficientes de detalles son pequeños, pueden ser omitidos parcial o totalmente en el momento de la reconstrucción, sin que esto afecte significativamente las características principales de la imagen. Si se considera que una imagen está afectada por ruido, éste presenta componentes pequeñas en varias frecuencias, entonces, el uso de la técnica llamada *thresholding* es una buena opción para reducir el ruido. Consiste en igualar a cero todos los coeficientes que sean menores que un *umbral* predeterminado, y si se considera que los coeficientes pequeños corresponden al ruido, éste será suprimido, o al menos, reducido, mientras los coeficientes grandes corresponderán a la información importante de la imagen, que será conservada. Cuando la imagen se reconstruye se verifica que la contribución del ruido ha disminuido notablemente. La característica más importante de esta técnica es que remueve el ruido sin *suavizar* los contornos de la imagen, de tal suerte que ésta no pierde definición. Sin embargo, resulta evidente que esta técnica no da buenos resultados cuando se tiene una señal con una relación señal a ruido pequeña, ya que en la descomposición, los coeficientes de ruido pueden ser iguales o mayores que los que portan información importante de la imagen, de manera de que en la reconstrucción se apreciaría una imagen deteriorada. En este trabajo se tomará en cuenta esta situación.

Capítulo IV. Algoritmo de reducción de ruido

4.1 De la compresión a la reducción de ruido

A partir de la necesidad de reducir la cantidad de datos con el fin de ahorrar espacio en el almacenamiento de archivos y de lograr una transmisión rápida de los mismos, una gran parte del procesamiento de imágenes se enfoca a la compresión. La base fundamental de este proceso es la eliminación de datos redundantes. Desde el punto de vista matemático, se busca transformar una imagen en un arreglo de datos estadísticos no correlacionados. El desarrollo de algoritmos nuevos es importante, ya que se pretende que el proceso sea cada vez más eficiente, es decir, que la compresión sea mayor sin que haya una pérdida significativa de información. Es por eso que, históricamente, la compresión de imágenes ha sido el área de más actividad en cuanto a investigación y desarrollo de técnicas de procesamiento, y el resultado de esto es que, aparentemente, es un campo mucho más maduro que el de la reducción de ruido. Sin embargo, la gran relación que existe entre ambos campos, implica que el exitoso avance de la primera línea de investigación bien puede ser aprovechado por la segunda. De hecho, Natarajan, en 1993 [16], propuso el uso un algoritmo de compresión para hacer reducción de ruido en señales contaminadas con ruido blanco. La estrecha relación entre la compresión y la reducción de ruido parte del hecho de que una señal correlacionada es compresible, pero el ruido no correlacionado no lo es. Este principio ha constituido la base de varios algoritmos para eliminar el ruido, los cuales están basados principalmente en modelos estadísticos.

4.2 Árboles de ceros

Con base en la transformación *wavelet*, Shapiro presentó en 1993 [20] un algoritmo simple pero muy efectivo para compresión de imágenes, llamado codificación embebida por árboles de ceros, el cual genera una cadena de bits ordenados progresivamente de acuerdo a su importancia (cadena embebida). Esta técnica toma ventaja de la correlación que existe entre los coeficientes *wavelet* de sub-bandas distintas; es decir, cuando un coeficiente $c[k_1/2, k_2/2]_{j+1}$ es irrelevante a una escala $j+1$, también lo será el coeficiente $c[k_1, k_2]_j$ en la escala j . De esta manera es posible encontrar los coeficientes portadores de información más importante y darles prioridad durante la codificación. Un coeficiente es irrelevante cuando es menor que un umbral T predeterminado, dicho de otra manera, los coeficientes grandes son más importantes que los pequeños porque contienen más información. La expresión matemática para describir lo anterior es la siguiente:

$$P\left(c_j[k_1, k_2] \leq T \mid c_{j+1}\left[\left\lceil \frac{k_1}{2} \right\rceil, \left\lceil \frac{k_2}{2} \right\rceil\right] \leq T\right) \approx 1 \quad (4.1)$$

donde $P(A|B)$ es la probabilidad condicionada de A dado el evento B .

Después de transformar una imagen con la DWT, es posible representar a la transformada utilizando estructuras de árbol en las que cada raíz corresponde a un coeficiente *padre*

$c[k_1/2, k_2/2]_{j+1}$ y tiene cuatro ramas, una para cada coeficiente *hijo* $c[k_1, k_2]_j$, como se ilustra a continuación.

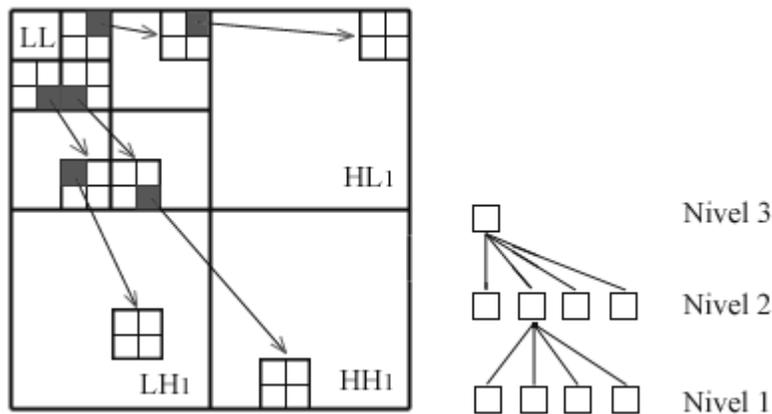


Figura 4.1 a) Relación de coeficientes de distintas sub-bandas, b) estructura de árbol.

Todos los elementos *padre* tienen cuatro *hijos*, excepto los de la escala más grande, que solamente tienen tres.

El análisis de los coeficientes debe cumplir con que ningún elemento *hijo* puede ser analizado antes que su *padre*, por lo tanto, solo hasta que una sub-banda ha sido totalmente barrida puede empezarse con la siguiente. Se propone el siguiente esquema de barrido, en el que se comienza por la banda de menor frecuencia LL_N , se continúa con HL_N , LH_N , HH_N , posteriormente se barren los coeficientes del nivel $N-1$, y así, sucesivamente.

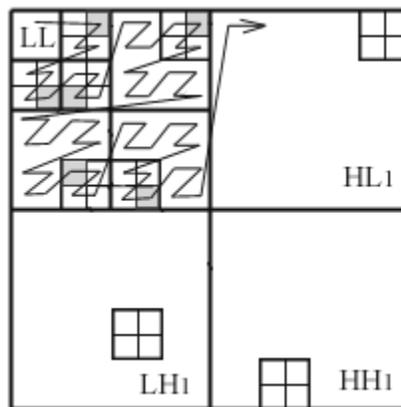


Figura 4.2 Barrido de coeficientes.

El funcionamiento detallado del algoritmo no se presenta en este documento porque está fuera del objetivo de la tesis, pero puede encontrarse en [20] y [26]. Lo explicado hasta este punto es suficiente para el desarrollo del algoritmo de reducción de ruido que se pretende implementar.

4.3 Modelado estadístico de los coeficientes *wavelet*

Muchos algoritmos de compresión se basan en la DWT, ya que logra concentrar la mayor parte de la información en algunos coeficientes, además de que se conserva cierta estructura al interior de los mismos. Los coeficientes *wavelet* pueden modelarse como variables aleatorias independientes idénticamente distribuidas en forma gaussiana [12], y esta característica ha sido el fundamento de muchos algoritmos de compresión y de reducción de ruido. Cuando se toma en cuenta la interdependencia espacial entre los coeficientes, tanto los de la misma sub-banda como los de sub-bandas distintas, puede llegarse a algoritmos más sofisticados y, probablemente, de complejidad baja. En el caso particular de los métodos de reducción de ruido, se ha observado que cuando se toma en cuenta esta interdependencia o *adaptividad* espacial entre los coeficientes, los resultados obtenidos son de mejor calidad, particularmente cuando se trabaja con imágenes reales.

4.3.1 Estimador del mínimo error cuadrático medio (MMSE)

En cada escala, los coeficientes *wavelet* presentan un comportamiento de acuerdo a una distribución de Gauss, con media igual a cero. Cuando se calcula la DWT para una imagen con ruido $g[x_1, x_2]$, si se considera que la DWT es una transformación ortogonal, entonces la ecuación (3.3) puede expresarse en el dominio *wavelet* de la siguiente manera:

$$c_g[k_1, k_2] = c_f[k_1, k_2] + n[k_1, k_2] \quad (4.2)$$

donde c_g , c_f y n corresponden, respectivamente, a los coeficientes de $g[x_1, x_2]$, a los de la señal sin ruido y a los de ruido blanco con distribución gaussiana. Con base en el hecho de que la suma de dos variables aleatorias gaussianas e independientes es otra variable aleatoria gaussiana con varianza igual a la suma de sus varianzas, entonces:

$$\sigma_g^2 = \sigma_f^2 + \sigma_n^2 \quad (4.3)$$

Como σ_g es un parámetro que puede calcularse, y con la suposición de que es posible obtener un estimado de σ_n , entonces σ_f puede encontrarse de la siguiente manera:

$$\sigma_f^2 = \sigma_g^2 - \sigma_n^2 \quad (4.4)$$

$$= \frac{1}{MN} \left(\sum_{i=1}^M \sum_{j=1}^N \left(c_g[i, j] - \mu_{c_g} \right)^2 \right) - \sigma_n^2 \quad (4.5)$$

$$= \frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N c_g^2[i, j] - \sigma_n^2 \quad (4.6)$$

Ahora, con base en la teoría de la estimación por mínimos cuadrados, aplicada a una señal alterada por ruido blanco, es posible obtener un valor aproximado de cada coeficiente c_f con la siguiente ecuación:

$$c_f[k_1, k_2] = \frac{\sigma_f}{\sigma_f^2 + \sigma_n^2} c_g[k_1, k_2] \quad (4.7)$$

4.3.2 Estimador de máxima similitud (ML)

En 1999, Mihçak, Konzintsev y Ramchandran [15] introdujeron un nuevo modelo para los coeficientes *wavelet* de las imágenes, partiendo de un método de compresión previamente propuesto por LoPresto y Ramchandran [9]. En este modelo, se asume que existe un *campo de varianza* desconocido y variante en el espacio. Igualmente, se propone que el campo de varianza no presenta cambios drásticos. Bajo estas suposiciones, los coeficientes se modelan como variables aleatorias independientes distribuidas localmente de forma idéntica, de manera que puede esperarse una correlación alta entre las varianzas de coeficientes contiguos. Matemáticamente, si se forma un vecindario $N_{m \times m}$ alrededor de cada coeficiente $c_g[k_1, k_2]$, entonces:

$$\sigma_g^2(k_1, k_2) \approx \sigma_g^2(k_1 + i, k_2 + j) \quad i, j = -\frac{m-1}{2}, \dots, 1, 0, 1, \dots, \frac{m-1}{2}. \quad (4.8)$$

En el vecindario, m debe ser un número impar para asegurar que $c_g[k_1, k_2]$ está en el centro. También es recomendable que m no sea muy grande ya que la consideración anterior deja de ser válida para coeficientes muy lejanos de $c_g[k_1, k_2]$, pues hay que recordar que el campo de varianza es ligeramente variante en el espacio, de modo que puede considerarse estático solamente para una región limitada. Si la suposición de la existencia del campo de varianza es correcta, entonces puede proponerse un estimador de máxima similitud para determinar las varianzas locales de los coeficientes de f , y según el resultado de la ecuación (4.6), se propone la siguiente ecuación:

$$\sigma_f^2(k_1, k_2) \approx \text{máx} \left(0, \frac{1}{m^2} \sum_{i, j=-m}^m c_g^2[k_1 + i, k_2 + j] - \sigma_n^2 \right) \quad (4.9)$$

En este punto es pertinente hacer dos comentarios, uno en relación con la calidad del estimador y otro con respecto a σ_n^2 . Con base en resultados experimentales, puede afirmarse que la calidad de los estimados de $\sigma_f^2(k_1, k_2)$ y de σ_n^2 no afecta significativamente en el proceso de reducción de ruido, es por eso que se eligió la ecuación (4.9) sobre otras alternativas más elaboradas, ya que el sacrificio en la calidad del procesamiento es despreciable y, en cambio, el ahorro en el tiempo de ejecución es notable. En cuanto a σ_n^2 , usualmente se asume en la literatura que se trata de un parámetro conocido cuando en realidad no lo es. En el caso del problema que se aborda en este trabajo, se sabe que el

ruido tiene una distribución de Poisson y que su intensidad es aproximadamente igual a la raíz cuadrada de la intensidad de la señal contaminada $g(x_1, x_2)$ [24]. Por otro lado, se ha comprobado que para una imagen de ruido $N(x_1, x_2)$, generada con base en la distribución y las características mencionadas, los coeficientes *wavelet* $n(k_1, k_2)$ resultantes de la DWT presentan una distribución *gaussiana* con media igual a cero. Además, si $g(x_1, x_2)$ es una imagen con poca información en frecuencias altas (sin contornos bien definidos, como las imágenes comunes en la Astronomía infrarroja), entonces la varianza de los coeficientes *wavelet* de $N(x_1, x_2)$ en la escala más fina es muy cercana a la de los de $g(x_1, x_2)$ en la misma escala, con lo que puede proponerse un estimador confiable para σ_n^2 dado por:

$$\sigma_n^2 \approx \text{var}(c_{g,j}) \quad (4.10)$$

donde se considera el caso en el que $j=1$, es decir, se toman en cuenta únicamente los coeficientes del primer nivel de la descomposición, en las bandas LH_1 , HL_1 y HH_1 .

4.4 Integración del algoritmo

El funcionamiento del algoritmo de reducción de ruido se describe a continuación en forma resumida.

- i) Se realiza la DWT de L niveles. $L=2$ es un número de niveles de descomposición adecuado, ya que es suficiente para distinguir los coeficientes de ruido de los de la información más importante. Puede utilizarse cualquier tipo de *wavelet* que sea ortogonal. En este trabajo se hicieron pruebas con *wavelets* biortogonales, *symlets*, *coiflets* y de Daubechies [4].
- ii) Se define una matriz de ceros D_j para cada sub-banda en cada escala. Las dimensiones de D_j deben ser idénticas a las de la sub-banda correspondiente.
- iii) Todos los coeficientes de la banda de aproximación se comparan con un umbral T , tomando en cuenta la siguiente simplificación de la ecuación (4.1), en la cual los coeficientes han sido sustituidos por sus respectivas varianzas con el fin de ajustarla a este algoritmo.

$$P\left(\sigma_j^2(k_1, k_2) = 0 \mid \sigma_{j+1}^2\left(\left[\frac{k_1}{2}\right], \left[\frac{k_2}{2}\right]\right) \leq T\right) \approx 1 \quad (4.11)$$

- iv) El cálculo de la varianza estimada de los coeficientes de f solamente se llevará a cabo para aquellos coeficientes de aproximación que resulten mayores que T , así como para sus *hijos*. El umbral T puede determinarse con base en la varianza del ruido obtenida con la ecuación (4.10) y de acuerdo con la siguiente relación:

$$T = \sqrt{\frac{\sigma_n^2}{3 \times 10^{-3}}} \quad (4.12)$$

Sin embargo, los resultados al automatizar el cálculo de T , con esta ecuación o por alguno de los métodos empíricos propuestos en la literatura, son muy variables, muchas veces de baja calidad, por lo que lo más recomendable es que T sea fijado por el usuario. Un umbral ideal depende tanto del ruido de la señal como de la *wavelet* utilizada en la descomposición, de tal suerte que el desarrollo de una rutina más confiable para su cálculo sería complicado y requeriría un trabajo adicional al realizado en esta tesis. Cada varianza estimada se adjudica al elemento correspondiente de la matriz \mathbf{D}_j , con lo cual, cuando todos los coeficientes de aproximación han sido analizados, \mathbf{D}_j contendrá todos los valores σ_f necesarios para el cálculo de los coeficientes c_f .

- v) Con la ecuación (4.7) se obtiene el estimado de los coeficientes de la señal sin ruido. En seguida puede hacerse la DWT inversa.
- vi) Finalmente, aunque en este punto el ruido ha disminuido notablemente, aún hace falta compensar el desajuste originado por el nivel medio del ruido en la señal observada, lo cual se logra al utilizar la siguiente ecuación:

$$f[x_1, x_2] = iDWT\{c_f[k_1, k_2]\} - \left(\frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N g[i, j] \right)^{\frac{1}{2}} \quad (4.13)$$

Capítulo V. Resultados y conclusiones

5.1 Discusión de los resultados

Con el fin de hacer una comparación adecuada de la calidad de la reducción de ruido, es necesario contar con una imagen conocida de referencia para saber cómo debe quedar después del procesamiento, por lo tanto una imagen astronómica es poco útil en este caso ya que de ella solamente se conoce la imagen ruidosa detectada y no se conoce la fuente. Por esta razón se utilizará la imagen de la figura 3.7, que puede ser contaminada con ruido generado de manera artificial, como se ilustra en la figura siguiente. Por concordancia con lo expuesto en los capítulos anteriores, se usará ruido con una distribución de Poisson. Además, se ha mencionado que las imágenes infrarrojas presentan contornos poco definidos, lo cual fue una consideración para el desarrollo del algoritmo y para escoger esta imagen de referencia que se aprecia empañada o borrosa.



Figura 5.1 a) Imagen original, b) Imagen contaminada con $\sigma_n^2 = 10$.

Existen medidas subjetivas y objetivas para determinar la calidad de una imagen resultante de un proceso determinado. Una medida subjetiva es simplemente dar una opinión de la calidad en una escala de 1 a 5: muy baja, baja, media, alta, muy alta. Por otro lado, una medida objetiva es el *Valor Pico de la Relación Señal a Ruido* (PSNR), expresada en decibeles (dB), que puede calcularse según la siguiente ecuación matemática:

$$PSNR = 10 \log_{10} \frac{255^2}{\frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N (f[i,j] - f_E[i,j])^2} \quad (5.1)$$

El denominador en el argumento del logaritmo es error cuadrático medio (MSE) entre la función original f y su valor estimado f_E . Usualmente el PSNR es un valor entre 20 dB y 50 dB. En realidad, este valor no es significativo por sí mismo, en cambio, la comparación

entre dos valores correspondientes a métodos distintos de procesamiento da una idea de la calidad de las imágenes.

En relación con los resultados, en primer lugar es interesante verificar el funcionamiento del algoritmo desarrollado en el capítulo anterior (ACA) para varias funciones *wavelets* distintas. En la tabla 5.1 se presentan los valores de PSNR para ocho *wavelets* ortogonales utilizando vecindarios de 3x3, 5x5 y 7x7 píxeles. Esta comparación ayudará a determinar qué *wavelet* es la más adecuada.

	PSNR (dB)							
	Haar	Db10	Sym5	Sym10	Bior3.9	Bior6.8	Coif2	Coif5
ACA 3x3	46.0846	45.4184	45.9975	46.6167	42.0080	46.2073	46.5592	46.7557
ACA 5x5	47.1741	47.7765	47.1766	48.0183	42.3583	47.5591	47.9332	48.2339
ACA 7x7	47.5160	48.2558	47.6229	48.5069	42.5197	48.0743	48.4343	48.7841

Tabla 5.1 Comparación de los resultados del algoritmo presentado en el capítulo 4, utilizando diez *wavelets* distintas. La imagen fue contaminada con ruido de distribución de Poisson con $\sigma_n^2 = 10$.

La *coiflet* de quinto orden parece aventajar ligeramente al resto. Contrasta la *wavelet* biortogonal de órdenes 3 y 9, con un PSNR notablemente menor que las demás. Hay que mencionar que los resultados anteriores no son definitivos, pues se observó que pueden variar considerablemente en función del umbral seleccionado en el tercer paso del algoritmo. También cabe aclarar que en la prueba anterior se eligió el umbral que proporcionara los mejores resultados, partiendo de una serie de valores situados alrededor del umbral calculado con la ecuación (4.12).

En la tabla 5.2 se presenta una comparación de ACA con otros algoritmos. Los dos primeros son instrucciones de Matlab para reducción de ruido en imágenes. El método de Wiener fue mencionado en el capítulo tercero de este trabajo. La instrucción *wdencomp* es híbrida, ya que puede usarse tanto para reducción de ruido como para compresión de imágenes. Ambas modalidades realizan *thresholding* de los coeficientes *wavelet*. El tercer método de comparación es el *modelado estadístico espacialmente adaptativo* (SASM) propuesto por Mihçak, Konzintsev y Ramchandran [16], un algoritmo novedoso a partir del cual han sido desarrollados algunos métodos semejantes cuyos resultados no varían mucho del original. También se definen vecindarios locales para los coeficientes *wavelet*.

En la tabla puede observarse que el PSNR de los comandos de Matlab resulta muy bajo comparado con los otros dos aún cuando la *wavelet* utilizada es la que proporcionó los menores niveles de calidad en la primera prueba. También puede verificarse que el algoritmo ACA resulta superior al SASM en todos los casos, lo cual confirma la validez de las consideraciones hechas durante su diseño, ya que si se aplica a imágenes con buena definición la diferencia no es tan grande y, en varias ocasiones, el SASM presenta mejores resultados. Debido a que los valores de PSNR son promedios, el funcionamiento de estos cuatro métodos se aprecia mejor subjetivamente porque así puede compararse directamente la imagen resultante con las de la figura 5.1.

	PSNR (dB)		
	$\sigma_n^2 = 10$	$\sigma_n^2 = 20$	$\sigma_n^2 = 30$
Matlab (wiener2)	28.0967	22.1190	18.5987
Matlab (wdencomp)	28.1280	22.1333	18.6065
SASM 3x3	40.1910	37.2487	35.4616
ACA 3x3	42.0080	41.1408	37.6581
SASM 5x5	39.3338	36.3719	34.5924
ACA 5x5	42.3583	41.6894	38.0173
SASM 7x7	38.8709	35.9101	34.1392
ACA 7x7	42.5197	41.9484	38.1847

Tabla 5.2 Comparación de los resultados para varios algoritmos para una imagen contaminada con tres niveles de ruido diferentes. Para SASM y ACA se utilizó la *wavelet* biortogonal 3.9.

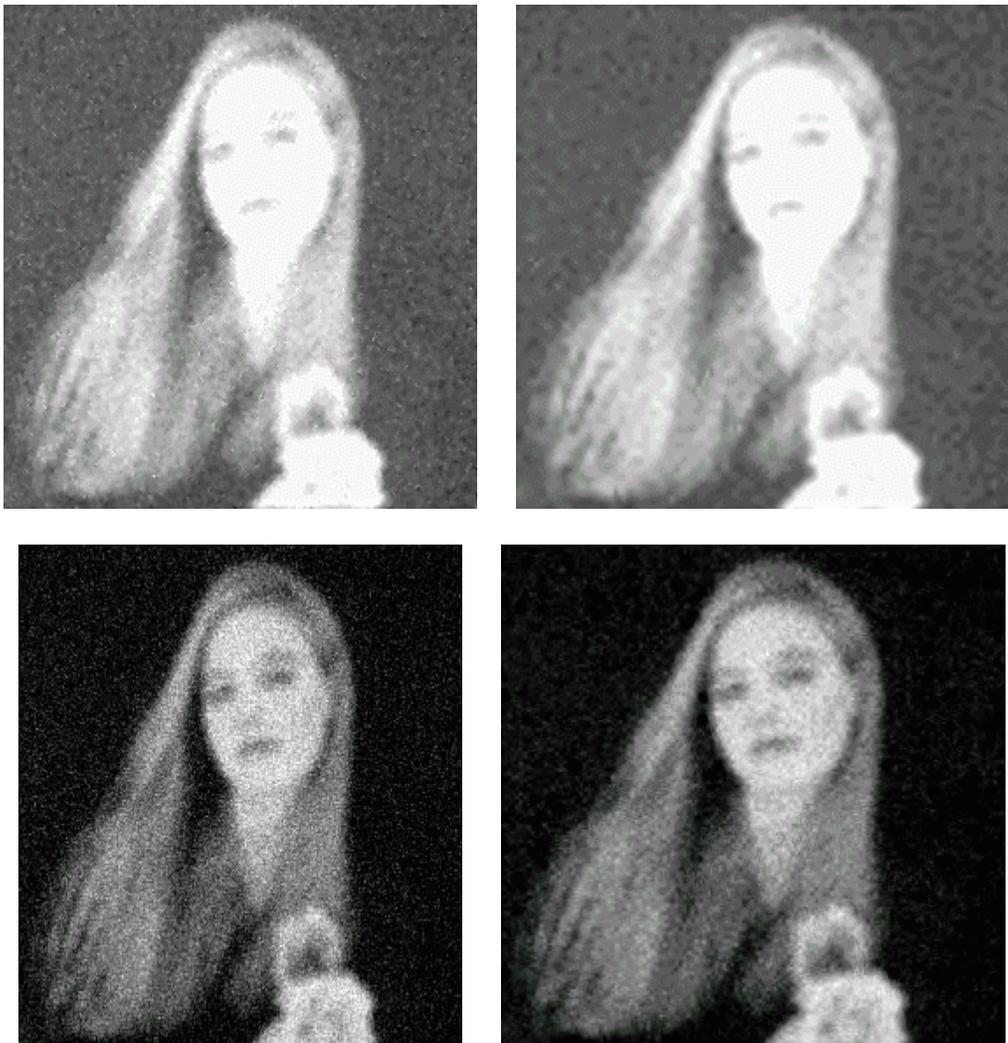


Figura 5.2 Resultado de la reducción de ruido utilizando a) Wiener2, b) Wdencomp,

c) SASM 7x7, d) ACA 7x7.

La diferencia puede apreciarse claramente en el fondo de la imagen, ya que en las imágenes a) y b) no es negro, sino que todavía tiene una gran cantidad de ruido. En c) se ve mejor, aunque todavía presenta cierto ruido. Finalmente, en d) el ruido del fondo ha disminuido hasta sólo algunos puntos. Es posible disminuir la cantidad de puntos casi hasta cero si se aumenta el umbral, sin embargo, esta acción provocaría un decremento en el PSNR, en otras palabras, la medida subjetiva de la calidad aumentaría mientras la medida objetiva disminuiría. Es importante mencionar que el umbral no debe incrementarse demasiado si no se quiere eliminar información importante en la imagen. La figura 5.3 ejemplifica lo anterior. En ambas imágenes el ruido de fondo ha desaparecido, sin embargo, en b) la imagen se aprecia deteriorada.



Figura 5.3 Reducción de ruido con ACA utilizando a) un umbral adecuado, b) un umbral demasiado grande.

5.2 Conclusiones

Los objetivos fundamentales del proyecto se cumplieron satisfactoriamente. El diseño y la programación de un algoritmo de reducción de ruido en imágenes infrarrojas será un apoyo importante para la comunidad astronómica. Se ha demostrado el buen funcionamiento del programa tanto cualitativa como cuantitativamente, superando los resultados logrados por otros métodos en virtud del conocimiento de las características que presenta el problema particular de la visión infrarroja. Hay que aclarar que, si bien, las consideraciones de diseño fueron hechas con la finalidad de obtener los mejores resultados para un tipo específico de imágenes, no se trata de la única aplicación del método desarrollado, puesto que también presenta buenos resultados en tipos de imágenes diversos. Así mismo, aunque en la literatura se habla únicamente del ruido blanco como fuente de alteraciones, se comprobó que el método funciona igual para imágenes contaminadas con ruido de Poisson. Para reforzar lo anterior, en el Anexo A se presentan resultados obtenidos en varias imágenes corruptas por distintos tipos de ruido.

Además del buen cumplimiento de los objetivos planteados originalmente, es importante señalar las conclusiones particulares del trabajo de tesis. En primer lugar hay que mencionar que la investigación que fue necesaria para situar el proyecto en un marco de referencia, que es la instrumentación astronómica, constituye definitivamente una motivación a seguir participando en esa actividad. Detrás de las observaciones, de los descubrimientos y de las teorías existe un gran número de horas de experimentación y desarrollo de tecnología de punta en materia de óptica, análisis de estructuras, diseño y construcción de controladores y otros circuitos electrónicos, también investigación sobre métodos de control y su implementación, estudio de física de semiconductores para el diseño y fabricación de nuevos dispositivos y un sin número de disciplinas más. El estudio del funcionamiento de los detectores de radiación electromagnética resultó particularmente interesante, en especial el grado de tecnología necesario para alcanzar la precisión adecuada en el proceso de fabricación. Por otro lado, es motivador encontrar que la astronomía mexicana ocupa un lugar importante a nivel mundial y que constituye un catalizador para el desarrollo tecnológico y la investigación en el país.

Uno de los motivos principales para la elección de este proyecto como tema del trabajo de tesis fue el reto que representa el incursionar en un campo que no se domina por completo, pero teniendo una base en los conocimientos conseguidos en la carrera. El procesamiento de imágenes era una opción tan atractiva como interesante, y la oportunidad de estudiar una teoría moderna como herramienta para abordar el problema influyó de manera definitiva en la decisión. La investigación de los temas de procesamiento de señales con *wavelets* consumió la mayor parte del tiempo invertido en el proyecto, pero en retribución quedan la experiencia adquirida y la posibilidad de su posterior aplicación en la búsqueda de soluciones a cuestiones pendientes en lo que se refiere al tratamiento de imágenes, en particular para la astronomía. Desde luego los resultados obtenidos en esta área ya son muy alentadores.

5.3 Trabajo a futuro

La primera fase del trabajo futuro es a corto plazo. Consiste en el diseño de una presentación gráfica interactiva del programa en la que el usuario pueda verificar el funcionamiento del algoritmo y familiarizarse con el proceso de reducción de ruido. Será posible el control de parámetros como el umbral, el nivel de ruido y la selección de la función *wavelet*, así como la comparación del ACA con los métodos de Matlab. Al igual que el programa, la interfaz será diseñada en Matlab. La figura 5.4 es un esquema de la pantalla de trabajo que se pretende integrar. La parte más importante del trabajo a futuro es la integración del programa como un módulo de IRAF (Facilidad para el Análisis y Reducción de Imágenes). IRAF es un programa de análisis de datos astronómicos cuyo uso se ha generalizado en institutos de investigación y observatorios en todo el mundo. Tiene su propio formato de datos, lenguaje de programación, comandos y un sistema de trabajo y ayuda en línea. Consiste en una serie de módulos alrededor de un núcleo central que siempre está disponible y ofrece rutinas generales para el procesamiento de imágenes. El acceso a tareas más específicas es posible a través de los módulos o paquetes que pueden descargarse durante la sesión de trabajo. IRAF está desarrollado en el ambiente de

programación SPP/VOS, similar al lenguaje C, que puede utilizarse para generar nuevos operadores. La inclusión del ACA en el ambiente de IRAF proveerá a la observación astronómica de una herramienta novedosa para el tratamiento de las imágenes, además de abrir la posibilidad de conocer otras técnicas de procesamiento así como necesidades actuales y problemas particulares.

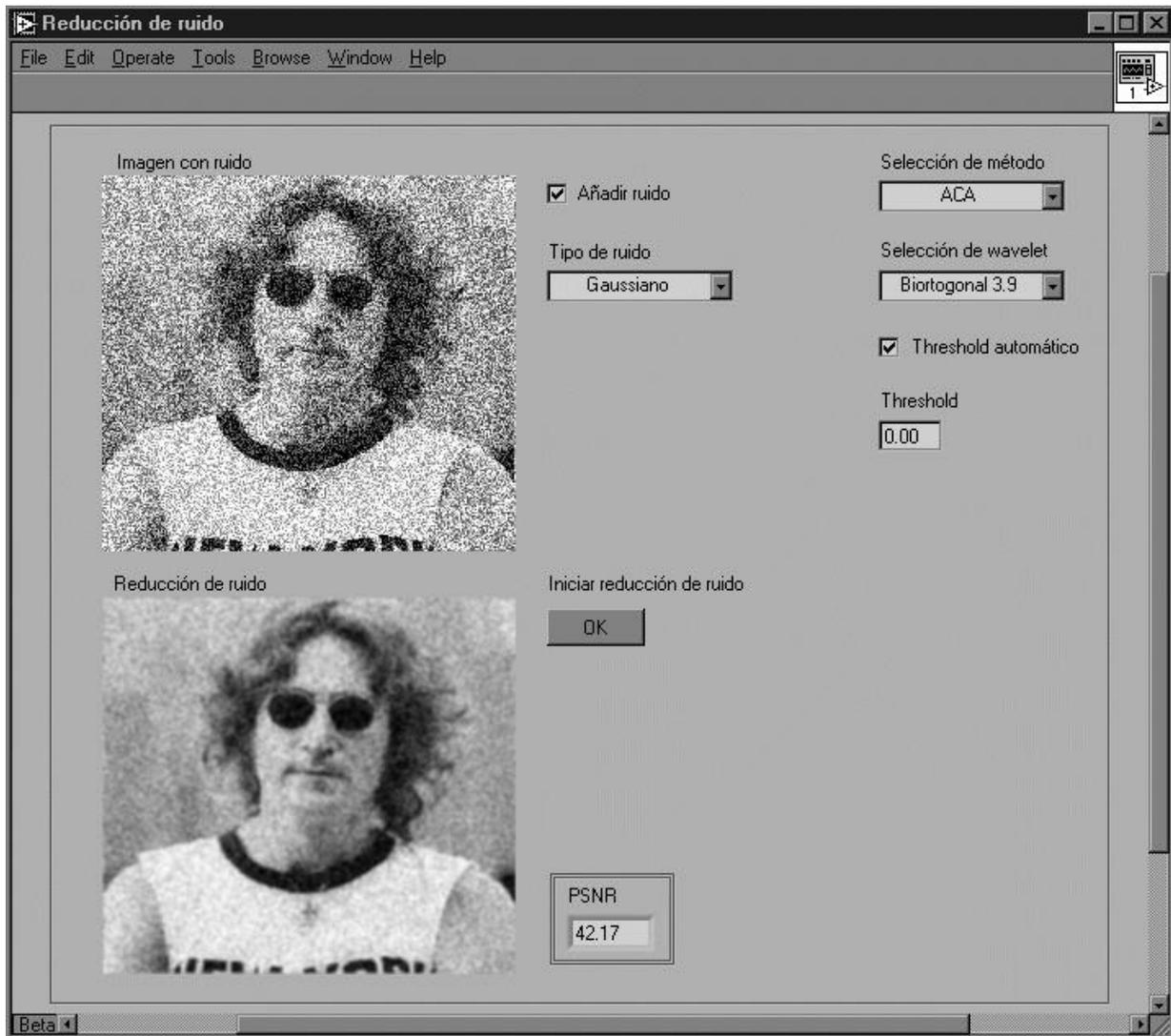


Figura 5.4 Interfaz gráfica.

Apéndice A

A.1 Resultados en diversas imágenes

A continuación se presentan los resultados obtenidos para cuatro imágenes distintas. Los algoritmos utilizados son los mismos que los del capítulo quinto de este trabajo, excepto que en los casos del SASM y ACA la descomposición *wavelet* fue implementada con *symlet10* y únicamente se propusieron ventanas de 5x5 píxeles. Las imágenes fueron contaminadas con ruido generado según la distribución de Poisson, con $\sigma_1^2 = 10$, $\sigma_2^2 = 20$ y $\sigma_3^2 = 30$. Las imágenes que se muestran en las figuras A.1, A.2, A.3 y A.4 corresponden a los casos en los que la varianza es igual a 20.

PSNR Café.gif (dB)	$\sigma^2 = 10$	$\sigma^2 = 20$	$\sigma^2 = 30$
Matlab (wiener2)	28.0558	22.0999	18.5988
Matlab (wdencmp)	27.9595	22.0782	18.5937
SASM 5x5	39.0423	35.9375	34.1555
ACA 5x5	43.5585	42.3056	41.0594

Tabla A.1 Comparación de los resultados para varios algoritmos para la imagen ‘Café.gif’ contaminada con tres niveles de ruido.

PSNR Gato.gif (dB)	$\sigma^2 = 10$	$\sigma^2 = 20$	$\sigma^2 = 30$
Matlab (wiener2)	28.0952	22.1115	18.6044
Matlab (wdencmp)	28.0819	22.1085	18.6076
SASM 5x5	39.0635	35.9715	34.1957
ACA 5x5	43.4513	42.1717	41.6365

Tabla A.2 Comparación de los resultados para varios algoritmos para la imagen ‘Gato.gif’ contaminada con tres niveles de ruido.

PSNR Lennon.gif (dB)	$\sigma^2 = 10$	$\sigma^2 = 20$	$\sigma^2 = 30$
Matlab (wiener2)	28.0838	22.1181	18.6087
Matlab (wdencmp)	28.0068	22.0866	18.5974
SASM 5x5	39.1015	35.9688	34.1863
ACA 5x5	44.0071	42.0603	41.6815

Tabla A.3 Comparación de los resultados para varios algoritmos para la imagen ‘Lennon.gif’ contaminada con tres niveles de ruido.

PSNR Meteorito.gif (dB)	$\sigma^2 = 10$	$\sigma^2 = 20$	$\sigma^2 = 30$
Matlab (wiener2)	28.0849	22.1188	18.6094
Matlab (wdencmp)	28.0813	22.1101	18.6086
SASM 5x5	39.0157	35.9512	34.1814
ACA 5x5	43.7348	42.9044	42.4615

Tabla A.4 Comparación de los resultados para varios algoritmos para la imagen ‘Meteorito.gif’ contaminada con tres niveles de ruido.

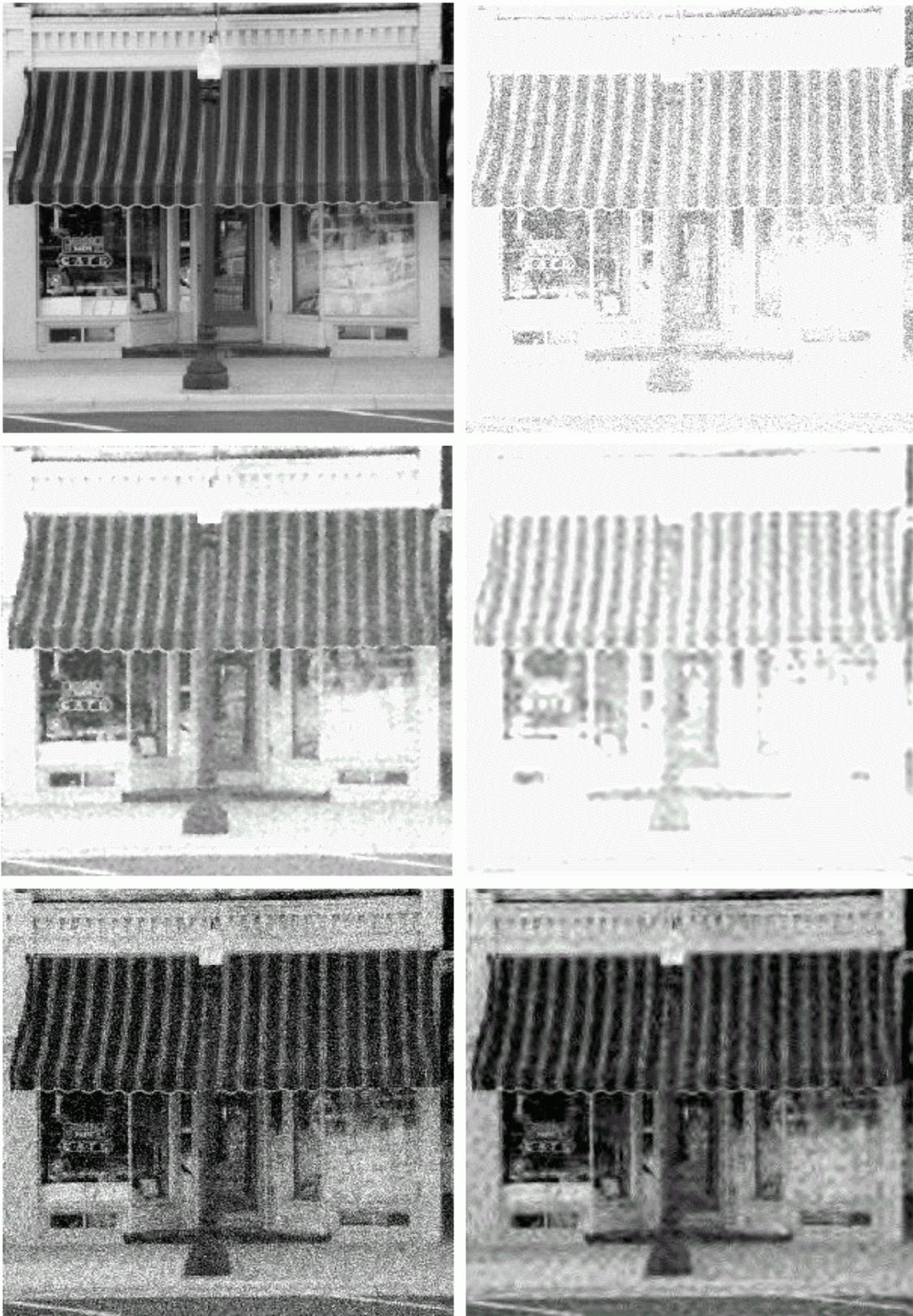


Figura A.1 a) Café.gif, b) imagen con ruido, c) reducción del ruido con Matlab (wiener2), d) con Matlab (wdencomp), e) con SASM 5x5 y f) con ACA 5x5.

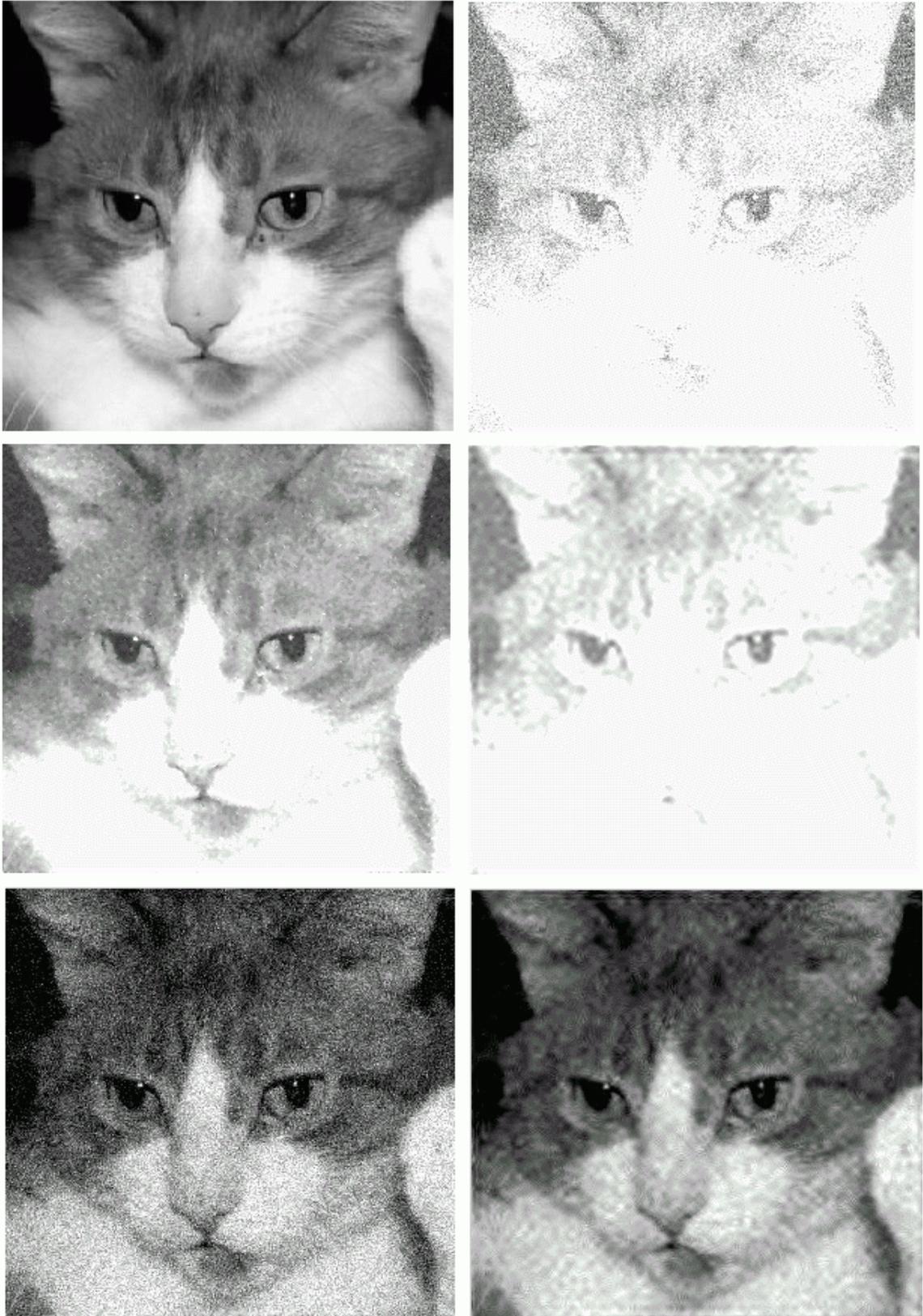


Figura A.2 a) Gato.gif, b) imagen con ruido, c) reducción del ruido con Matlab (wiener2), d) con Matlab (wdencomp), e) con SASM 5x5 y f) con ACA 5x5.

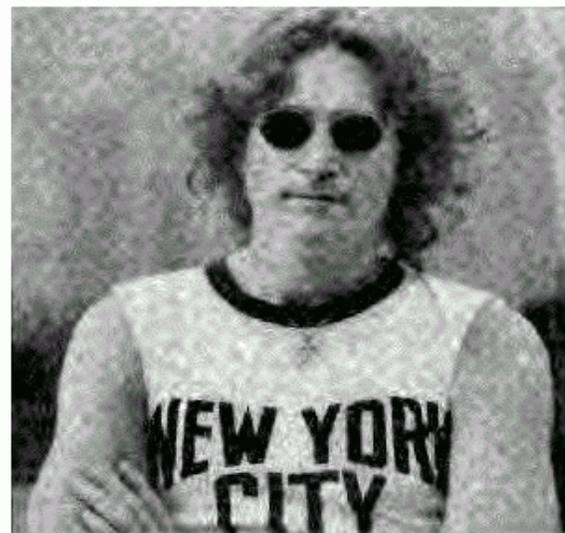
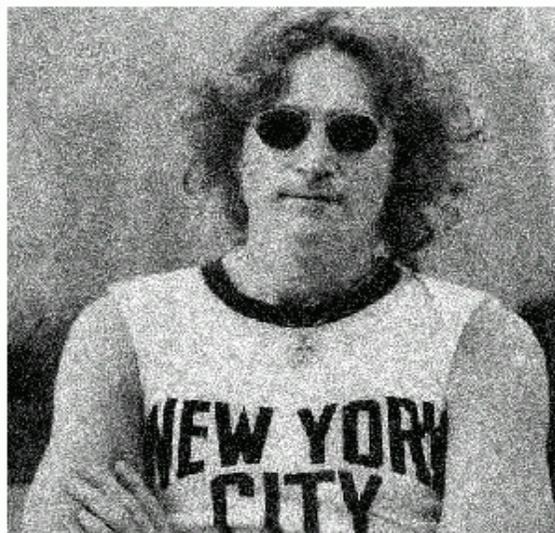
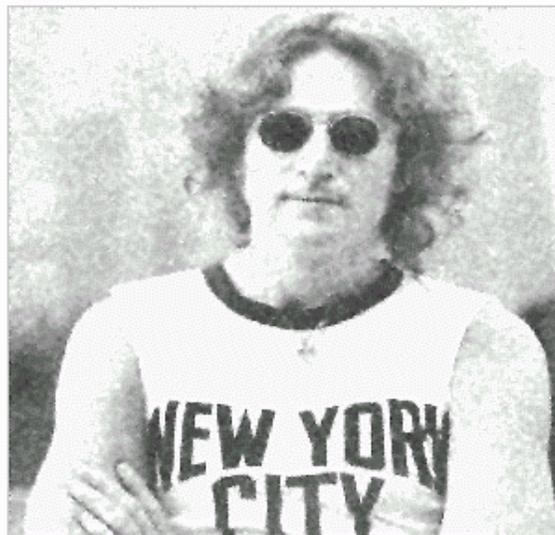


Figura A.3 a) Lennon.gif, b) imagen con ruido, c) reducción del ruido con Matlab (wiener2), d) con Matlab (wdencmp), e) con SASM y f) con ACA 5x5.

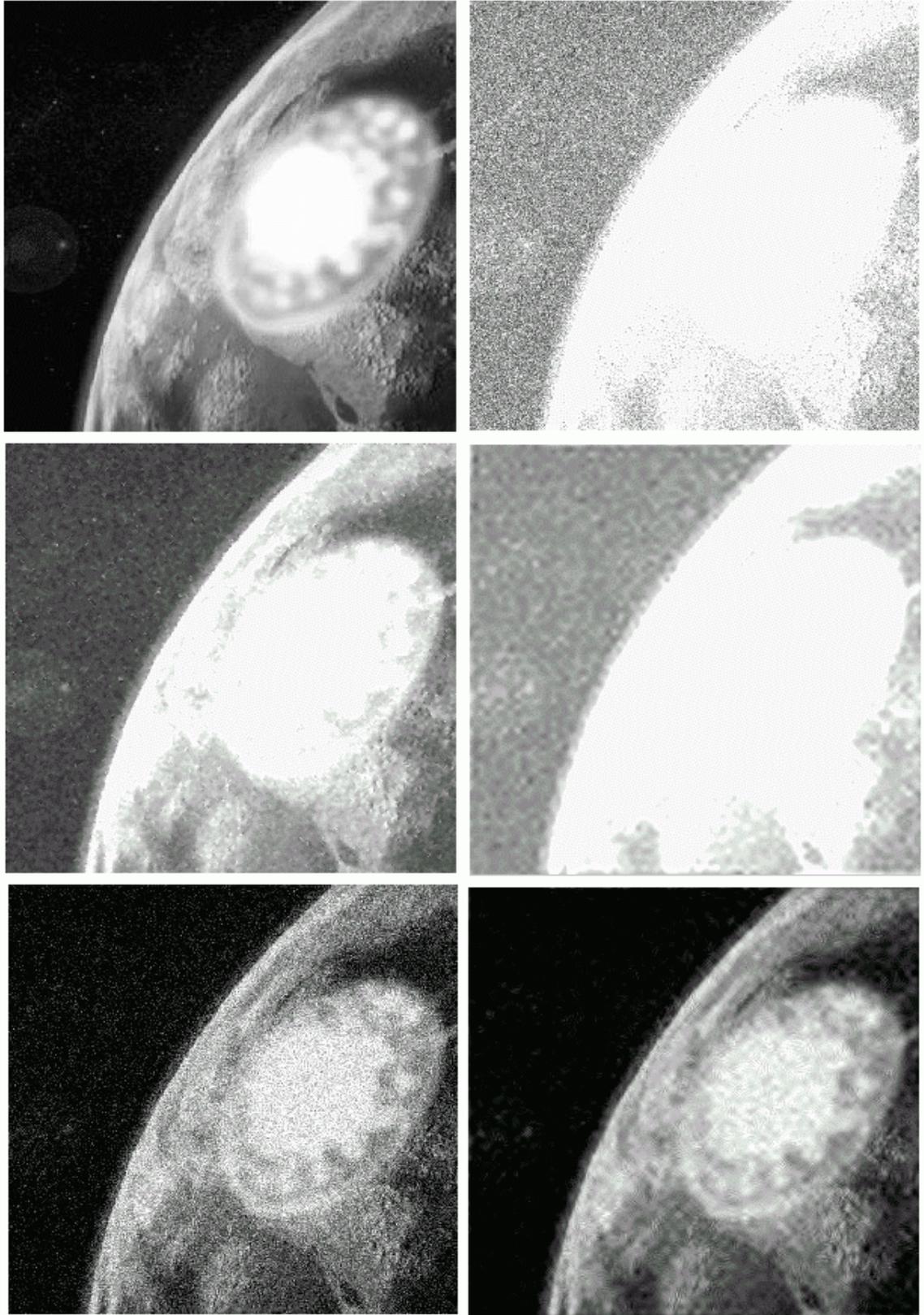


Figura A.4 a) Meteorito.gif, b) imagen con ruido, c) reducción del ruido con Matlab (wiener2), d) con Matlab (wdencmp), e) con SASM y f) con ACA 5x5.

A.2 Pruebas con otros tipos de ruido

Para ejemplificar el funcionamiento del algoritmo desarrollado en este trabajo en imágenes contaminadas con ruido distinto al de Poisson, en las figuras siguientes se muestran, respectivamente, los resultados obtenidos en Meteorito.gif con ruido de distribuciones de Gauss, exponencial y χ cuadrada.

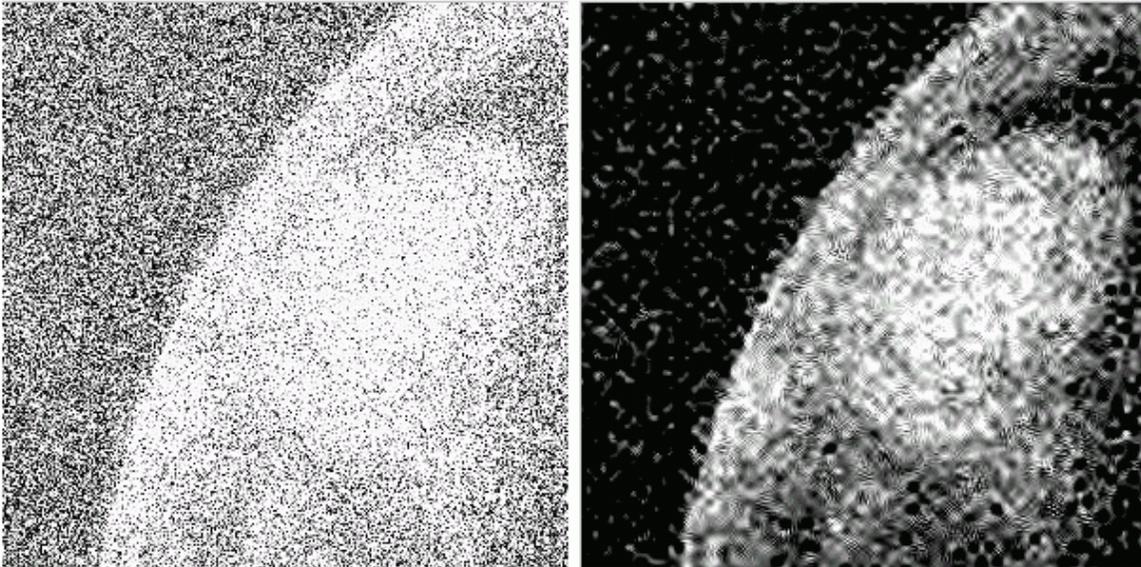


Figura A.5 a) Meteorito.gif con ruido gaussiano con $\mu = 20$ y $\sigma^2 = 20$, b) resultado de la reducción de ruido con ACA 5x5 y *symlet10*. Se obtuvo un PSNR = 26.7583.

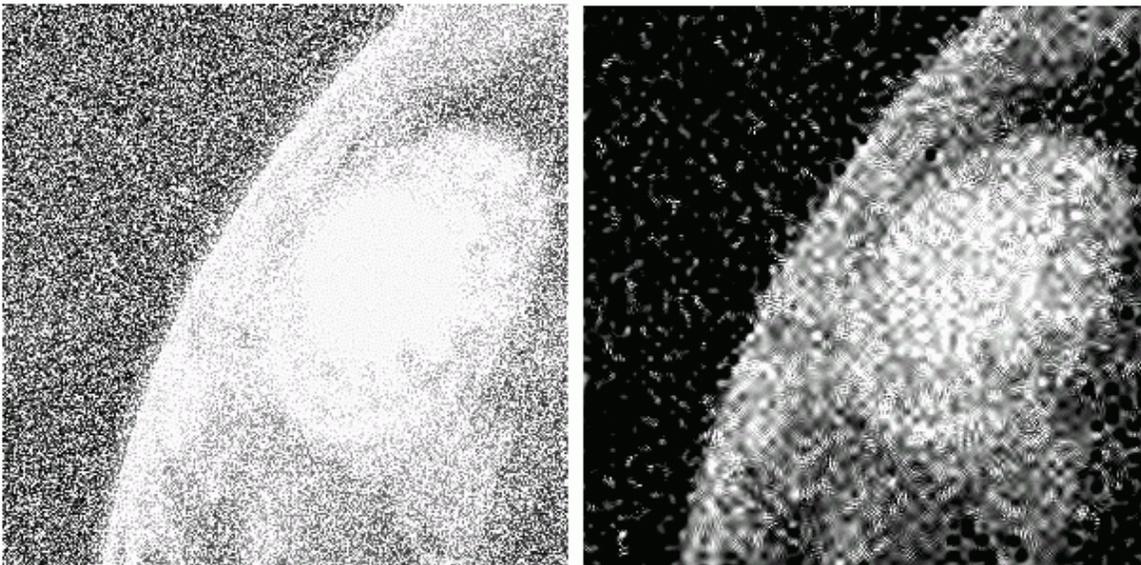


Figura A.6 a) Meteorito.gif con ruido exponencial con $\mu = 20$, b) resultado de la reducción de ruido con ACA 5x5 y *symlet10*. Se obtuvo un PSNR = 26.5415.

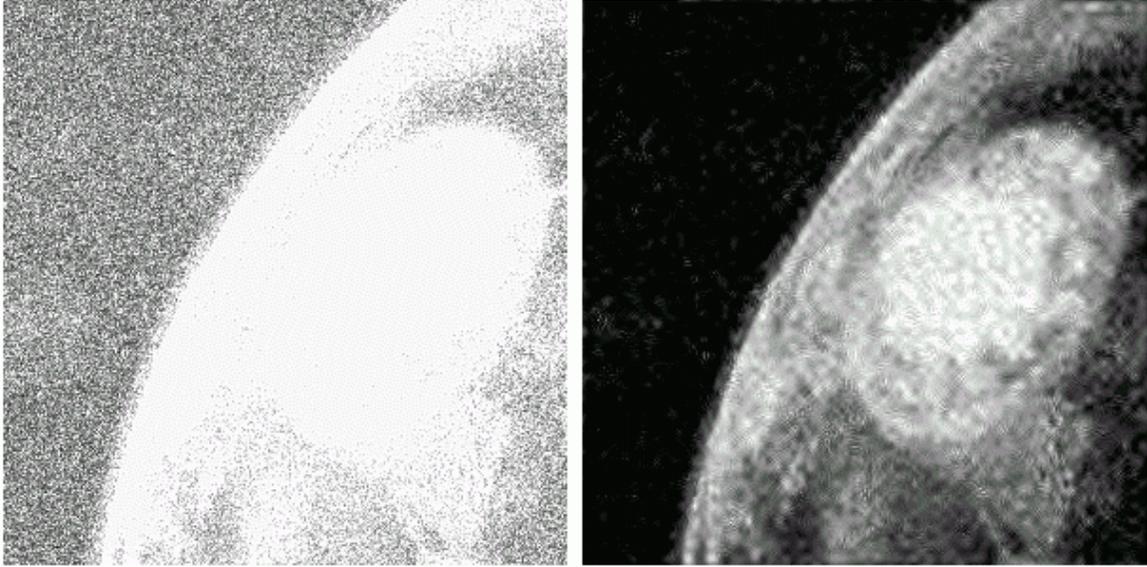


Figura A.7 a) Meteorito.gif con ruido X_i cuadrado con $\mu = 20$ ($\sigma = 20$), b) resultado de la reducción de ruido con ACA 5x5 y *symlet10*. Se obtuvo un PSNR = 41.4361.

Referencias bibliográficas

- [1] Andrews, H. C. & Hunt, B. R. *Digital image restoration*. Prentice Hall. NJ, 1997.
- [2] Crouse, M., et-al. *Wavelet-based statistical signal processing using Hidden Markov Models*. IEEE. Transactions on Signal Processing. Vol. 46. 1998.
- [3] Cruz-González, I., et-al. *Proyecto del Telescopio Infrarrojo Mexicano de nueva tecnología: TIM*. Instituto de Astronomía, UNAM, 1998.
- [4] Daubechies, I. *Ten lectures on wavelets*. Society for Industrial and Applied Mathematics. PA, 1992.
- [5] Farah, A. *Diseño y análisis estructural de la celda para el espejo primario del telescopio TIM, usando el método del elemento finito*. UNAM, 2000.
- [6] González, R. & Woods, R. *Digital image processing*. Adison-Wesley. USA, 1992.
- [7] Hecht, E. *Optics*. Addison-Wesley. USA, 1998.
- [8] Jain, A. K. *Fundamentals of digital image processing*. Prentice Hall. NJ, 1989.
- [9] Lo Presto, S. & Ramchandran, K. *Image coding based on mixture modelling of wavelet coefficients and a fas estimation-quantization framework*. IEEE, Data compresion conference. 1997.
- [10] Mackay, C. D. *Charged-coupled Devices in Astronomy*. Annual Rev. Astronomy and Astrophysics. Vol. 24. 1986.
- [11] Malacara, D., et-al. *Handbook of optical engineering*. Marcel Dekker, Inc. NY, 2001.
- [12] Mallat, S. *A wavelet tour of signal processing*. Academic Press. Cambridge, 1999.
- [13] McLean, I. *Electronic and computer aided Astronomy*. Ellis Horwood series in space science and technology. Sussex, 1994.
- [14] McLean, I. *Electronic imaging in Astronomy: detectors and instrumentation*. Wiley-Praxis, UK, 1997.
- [15] Mihçak, M. K., et-al. *Spatially adaptive statistical modelling of wavelet image coefficients and its application to denoising*. IEEE. Conference in Acoutics, Speech and Signal Processing. Vol. 6. 1999.
- [16] Natarajan, B. *Filtering random noise from deterministic signal via data compression*. IEEE. Transactions on Signal Processing. Vol. 43. 1995.

- [17] Rieke, G. H. *Detection of light*. Cambridge University Press. UK, 1994.
- [18] Rogalsky, A., et-al. *Infrared photon detectors*. SPIE Optican Engineering Press. WA, 1995.
- [19] Salas, L., et-al. *Manual del instrumento cámara infrarroja doble del OAN: CID*. Instituto de Astronomía, UNAM, OAN, 2001.
- [20] Shapiro, J. M. *Embedded image coding zerotrees of wavelet coefficients*. IEEE. Transactions on Signal Processing. Vol. 41. 1993.
- [21] Sirohi, R. & Kothial, M. P. *Optical components, systems and measurement techniques*. Marcel Dekker, Inc. NY, 1991.
- [22] Sohn, E. *Sistema para la detección de imágenes infrarrojas astronómicas*. UIA, 1996.
- [23] Sohn, E., et-al. *Mid-infrared camera/spectrograph for OAN/SPM*. . SPIE Conference on Infrared Astronomical Instrumentation. Vol. 3354. 1998.
- [24] Sohn, E., et-al. *Noise reduction in BIB type detectors*. SPIE Conference on Infrared Astronomical Instrumentation. Vol. 4454. 2001.
- [25] Strobel, N. *Astronomy notes*. McGraw-Hill. CA, 1998.
- [26] Valens, C. *Embedded zerotree wavelet encoding*. USA, 1999.